Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Bootstrap Testing of the Rank of a Matrix via Least-Squared Constrained Estimation

François Portier & Bernard Delyon

To cite this article: François Portier & Bernard Delyon (2014) Bootstrap Testing of the Rank of a Matrix via Least-Squared Constrained Estimation, Journal of the American Statistical Association, 109:505, 160-172, DOI: 10.1080/01621459.2013.847841

To link to this article: https://doi.org/10.1080/01621459.2013.847841

Published online: 19 Mar 2014.

Submit your article to this journal

Article views: 549

View related articles

View Crossmark data

Citing articles: 2 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=uasa20

Bootstrap Testing of the Rank of a Matrix via Least-Squared Constrained Estimation

François PORTIER and Bernard DELYON

To test if an unknown matrix M_0 has a given rank (null hypothesis noted H_0), we consider a statistic that is a squared distance between an estimator $\widehat{M}$ and the submanifold of fixed-rank matrix. Under H_0, this statistic converges to a weighted chi-squared distribution. We introduce the constrained bootstrap (CS bootstrap) to estimate the law of this statistic under H_0. An important point is that even if H_0 fails, the CS bootstrap reproduces the behavior of the statistic under H_0. As a consequence, the CS bootstrap is employed to estimate the nonasymptotic quantile for testing the rank. We provide the consistency of the procedure and the simulations shed light on the accuracy of the CS bootstrap with respect to the traditional asymptotic comparison. More generally, the results are extended to test whether an unknown parameter belongs to a submanifold of the Euclidean space. Finally, the CS bootstrap is easy to compute, it handles a large family of tests and it works under mild assumptions.

KEY WORDS: Dimension reduction; Hypothesis testing; Rank estimation.

1. INTRODUCTION

Let $M_0 \in \mathbb{R}^{p\times H}$ be an unknown matrix. To infer about the rank of M_0 with hypothesis testing, the general framework usually considered is the following: there exists an estimator $\widehat{M} \in \mathbb{R}^{p\times H}$ of M_0 such that

$$n^{1/2}(\widehat{M} - M_0) \xrightarrow{\mathrm{d}} W, \quad \text{with} \quad \mathrm{vec}\,(W) = \mathcal{N}(0, \Gamma), \quad \text{(A1)}$$

where vec$(\cdot)$ vectorizes a matrix by stacking its columns. In the whole article, the hatted quantities are random sequences that depend on the sample number n, all the limits are taken with respect to n. Moreover, there exists an estimator $\widehat{\Gamma}$ such that

$$\widehat{\Gamma} \xrightarrow{\mathbb{P}} \Gamma, \quad \text{(A2)}$$

and in some cases, one may ask that

$$\Gamma \text{ is full rank.} \quad \text{(A3)}$$

Let d_0 be the rank of M_0 and $m \in \{1, \ldots, p\}$, we consider the set of hypotheses

$$H_0 \ : \quad d_0 = m \quad \text{against} \quad H_1 \ : \quad d_0 > m. \quad (1)$$

Thus, d_0 would be estimated in the following way: we start by testing $m = 0$, if H_0 is rejected we go a step further $m := m + 1$, if not we stop the procedure and the estimated rank is $\widehat{d} = m$ (see Robin and Smith 2000 for more details on this procedure). In this article, by considering the hypotheses (1), we focus on each step of this procedure.

To test (1), many different statistics have been proposed in the literature. For instance, Cragg and Donald (1996) introduced a statistic based on the Lower Upper (LU) decomposition of $\widehat{M}$, Kleibergen and Paap (2006) studied the asymptotic behavior of some transformation of the singular values of $\widehat{M}$, and Cragg and Donald (1997) considered the minimum of a squared distance between $\widehat{M}$ and the submanifold of fixed-rank matrix. In some other fields with similar issues, close ideas have been developed: Li (1991) proposed a statistic equal to the sum of squared singular values of $\widehat{M}$, Bura and Yang (2011) examined a normalized version of the Li's statistic, and Cook and Ni (2005) also considered the minimum of a squared distance under rank constraint. For comprehensiveness, in this article, we consider the following three statistics. The first one is introduced in Li (1991) as

$$\widehat{\Lambda}_1 = n \sum_{k=m+1}^{p} \widehat{\lambda}_k^2, \quad (2)$$

where $(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_p)$ are the singular values of $\widehat{M}$ arranged in a descending order. Under H_0 and (A1), this statistic converges in law to a weighted chi-squared distribution (Bura and Yang 2011). The main drawback of such a test is that $\widehat{\Lambda}_1$ is not pivotal, that is, its asymptotic law depends on unknown quantities that are M_0 and Γ. Accordingly, the consistency of the associated test requires assumptions (A1) and (A2). In Bura and Yang (2011), a standardized version of $\widehat{\Lambda}_1$ is studied with

$$\widehat{\Lambda}_2 = n\,\mathrm{vec}(\widehat{Q}_1\widehat{M}\widehat{Q}_2)^T[(\widehat{Q}_2 \otimes \widehat{Q}_1)\widehat{\Gamma}(\widehat{Q}_2 \otimes \widehat{Q}_1)]^+ \times \mathrm{vec}(\widehat{Q}_1\widehat{M}\widehat{Q}_2), \quad (3)$$

where M^+ stands for the Moore–Penrose inverse of M and $\widehat{Q}_1$ and $\widehat{Q}_2$ are, respectively, the orthogonal projectors on the left and right singular spaces of $\widehat{M}$ associated with the $p - m$ smallest singular values. The authors proved that under H_0, if (A1) and (A2) hold, the Wald-type statistic $\widehat{\Lambda}_2$ is asymptotically chi-squared distributed. Besides, Cragg and Donald (1997) and Cook and Ni (2005) proposed

$$\widehat{\Lambda}_3 = n \min_{\mathrm{rank}(M)=m} \|\widehat{\Gamma}^{-1/2}\,\mathrm{vec}(\widehat{M} - M)\|^2, \quad (4)$$

which is also asymptotically chi-squared distributed under H_0, assuming (A1), (A2), and (A3). We will refer the minimum discrepancy approach as in Cook and Ni (2005), where the authors argued for the optimality of this approach. Although the statistics $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$ have the convenience of being asymptotically

François Portier is Post-Doctoral Researcher, Department of Statistics, Université Catholique de Louvain and a member of Institut de Statistique, Biostatistique et Sciences Actuarielles (ISBA), Voie du Roman Pays 20, B-1348 Louvain-La-Neuve, Belgium (E-mail: *francois.portier@gmail.com*). Bernard Delyon is Full Professor, Department of Statistics, Université de Rennes 1 and a member of Institut de Recherche Mathématiques de Rennes (IRMAR), UMR 6625, Campus de Beaulieu, CS 74205, 35042 Rennes Cedex, France (E-mail: *bernard.delyon@univ-rennes1.fr*).

© 2014 American Statistical Association
Journal of the American Statistical Association
March 2014, Vol. 109, No. 505, Theory and Methods
DOI: 10.1080/01621459.2013.847841

Table 1. Values of $\widehat{A}$ and $\widehat{B}$ in (5) and (6) for computing $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$, and $\widehat{\Lambda}_3$

	$\widehat{\Lambda}_1$	$\widehat{\Lambda}_2$	$\widehat{\Lambda}_3$
$\widehat{A}$	I	I	$\widehat{\Gamma}^{-1}$
$\widehat{B}$	I	$[(\widehat{Q}_2 \otimes \widehat{Q}_1)\widehat{\Gamma}(\widehat{Q}_2 \otimes \widehat{Q}_1)]^+$	$\widehat{\Gamma}^{-1}$

pivotal, they both require the inversion of a large matrix and this may cause robustness problems when the sample number is not large enough. For $\alpha \in (0, 1)$ and under the relevant assumptions, each of these statistics $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$, and $\widehat{\Lambda}_3$ is consistent at level $1 - \alpha$ in testing (1), that is, the level goes to $1 - \alpha$ and the power goes to 1 as n goes to ∞.

Nevertheless the estimation of the quantile is difficult because either the asymptotic distribution depends on the data ($\widehat{\Lambda}_1$ is not pivotal) or the true distribution may be quite different than the asymptotic one (slow rates of convergence of $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$). The objective of the article is to propose a bootstrap method for quantile estimation in this context.

An important remark that instigates the sketch of the article is that all the previous statistics share the form

$$\widehat{\Lambda} = n\|\widehat{B}^{1/2}\,\mathrm{vec}(\widehat{M} - \widehat{M}_c)\|^2, \tag{5}$$

with

$$\widehat{M}_c = \underset{\mathrm{rank}(M)=m}{\mathrm{argmin}} \|\widehat{A}^{1/2}\,\mathrm{vec}(\widehat{M} - M)\|^2, \tag{6}$$

and $\|\cdot\|$ is the Euclidean norm, $\widehat{A} \in \mathbb{R}^{pH\times pH}$, $\widehat{B} \in \mathbb{R}^{pH\times pH}$. The values of $\widehat{A}$ and $\widehat{B}$ corresponding to the statistics $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$, and $\widehat{\Lambda}_3$ are summarized in Table 1.

We refer to traditional testing (resp. bootstrap testing) when the statistic is compared to its asymptotic quantile (resp. bootstrap quantile). The bootstrap test is said to be consistent at level α if

$$\mathbb{P}_{H_0}(\widehat{\Lambda} > \widehat{q}(\alpha)) \longrightarrow 1 - \alpha \quad \text{and} \quad \mathbb{P}_{H_1}(\widehat{\Lambda} > \widehat{q}(\alpha)) \longrightarrow 1, \tag{7}$$

where $\widehat{q}(\alpha)$ is the quantile of level α calculated by bootstrap.

The advantages of bootstrap testing with respect to traditional testing can be summarized with the two following arguments. From a practical point of view, the bootstrap approach is simpler than the traditional approach because it alleviates the sometimes difficult derivation of the asymptotic law of the statistic. From a theoretical point of view, bootstrap testing enjoys a high level of accuracy under H_0. This fact is emphasized by considering the two possibilities: when the statistic is pivotal (as $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$) and when the asymptotic law of the statistic depends on unknown quantities (as $\widehat{\Lambda}_1$). First, as highlighted by Hall (1992), when the statistic is pivotal, under some conditions the gap between the distribution of the statistic and its bootstrap distribution is $O_{\mathbb{P}}(n^{-1})$. Since the normal approximation leads to a difference $O(n^{-1/2})$, the bootstrap enjoys a better level of accuracy. Second, if the asymptotic law of the statistic is unknown, the bootstrap appears as an alternative even more convenient because it avoids its estimation. In Hall and Wilson (1991), the authors gave two advice for the use of the bootstrap testing:

(a) Whatever the sample is under H_0 or H_1, the bootstrap estimates the law of the statistic under H_0.
(b) The statistic is pivotal.

The first guideline is the most crucial one because if it fails it may lead to inconsistency of the test (see Hall and Wilson 1991 for some examples). The second guideline aims at improving the accuracy of the test by taking full advantage of the accuracy of the bootstrap.

In this article, we introduce the constrained (CS) bootstrap whose procedure is as follows when testing (1).

The CS bootstrap procedure for testing (1)

Required:
- An estimator $\widehat{M}$ of M_0.
- Bootstrap values of $\sqrt{n}(\widehat{M} - M_0)$ noted W^*.

Test statistic: $\widehat{\Lambda} = n\|\widehat{B}^{1/2}\mathrm{vec}(\widehat{M} - \widehat{M}_c)\|^2$, with $\widehat{M}_c = \underset{\mathrm{rank}(M)=m}{\mathrm{argmin}} \|\widehat{A}^{1/2}\mathrm{vec}(\widehat{M} - M)\|^2$, where $\widehat{A}$ and $\widehat{B}$ are design matrix.

Bootstrapped statistic: $\Lambda^* = n\|B^{*1/2} \times \mathrm{vec}(\widehat{M}_c + n^{-1/2}W^* - M_c^*)\|^2$, with $M_c^* = \underset{\mathrm{rank}(M)=m}{\mathrm{argmin}} \|A^{*1/2} \times \mathrm{vec}(\widehat{M}_c + n^{-1/2}W^* - M)\|^2$, where A^* and B^* are bootstrapped version of $\widehat{A}$ and $\widehat{B}$.

Test: Compare $\widehat{\Lambda}$ to the empirical quantile of the sample of Λ^*.

The choice of the bootstrap for W^* is free. For instance in Section 4, the estimate $\widehat{M}$ will have the form

$$\widehat{M} = n^{-1}\sum_{i=1}^{n} M_i,$$

where (M_i) is a sequence of random variables, and we will use the bootstrap

$$W^* = n^{-1}\sum_{i=1}^{n} w_i(M_i - \widehat{M}),$$

where (w_i) is an iid sequence of random variables with variance 1.

More generally, the CS bootstrap can be used to test whether a parameter belongs to a submanifold of the Euclidean space. As a result, the test (1) becomes a particular case. This more general context does not involve too much difficulties with respect to rank estimation since the statistic has the same form as (5) and so involves *least-squared constraint estimation* (LSCE) (see, e.g., Boos 1990). The sketch of the article is as follows:

- The CS bootstrap in LSCE (Section 2).
- Bootstrap testing procedure in rank estimation for $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$, and $\widehat{\Lambda}_3$ (Section 3).
- Application to *sufficient dimension reduction* (SDR) (Section 4).

2. THE CONSTRAINED BOOTSTRAP IN LSCE FOR HYPOTHESIS TESTING

Because of (5), LSCE has a central place in the article. Moreover, since LSCE intervenes in many statistical fields as M-estimation or hypothesis testing, this section is independent from the rest of the article.

2.1 LSCE

Let $\theta_0 \in \mathbb{R}^p$ be called the parameter of interest, and let $\widehat{\theta} \in \mathbb{R}^p$ be an estimator of θ_0. We define the constrained estimator of θ_0 as

$$\widehat{\theta}_c = \underset{\theta \in \mathcal{M}}{\operatorname{argmin}} \, \|\widehat{A}^{1/2}(\widehat{\theta} - \theta)\|^2, \tag{8}$$

where $\mathcal{M}$ is a submanifold of $\mathbb{R}^p$ with co-dimension q, and $\widehat{A} \in \mathbb{R}^{p \times p}$. The constrained statistic is defined as

$$\widehat{\Lambda} = n\|\widehat{B}^{1/2}(\widehat{\theta} - \widehat{\theta}_c)\|^2, \tag{9}$$

where $\widehat{B} \in \mathbb{R}^{p \times p}$. Note that if $\widehat{A}$ is full rank, the unique minimizer of (8) without constraint is $\widehat{\theta}$, hence it could be understood as the unconstrained estimator. We introduce now the notion of nonsingular point in $\mathcal{M}$. This one is needed to express the Lagrangian first-order condition of the optimization (8). For any function $g = (g_1, \ldots, g_q) : \mathbb{R}^p \to \mathbb{R}^q$, we define its Jacobian as $J_g = (\nabla g_1, \ldots, \nabla g_q)$, where ∇ stands for the gradient operator.

Definition 1. We say that θ is $\mathcal{M}$-nonsingular if $\theta \in \mathcal{M}$ and if there exists a neighborhood V of $\mathbb{R}^p$ and a function $g : V \to \mathbb{R}^q$ continuously differentiable on V with $J_g(\theta)$ of full rank such that $V \cap \mathcal{M} = \{g = 0\}$.

Using the vocabulary of the book of Lee (2003), any point of an embedded submanifold is nonsingular (see Proposition 8.12 of Lee 2003). Moreover, by example 8.14 of Lee (2003), the submanifold $\{M \in \mathbb{R}^{p \times H}, \ \operatorname{rank}(M) = m\}$ is an embedded submanifold. As a consequence, the introduction of the notion of nonsingular points in a submanifold of $\mathbb{R}^p$ is a great generalization for rank testing. We are interested in the hypothesis test

$$H_0 \ : \quad \theta_0 \in \mathcal{M} \quad \text{against} \quad H_1 \ : \quad \theta_0 \notin \mathcal{M}, \tag{10}$$

and the decision rule to reject H_0 if $\widehat{\Lambda}$ is larger than a quantile of its asymptotic law. The previous framework can be seen as an extension of the Wald test statistic that handles the simple hypothesis $\theta_0 = \theta$ with the statistic $n\|A^{-1/2}(\widehat{\theta} - \theta)\|^2$.

2.2 The Bootstrap in LSCE

Since LSCE is a particular case of estimating equation, we shortly review the bootstrap literature with two principal directions: estimating equation and hypothesis testing.

The *Efron's resampling plan* (C bootstrap) has been studied in the M- and Z-estimation theory (see Arcones and Giné 1992; Wellner and Zhan 1996 among others), but such procedures are not adapted to hypothesis testing since it does not satisfy guideline (a).

By contrast, the *biased bootstrap* (B bootstrap) introduced in Hall and Presnell (1999) is directly motivated by the test of equal mean. The original idea of their work is to resample with respect to a distribution that satisfies the constraint. The main drawback of the B bootstrap deals with algorithmic difficulties when computing this distribution. To our knowledge, the study of the B bootstrap has not been extended to other tests than the test of equal mean.

Some other ideas about the bootstrap of Z-estimators can be found in Lele (1991) and Hu and Kalbfleisch (2000) where the *estimating function bootstrap* is studied. The authors also provided a procedure for testing whether $g(\theta_0) = 0$, which is a particular case of (10).

Their bootstrap of $\widehat{\Lambda}$ when $\widehat{A} = \widehat{\Gamma}^{-1}$ is carried out by

$$n(\theta^* - \widehat{\theta})^T J_g(\widehat{\theta})(J_g(\widehat{\theta})^T \Gamma^* J_g(\widehat{\theta}))^{-1} J_g(\widehat{\theta})^T(\theta^* - \widehat{\theta}).$$

Although it verifies the guideline (a), one can see that the good behavior of such an approach is more based on the rank deficiency of $J_g(\widehat{\theta})$ than on the bootstrap of $\sqrt{n}(\widehat{\theta} - \widehat{\theta}_c)$. Indeed $\sqrt{n}(\theta^* - \widehat{\theta})$ bootstraps the nonconstrained estimator $\sqrt{n}(\widehat{\theta} - \theta_0)$. Then as the authors noticed, it is first of all a bootstrap of the Wald-type statistic $n(\widehat{\theta} - \theta_0)^T J_g(\theta_0)\left(J_g(\theta_0)^T \widehat{\Gamma} J_g(\theta_0)\right)^{-1} J_g(\theta_0)^T(\widehat{\theta} - \theta_0)$, which has fortunately the same asymptotic law than the targeted one. This may induce some loss in accuracy. However, it requires the knowledge of the function J_g, which is not the case for fixed-rank constraints where g depends on the limit M_0.

2.3 The Constrained Bootstrap

The CS bootstrap is introduced to solve all the issues we have raised through the previous little review, which are essentially: computational difficulties and small scope of the existing methods. The CS bootstrap targets an estimation $\widehat{q}(\alpha)$ of the quantile under H_0 of $\widehat{\Lambda}$. The consistency of the procedure, that is, (7), forms the main result about the CS bootstrap. Another important issue that occurs beforehand in the section is the bootstrap of the law of

$$n^{1/2}(\widehat{\theta}_c - \theta_0) \quad \text{under } H_0.$$

Basically, we show that a bootstrap of the unconstrained estimator $\sqrt{n}(\widehat{\theta} - \theta_0)$ allows a bootstrap of the constrained estimator $\sqrt{n}(\widehat{\theta}_c - \theta_0)$ under H_0. We point out that the heuristic in CS bootstrap is rather different than the C and EF bootstrap. Otherwise it shares the idea to "reproduce" H_0 even if H_1 is realized with the B bootstrap. Assuming that we can bootstrap $\sqrt{n}(\widehat{\theta} - \theta_0)$, the CS bootstrap calculation of the statistic is realized as follows.

The CS bootstrap procedure

Compute

$$\theta_0^* = \widehat{\theta}_c + n^{-1/2}W^*, \quad \text{with}$$
$$\mathcal{L}_\infty(W^*|\widehat{P}) = \mathcal{L}_\infty(n^{1/2}(\widehat{\theta} - \theta_0)) \quad \text{a.s.}, \tag{11}$$

where W^* can be obtained by any standard bootstrap procedure. Calculate

$$\theta_c^* = \underset{\theta \in \mathcal{M}}{\operatorname{argmin}} \, \|A^{*1/2}(\theta_0^* - \theta)\|^2, \quad \text{and}$$
$$\Lambda^* = n\|B^{*1/2}(\theta_0^* - \theta_c^*)\|^2, \tag{12}$$

where $A^* \in \mathbb{R}^{p \times p}$ and $B^* \in \mathbb{R}^{p \times p}$ (see Theorem 2 for more details).

Intuitively, this choice appears natural because θ_0^* equals $\widehat{\theta}_c$ plus a small perturbation going to 0. Accordingly θ_0^* is somewhat reproducing the behavior of $\widehat{\theta}$ under H_0, especially because W^* has the right asymptotic variance. As we should notice, A^* and B^* could be chosen as $\widehat{A}$ and $\widehat{B}$ but this is not the best choice in practice. As highlighted by Hall (1992), we should normalize by the associated bootstrap quantities (e.g., the variance computed on the bootstrap sample). The following lemma gives a first-order decomposition of the bootstrap law $\sqrt{n}(\theta_c^* - \widehat{\theta}_c)$ under mild conditions. The following lemma is proved in the Appendix.

Lemma 1. Let $\mathcal{M}$ be a submanifold of $\mathbb{R}^p$. Assume there exists $\widehat{\theta}_c \in \mathcal{M}$ and θ_c a $\mathcal{M}$-nonsingular point such that $\widehat{\theta}_c \xrightarrow{\text{a.s.}} \theta_c$. If moreover $\mathcal{L}_\infty(\sqrt{n}(\theta_0^* - \widehat{\theta}_c)|\widehat{P})$ exists a.s. and conditionally a.s. $A^* \xrightarrow{\mathbb{P}} A$ is full rank, then we have conditionally a.s.

$$n^{1/2}(\theta_c^* - \widehat{\theta}_c) = (I - P)n^{1/2}(\theta_0^* - \widehat{\theta}_c) + o_{\mathbb{P}}(1),$$

with $P = A^{-1}J_g(\theta_c)(J_g(\theta_c)^T A^{-1} J_g(\theta_c))^{-1} J_g(\theta_c)^T$.

Note that if θ_0 is $\mathcal{M}$-nonsingular and $\mathcal{L}_\infty(\sqrt{n}(\widehat{\theta} - \theta_0)|\widehat{P})$ exists, we can apply Lemma 1 with $\widehat{\theta}_c = \theta_c = \theta_0$. This gives the following proposition:

Proposition 1. Let $\mathcal{M}$ be a submanifold of $\mathbb{R}^p$. Assume that $\mathcal{L}_\infty(\sqrt{n}(\widehat{\theta} - \theta_0)|\widehat{P})$ exists with θ_0 $\mathcal{M}$-nonsingular. Assume also that $\widehat{A} \xrightarrow{\mathbb{P}} A$ is full rank, then we have

$$n^{1/2}(\widehat{\theta}_c - \theta_0) = (I - P)n^{1/2}(\widehat{\theta} - \theta_0) + o_{\mathbb{P}}(1),$$

with $P = A^{-1}J_g(\theta_0)(J_g(\theta_0)^T A^{-1} J_g(\theta_0))^{-1} J_g(\theta_0)^T$.

Proposition 1 leads easily to the weak convergence of $\widehat{\Lambda}$ under H_0 and extends classical results of Boos (1990) about constrained estimation with constraint $\{g = 0\}$ to manifold type constraints. Besides statements of Lemma 1 and Proposition 1 together explain the preceding definition of θ_0^* in (11). They also lead to the following theorem.

Theorem 1. Let $\mathcal{M}$ be a submanifold of $\mathbb{R}^p$. Assume that $\widehat{\theta} \xrightarrow{\text{a.s.}} \theta_0$ with θ_0 $\mathcal{M}$-nonsingular and $\widehat{A} \xrightarrow{\mathbb{P}} A$ hold. If moreover (11) holds and conditionally a.s. $A^* \xrightarrow{\mathbb{P}} A$ is full rank, then we have

$$\mathcal{L}_\infty(n^{1/2}(\theta_c^* - \widehat{\theta}_c)|\widehat{P}) = \mathcal{L}_\infty(n^{1/2}(\widehat{\theta}_c - \theta_0)) \quad \text{a. s.}$$

Essentially, Theorem 1 is an application of Lemma 1 under H_0, indeed as we saw in the proof of Lemma 1, Equation (A.1), the assumption $\widehat{\theta} \xrightarrow{\text{a.s.}} \theta_0 \in \mathcal{M}$ implies that $\widehat{\theta}_c \xrightarrow{\text{a.s.}} \theta_c$. Nevertheless under H_1 nothing guarantee such a convergence (see Example 1). Roughly speaking, asking for an equality in law under H_1 as in Theorem 1 may be too much to ask. However, as stated in the following theorem, we do not require that $\widehat{\theta}_c$ converges a.s. to a constant to provide that the power of the corresponding test goes to 1. This leads to the consistency of the CS bootstrap for hypothesis testing. For the statement of the consistency theorem, we need to define the quantile function of the bootstrap statistic

$$\widehat{q}(\alpha) = \inf\{x \ : \ \widehat{F}(x) \geq 1 - \alpha\},$$

where $\widehat{F}$ is the cdf of Λ^* conditionally on the sample.

Theorem 2. Let $\mathcal{M}$ be a submanifold of $\mathbb{R}^p$. Assume that $\widehat{\theta} \xrightarrow{\text{a.s.}} \theta_0$ with θ_0 $\mathcal{M}$-nonsingular under H_0. We assume also that $\widehat{A} \xrightarrow{\mathbb{P}} A$ is full rank, $\widehat{B} \xrightarrow{\mathbb{P}} B$ and $\sqrt{n}(\widehat{\theta} - \theta_0)$ converges in law to a distribution having a density. If moreover a.s. $\mathcal{L}_\infty(\sqrt{n}(\theta_0^* - \widehat{\theta}_c)|\widehat{P}) = \mathcal{L}_\infty(\sqrt{n}(\widehat{\theta} - \theta_0))$, and conditionally a.s. $A^* \xrightarrow{\mathbb{P}} A$, $B^* \xrightarrow{\mathbb{P}} B$, then we have

$$\mathbb{P}_{H_0}(\widehat{\Lambda} > \widehat{q}(\alpha)) \longrightarrow 1 - \alpha, \quad \text{and} \quad \mathbb{P}_{H_1}(\widehat{\Lambda} > \widehat{q}(\alpha)) \longrightarrow 1.$$

In other words, the test described in (10) with statistic $\widehat{\Lambda}$ and CS bootstrap calculation of quantile is consistent.

We provide the following example under H_1, where $\widehat{\theta}_c$ does not converge to a constant in probability. Although we cannot get the conclusion of Theorem 1, the least-squared constrained statistic still converges in distribution.

Example 1. Let $(X_i)_{i\in\mathbb{N}}$ be an iid sequence such that $X_1 \overset{\text{d}}{=} \mathcal{N}(0, 1)$. Define $\widehat{\theta} = \overline{X}$, and $H_0 : \ \theta_0^2 = 1$. Clearly H_0 does not hold and naturally the statistic $n \min_{\theta^2=1} \|\widehat{\theta} - \theta\|^2$ goes to infinity in probability. One can find that $\widehat{\theta}_c = \text{sign}(\overline{X})$, which does not converge. Since

$$\theta_c^* = \underset{\theta^2=1}{\text{argmin}} \|\theta_0^* - \theta\|^2 \quad \text{and} \quad \theta_0^* = \widehat{\theta}_c + n^{-1/2}W^*,$$

we get that $\theta_c^* = \widehat{\theta}_c$ a.s. and naturally, we do not have the asymptotic given by Theorem 1. Besides, the convergence to a chi-squared distribution holds for the quantity $n \min_{\theta^2=1} \|\theta_0^* - \theta\|^2$.

3. RANK ESTIMATION WITH HYPOTHESIS TESTING

In this section through a review of the literature about rank estimation, we apply the results obtained in Section 2 to provide a consistent bootstrap procedure for the test described by (1) associated with the statistics $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$, and $\widehat{\Lambda}_3$.

3.1 Application of the Theorem 2

We define $q_0 = p - d_0$ the dimension of the kernel of M_0^T. We denote by $(\lambda_1, \ldots, \lambda_p)$ the singular values of M_0 arranged in descending order and we write the singular value decomposition (SVD) of M_0 as

$$M_0 = (U_1\ U_0)\begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} V_1^T \\ V_0^T \end{pmatrix},$$

with $U_1 \in \mathbb{R}^{p\times d_0}$, $U_0 \in \mathbb{R}^{p\times q_0}$, $V_1 \in \mathbb{R}^{H\times d_0}$, $V_0 \in \mathbb{R}^{H\times q_0}$, and $D_1 = \text{diag}(\lambda_1, \ldots, \lambda_{d_0})$. For $m \in \{1, \cdots, p\}$, we note $q = p - m$ and we write the SVD of $\widehat{M}$ as

$$\widehat{M} = (\widehat{U}_1\ \widehat{U}_0)\begin{pmatrix} \widehat{D}_1 & 0 \\ 0 & \widehat{D}_0 \end{pmatrix}\begin{pmatrix} \widehat{V}_1^T \\ \widehat{V}_0^T \end{pmatrix},$$

with $\widehat{U}_1 \in \mathbb{R}^{p\times m}$, $\widehat{U}_0 \in \mathbb{R}^{p\times q}$, $\widehat{V}_1 \in \mathbb{R}^{H\times m}$, $\widehat{V}_0 \in \mathbb{R}^{H\times q}$, $\widehat{D}_1 = \text{diag}(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_m)$, and $\widehat{D}_0 = \text{diag}(\widehat{\lambda}_{m+1}, \ldots, \widehat{\lambda}_p)$. We also introduce the orthogonal projectors

$$Q_1 = I - P_1 = U_0U_0^T, \ Q_2 = I - P_2 = V_0V_0^T,$$
$$\widehat{Q}_1 = I - \widehat{P}_1 = \widehat{U}_0\widehat{U}_0^T, \text{ and } \widehat{Q}_2 = I - \widehat{P}_2 = \widehat{V}_0\widehat{V}_0^T.$$

Whereas the link between $\widehat{\Lambda}_3$ and LSCE is evident, the one connecting $\widehat{\Lambda}_1$ and $\widehat{\Lambda}_2$ to LSCE relies on the following classical lemma, whose proof is avoided.

Lemma 2. Let $\widehat{M} \in \mathbb{R}^{p\times H}$, it holds that

$$\operatorname*{argmin}_{\operatorname{rank}(M)=m} \|\widehat{M} - M\|_F^2 = \widehat{P}_1 \widehat{M} \widehat{P}_2,$$

$$\text{and} \quad \|\widehat{M} - \widehat{P}_1 \widehat{M} \widehat{P}_2\|_F^2 = \sum_{k=m+1}^{p} \widehat{\lambda}_k^2,$$

where $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_p$ are the singular values of $\widehat{M}$ arranged in descending order, and $\widehat{P}_1$ and $\widehat{P}_2$ are right and left singular orthogonal projectors (uniquely determined if and only if $\widehat{\lambda}_m \neq \widehat{\lambda}_{m+1}$) of $\widehat{M}$ associated with $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_m$.

Remark 1. Using the previous Lemma, we have that each statistic $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$, or $\widehat{\Lambda}_3$ has the form (5) and so belongs to the framework of LSCE. To apply Theorem 2, it remains to note that, by example 8.14 of Lee (2003), the submanifold $\{M \in \mathbb{R}^{p\times H},\ \operatorname{rank}(M) = m\}$ is an embedded submanifold and so any of its point is a nonsingular point (see Definition 1).

3.2 Nonpivotal Statistic

Following Section 2.3 and by using Lemma 2, we define

$$M_0^* = \widehat{P}_1 \widehat{M} \widehat{P}_2 + n^{-1/2} W^* \quad \text{with} \quad W^* | \widehat{P} \xrightarrow{\text{d}} W \quad \text{a. s.}, \quad (13)$$

with W defined in (A1). We introduce the CS bootstrap statistic

$$\Lambda_1^* = n \sum_{k=m+1}^{p} \lambda_k^{*2},$$

with $\lambda_{m+1}^*, \ldots, \lambda_p^*$ the smallest singular values of M^*.

Proposition 2. If (A1), (13), and $\widehat{M} \xrightarrow{\text{a.s.}} M_0$ hold, then the test described in (1) with the statistic $\widehat{\Lambda}_1$ and calculation of quantile with Λ_1^* is consistent.

The next proposition describes the asymptotic behavior of $\widehat{\Lambda}_1 = n \sum_{k=m+1}^{p} \widehat{\lambda}_k^2$. It was stated in Bura and Cook (2001) and some recent extension can be found in Bura and Yang (2011). Our statement goes further because we are also concerned about the estimation of the asymptotic law of $\widehat{\Lambda}_1$, that is, the estimation of the weights that intervenes in the weighted chi-squared asymptotic law. Besides, the proof we give in the Appendix is quite simple because we no longer need the results of Eaton and Tyler (1994) about the asymptotic behavior of singular values.

Proposition 3. (Bura and Cook 2001; Bura and Yang 2011) Under H_0, if (A1) holds we have

$$\widehat{\Lambda}_1 \xrightarrow{d} \sum \nu_k W_k^2,$$

where the ν_k's are the eigenvalues of the matrix $(Q_2 \otimes Q_1)\Gamma(Q_2 \otimes Q_1)$ and the W_k's are iid standard Gaussian variables. If moreover (A2) holds, we have

$$(\widehat{\nu}_1, \ldots, \widehat{\nu}_{pH}) \xrightarrow{\mathbb{P}} (\nu_1, \ldots, \nu_{pH}),$$

where the $\widehat{\nu}_k$'s are the eigenvalues of the matrix $(\widehat{Q}_2 \otimes \widehat{Q}_1)\widehat{\Gamma}(\widehat{Q}_2 \otimes \widehat{Q}_1)$.

Remark 2. Unlike Theorem 1 in Bura and Cook (2001) or Theorem 1 in Bura and Yang (2011), we prefer to state this theorem with the quantities Q_1 and Q_2 rather than with U_0 and V_0. Because we do not assume that the kernel of M has dimension 1, the vectors that form U_0 or V_0 are not unique because vector spaces with dimension larger than 2 have an infinite number of basis. As a consequence it does not make sense to estimate either U_0 or V_0. To characterize convergence of spaces, a suitable object is their associated orthogonal projectors.

Note that in contrast to the traditional test, the bootstrap test no longer requires a consistent estimator of Γ since (A2) is not needed anymore.

3.3 Wald-Type Statistic

The Wald-type statistic $\widehat{\Lambda}_2 = \operatorname{vec}(\widehat{Q}_1 \widehat{M} \widehat{Q}_2)^T [(\widehat{Q}_2 \otimes \widehat{Q}_1)\widehat{\Gamma}(\widehat{Q}_2 \otimes \widehat{Q}_1)]^+ \operatorname{vec}(\widehat{Q}_1 \widehat{M} \widehat{Q}_2)$ has been introduced in Bura and Yang (2011) to get a pivotal statistic. Following (12), we define the associated bootstrap statistic by

$$\widehat{\Lambda}_2^* = \operatorname{vec}(Q_1^* M_0^* Q_2^*)^T [(Q_2^* \otimes Q_1^*)\Gamma^*(Q_2^* \otimes Q_1^*)]^+ \times \operatorname{vec}(Q_1^* M_0^* Q_2^*),$$

where M_0^* is defined in (13), $\Gamma^* \in \mathbb{R}^{pH\times pH}$, $\widehat{Q}_1^*$, and $\widehat{Q}_2^*$ are the eigenprojectors associated with the smallest eigenvalues of $M_0^* M_0^{*T}$ and $M_0^{*T} M_0^*$. As Proposition 2, the following one is an easy application of Theorem 2.

Proposition 4. If (A1), (A2), (13), $\widehat{M} \xrightarrow{\text{a.s.}} M_0$, and $\Gamma^* \xrightarrow{\mathbb{P}} \Gamma$ hold, then the test described in (1) with the statistic $\widehat{\Lambda}_2$ and calculation of quantile with Λ_2^* is consistent.

For traditional testing, Bura and Yang (2011) obtained the following proposition for which we provide a different proof in the Appendix.

Proposition 5. (Bura and Yang 2011) If (A1) and (A2) hold, we have

$$\widehat{\Lambda}_2 \xrightarrow{\text{d}} \chi_s^2,$$

with $s = \min(\operatorname{rank}(\Gamma), (p-d)(H-d))$.

3.4 Minimum Discrepancy Approach

In general, a minimizer

$$\widehat{M}_c = \operatorname*{argmin}_{\operatorname{rank}(M)=m} \|\widehat{\Gamma}^{-1/2} \operatorname{vec}(\widehat{M} - M)\|^2$$

does not have an explicit form as it was for the constrained matrix associated with $\widehat{\Lambda}_1$ and $\widehat{\Lambda}_2$. Therefore, we define

$$M_0^* = \widehat{M}_c + n^{-1/2} W^* \quad \text{with} \quad W^* | \widehat{P} \xrightarrow{\text{d}} W \quad \text{a. s.}, \quad (14)$$

where W is defined in (A1). We also define the associated CS bootstrap statistic

$$\Lambda_3^* = n \min_{\operatorname{rank}(M)=m} \|\Gamma^{*-1/2} \operatorname{vec}(M_0^* - M)\|,$$

and applying Theorem 2 we have the following result.

Proposition 6. If (A1), (A2), (A3), (14), $\Gamma^* \xrightarrow{\mathbb{P}} \Gamma$, and $\widehat{M} \xrightarrow{\text{a.s.}} M_0$ hold, then the test described in (1) with the statistic $\widehat{\Lambda}_3$ and calculation of quantiles with Λ_3^* is consistent.

For traditional testing, noting that $\{\text{rank}(M) = m\}$ has codimension $(H-m)(p-m)$ and applying Proposition (1) we get the following proposition (see Cragg and Donald 1997 for the original proof).

Proposition 7. (Cragg and Donald 1997; Cook and Ni 2005) If (A1), (A2), and (A3) hold, we have

$$\widehat{\Lambda}_3 \xrightarrow{\text{d}} \chi^2_{(H-m)(p-m)}.$$

Remark 3. The set of assumptions needed to obtain Proposition 7 is stronger than the ones stated in propositions 3 and 5 ensuring the convergence of $\widehat{\Lambda}_1$ and $\widehat{\Lambda}_2$. As a consequence this is also true for Proposition 6 with respect to Propositions 2 and 4. The main difference is that we add the assumption on Γ to be nondeficient. This assumption cannot be alleviated in the statement but is not as restrictive in practice. On the one hand, if Γ is deficient the optimization under constraint has a free coordinate that implies the nonconvergence of the minimizer. On the other hand, because of the semidefinite character of Γ the projection of $\widehat{M}$ on the null space of Γ is null. Then one can apply the proposition to the restriction of $\widehat{M}$ on the range of Γ. This is the case in the application to SDR in Section 4.

Remark 4. Unlike $\widehat{\Lambda}_1$ and $\widehat{\Lambda}_2$, an optimization algorithm is needed to obtain $\widehat{\Lambda}_3$ and Λ_3^*, this points out an important issue of such a procedure. In Cook and Ni (2005), the authors noticed that

$$\widehat{\Lambda}_3 = n \min_{A\in H_d, B\in\mathbb{R}^{d\times l}} \|\widehat{\Gamma}^{-1/2}\,\text{vec}(\widehat{M} - AB)\|^2$$

where H_d is the set of orthogonal basis lying in $\mathbb{R}^p$ with dimension d. We follow their algorithm in the computation of $\widehat{\Lambda}_3$ (see Cook and Ni 2005, sec 3.3 for the details).

4. APPLICATION TO SUFFICIENT DIMENSION REDUCTION

Dimension reduction in regression is a modern statistical issue and a great field of application of rank estimation. There exist other applications such as principal component and factor analysis, or time series analysis. We refer to Gill and Lewbel (1992) for more details.

4.1 A Weighted Bootstrap for SIR

We focus on a particularly famous method in SDR called sliced inverse regression (SIR), which has been introduced in Li (1991) to deal with the regression model

$$Y = f(PX, \varepsilon), \tag{15}$$

where $\varepsilon \perp\!\!\!\perp X \in \mathbb{R}^p$, $Y \in \mathbb{R}$, and P is an orthogonal projector on the vector space E with dimension $d_0 < p$, called the central subspace. The objective is to estimate E. If X is elliptically distributed, then we have that $\Sigma^{-1}\mathbb{E}[(X - \mathbb{E}[X])\psi(Y)] \in E$ with $\Sigma = \text{var}(X)$, for any measurable function ψ. Accordingly, to recover the whole central subspace one needs to consider many functions ψ. For a given family of functions $(\psi_h)_{1\le h\le H}$, we define $\Psi = (\psi_1(Y), \ldots, \psi_H(Y))^T$. Under some additional conditions (Portier and Delyon 2013), the image of the matrix $\Sigma^{-1/2}\,\text{cov}(X, \Psi(Y))$ is equal to $\Sigma^{1/2}E$. Then one can make an SVD of an estimator of this matrix to obtain d_0 vectors that form an estimated basis of $\Sigma^{1/2}E$. Motivated by the curse of dimensionality, the estimation of d_0 is one of the most crucial points in SDR. A popular way consists in estimating the rank of $\Sigma^{-1/2}\,\text{cov}(X, \Psi)$ using the hypothesis testing framework given by (1) (see, e.g., Li 1991; Bura and Cook 2001; Cook and Ni 2005). Assume that $((X_1, Y_1), \ldots, (X_n, Y_n))$ is an iid sequence from model (15), denote by $\widehat{P}$ its associated empirical cdf and define the quantity

$$C = \mathbb{E}[K], \quad \text{with } K = (X - \mathbb{E}[X])(\Psi(Y) - \mathbb{E}[\Psi(Y)])^T,$$

associated with its empirical estimator

$$\widehat{C} = \overline{\widehat{K}}, \quad \text{with } \widehat{K}_i = (X_i - \overline{X})(\Psi_i - \overline{\Psi})^T, \quad \text{and } \Psi_i = \Psi(Y_i).$$

We apply the CS bootstrap to calculate the quantiles of each statistic. Facing (13) and (14), we use an independent weighted bootstrap to reproduce the asymptotic law of $\sqrt{n}(\widehat{C} - C)$, that is, we define the bootstrap matrix

$$C^* = \widehat{C}_c + \overline{K^*}, \quad \text{with} \quad K_i^* = w_i(\widehat{K}_i - \overline{\widehat{K}}), \tag{16}$$

where $\widehat{C}_c$ stands for the solution of an optimization problem depending on the selected statistic $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$, or $\widehat{\Lambda}_3$ (see Section 3 for the details) and (w_i) is a sequence of iid random variables. We also define

$$V = \text{var}(\text{vec}(K)) \quad \text{and}$$
$$V^* = \frac{1}{n}\sum_{i=1}^n \text{vec}(K_i^* - \overline{K^*})\,\text{vec}(K_i^* - \overline{K^*})^T.$$

To apply Propositions 2, 4, and 6, we need the following result that is of particular interest since it provides a new bootstrap procedure for SIR that is different than the one proposed in Barrios and Velilla (2007).

Proposition 8. Assume that $\mathbb{E}[\|X\|^2] < +\infty$, $\mathbb{E}[\|\Psi(Y)\|^2]$, and $\mathbb{E}[\|K\|_F^4]$ are finite, if moreover (w_i) is an iid sequence of real random variables with mean 0 and variance 1, then we have

$$\mathcal{L}_\infty(n^{1/2}\,\overline{K^*}|\widehat{P}) = \mathcal{L}_\infty(n^{1/2}(\widehat{C} - C)) \quad \text{a.s. and}$$
$$V^* \xrightarrow{\mathbb{P}} V \quad \text{conditionally a.s.}$$

Remark 5. Defining $\{I(h),\ h = 1, \ldots, H\}$ as a partition of the range of Y we recover the original SIR method with the family formed by the $p_h^{-1/2}\mathbb{1}_{\{Y\in I(h)\}}$'s with $p_h = \mathbb{P}(Y \in I(h))$. Then $C_{\text{SIR}} = \Sigma^{-1/2}\,\text{cov}(X, \mathbb{1})D^{-1/2}$ with $\mathbb{1} = (\mathbb{1}_{\{Y_i\in I(1)\}}, \ldots, \mathbb{1}_{\{Y_i\in I(H)\}})^T$ and $D = \text{diag}(p_h)$, and it is clear that estimated C_{SIR} is more complicated than estimated C. Since we are interested in estimating the rank, we prefer to deal directly with C to avoid the introduction of an additional noise due to the estimation of Σ and D.

Remark 6. For most of the SDR methods such as *sliced average variance estimation* (SAVE) (Cook and Weisberg 1991), *directional regression* (2007) (Li and Wang 2007), and *order 2 optimal function* (2013) (Portier and Delyon 2013), among others, the estimation of the dimension relies on the estimation of the rank of a candidate matrix. This makes the example of SIR quite general. Moreover, the CS bootstrap works for every methods for which a bootstrap of the candidate matrix is available

(for instance, Ye and Weiss 2003 used Efron's bootstrap to calibrate the methods SIR, SAVE, and *principal Hessian direction*, Li 1992).

4.2 Simulation Study

Recall that m is a nonnegative integer, for $k \in \{1, 2, 3\}$ and $B \in \mathbb{N}^*$ we calculate independent copies $\Lambda^*_{k,1}, \ldots, \Lambda^*_{k,B}$ with the CS bootstrap algorithm corresponding to each statistic. Then, we estimate the quantile with

$$q^*_k(\alpha) = \inf_{t\in\mathbb{R}}\{F^*_k(t) > \alpha\} = \Lambda^*_{k,(\lceil B\alpha\rceil)},$$

$$\text{where } F^*_k(t) = B^{-1}\sum_{b=1}^{B} \mathbb{1}_{\{\Lambda^*_{k,b}\le t\}},$$

$\lceil\cdot\rceil$ is the integer ceiling function and $\Lambda^*_{k,(\cdot)}$ stands for the rank statistic associated with the sample $\Lambda^*_{k,1}\ldots\Lambda^*_{k,B}$. On the one hand, we conduct the test described by (1) using the CS bootstrap, that is,

$$H_0 \text{ is rejected if} \qquad \widehat{\Lambda}_k > q^*_k(\alpha). \tag{17}$$

On the other hand, the traditional test is conducted by comparing the statistic $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$ to the quantile of their asymptotic law, respectively, given by propositions 5 and 7. For $\widehat{\Lambda}_1$, by Proposition 3 the limit in law is quite complicated in general (see also Bura and Cook 2001) so that we use approximations: the Wood's approximation (see Wood 1989) as it is computed in the R software, an adjusted version $\widehat{\Lambda}_{1,\text{adj.}} = \widehat{\Lambda}_1/a \xrightarrow{\text{d}} \chi^2_b$, with $a = \sum_{k=1}^s \omega^2/\sum_{k=1}^s \omega_k$, $b = (\sum_{k=1}^s \omega_k)^2/\sum_{k=1}^s \omega_k^2$, and a rescaled version $\widehat{\Lambda}_{1,\text{sc}} = \widehat{\Lambda}_1/c \xrightarrow{\text{d}} \chi^2_s$, $c = \overline{\omega}$ (see Bentler and Xie 2000 for these two corrections).

In all the simulations, we compute the matrix $\widehat{C}$ by taking $\Psi(t) = (\mathbb{1}_{\{y\in I(1),\ldots,y\in I(H)\}})$ where the $I(h)$'s form an equipartition of the range of the data $Y_1, \ldots, Y_n$. In the whole study, we put $(p, H) = (6, 5)$, $B = 1000$ and we consider $n = 50, 100, 200, 500$. Although the parameter H does not really affect the SIR method, we choose it globally good with respect to all the situations.

The first model we study is the following standard model:

$$\text{Model I:} \quad Y = X_1 + .1e \quad \text{with } e \perp\!\!\!\perp X, \quad X \overset{\text{d}}{=} \mathcal{N}(0, I),$$
$$e \overset{\text{d}}{=} \mathcal{N}(0, 1).$$

To highlight guidelines (a) and (b), we produce in Figure 1 two graphics each representing situations under H_0 and H_1 for the statistic $\widehat{\Lambda}_3$. Similar graphics dealing with $\widehat{\Lambda}_2$ have been drawn but are not presented here. On the second one, we see that even if the sample is under H_1, the bootstrap distribution reflects H_0. As a consequence, guideline (a) is satisfied and the power of the bootstrap test is going to 1. The first graph shows that the statistic distribution is closer to the bootstrap distribution than its asymptotic distribution. This has no reason to occur when the statistic is not pivotal (see the Introduction and Hall 1992 for the details). As a consequence, we believe that this good fitting is due to guideline (b).

In Figure 2, we analyze the asymptotic distribution of $\widehat{q}(\alpha)$ in model I for each statistic. To measure the error we consider the behavior of

$$F_n(\widehat{q}(\alpha)),$$

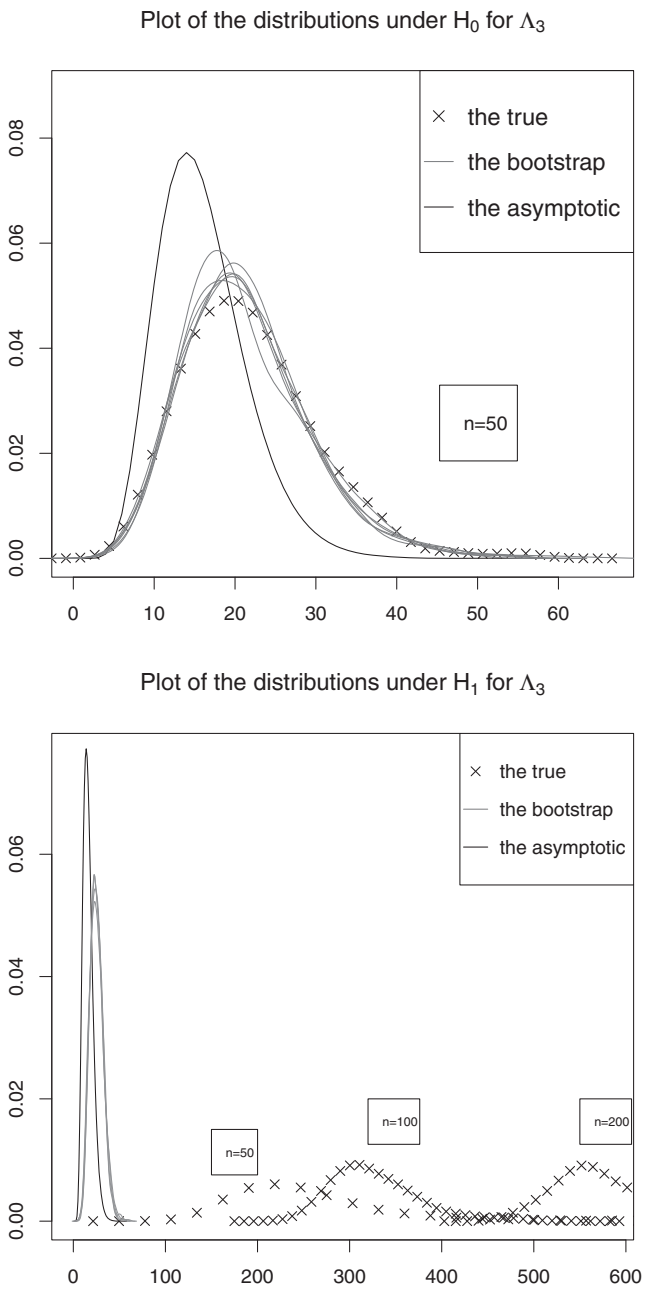

Figure 1. For $\widehat{\Lambda}_3$ in the case of Model I. Plot of the asymptotic distribution, the true distribution, and, for six different samples, the distribution of the bootstrap statistic.

which is optimally equal to $1 - \alpha$. To make that possible, F_n is estimated with a large sample size so that the estimation error is negligible. Then we run over 100 samples the CS bootstrap to provide, for each sample, a bootstrap estimation of the quantile $\widehat{q}(\alpha)$. The associated boxplot for $n = 100, 200, 500$ is provided in Figure 2. As a consequence, we may notice that the behavior of $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$ are quite similar facing the one of $\widehat{\Lambda}_1$. Even if every boxplot argues for convergence to $1 - \alpha$, testing with $\widehat{\Lambda}_1$ seems a better choice when n is small because of a quasi-immediate convergence of the bias. When n increase, this is no longer evident because the variance of either $\widehat{\Lambda}^*_2$ or $\widehat{\Lambda}^*_3$ is smaller.

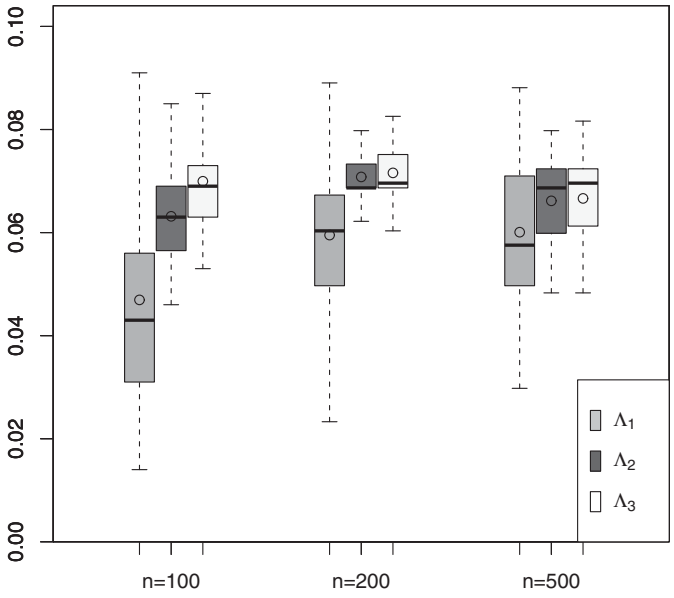

Figure 2. Boxplot over 100 samples of $\widehat{q}(\alpha)$ for $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$, $\widehat{\Lambda}_3$, and $\alpha = 0.95$ in the case of Model I for different values of n.

Furthermore, we go into details in Table 2 by running Model I over 5000 samples. For each of them and every statistic, we conduct the bootstrap test (17) and its traditional version. The table presents for each $m = 0, \ldots, d_0$, the proportion of rejected tests. This corresponds to either estimate of the power or estimate of the level of the test.

Although it has not the best power for $n = 50$, it seems that the most consistent tests over every choice of n are the ones based on $\widehat{\Lambda}_1$. Inside this group, for any sample number, the bootstrap and the rescaled version are the closest to the nominal level. Concerning $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$, the results are quite impressive when n is small: for $n = 50$, whereas traditional testing makes a Type I error 30% of the time, the bootstrap testing goes wrong around 7%. This confirms observation on the second graph of Figure 1.

In Tables 3 and 4, we consider the same model than Model I excepted that we change the distribution of the predictors: in Model Ia, X has independent coordinates with a student distribution with three degrees of freedom, in Model Ib, $X \overset{\mathrm{d}}{=} .1X_1\epsilon + X_2(1-\epsilon)$ with $\epsilon \overset{\mathrm{d}}{=} \mathcal{B}(1/2)$, $X_1 \overset{\mathrm{d}}{=} \mathcal{N}((6, 0, \ldots, 0), I)$, $X_2 \overset{\mathrm{d}}{=} \mathcal{N}(0, I)$. For these two models, we have similar conclusions to Model I concerning $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$, that is, the CS bootstrap really improves the accuracy of the test. For $\widehat{\Lambda}_1$, the rescaled version (which was the most serious competitor of the CS bootstrap in Model I) is not robust to the distribution of the predictors (Table 4, Model Ib).

We introduce a nonlinear relationship by considering the model

$$\text{Model II:} \quad Y = \tanh(X_1) + .1e \quad \text{with } e \perp X, \qquad X \overset{\mathrm{d}}{=} \mathcal{N}(0, I), \quad e \overset{\mathrm{d}}{=} \mathcal{N}(0, 1).$$

In Table 5, we present similar results as before with the difference that the nominal level is $\alpha = 1\%$ to highlight differences in the power of each test. Again, the CS bootstrap induces a large improvement of the accuracy of the test with $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$. At $n = 50$, the tests based on $\widehat{\Lambda}_1$ are less powerful than the others but they are more accurate under H_0. The more accurate under H_0 remains the CS bootstrap with $\widehat{\Lambda}_1$ in every considered sample number. A new important thing is that at $n = 500$, it seems better to use the CS bootstrap with $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$. Actually this is due to the variance of the formers, which is smaller than the variance of Λ_1^* as it was already highlighted in Figure 2.

We conclude by increasing difficulty considering the following model, introduced in Li (1991),

$$\text{Model III:} \quad Y = \frac{X_1}{0.5 + (X_2 + 2)^2} + .1e \quad e \perp X, \qquad X \overset{\mathrm{d}}{=} \mathcal{N}(0, I)$$

We still present in Table 6 the estimated level and power with the nominal level $\alpha = 2\%$ for each test. For such a model, the conclusions are quite mitigated because it induces a trade-off between high power and accurate level. Indeed when n is small, the better powers are provided by the traditional tests with $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$. Nevertheless the more accurate levels can be found looking at the CS bootstrap with $\widehat{\Lambda}_2$ ($n = 100$) or $\widehat{\Lambda}_1$ ($n = 200$). Moreover, the tests associated with $\widehat{\Lambda}_1$ without bootstrap are the worst concerning this model. Accordingly, the simulation study highlighted the good behavior of the CS bootstrap: in every model it improves the accuracy of the traditional test for each statistic. One may remember that the bias of the CS bootstrap with $\widehat{\Lambda}_1$ has the faster rate of convergence with respect to the CS bootstrap of $\widehat{\Lambda}_2$ or $\widehat{\Lambda}_3$. Otherwise, the variance of $\widehat{\Lambda}_1^*$ may be greater than the variance of $\widehat{\Lambda}_2^*$ or $\widehat{\Lambda}_3^*$. Finally, for simple models as Model I, or when the dimension of the matrices are high with respect to the sample number, it seems to be better to use the CS bootstrap with the statistic $\widehat{\Lambda}_1$.

Table 2. Estimated levels and power in Model I for $\alpha = 5\%$

		$\widehat{\Lambda}_1$				$\widehat{\Lambda}_2$		$\widehat{\Lambda}_3$	
n	m	Wood	Resc.	Adj.	CSBoot.	Trad.	CSBoot.	Trad.	CSBoot.
50	0	0.9988	0.9998	0.9988	0.9988	1.0000	1.0000	1.0000	1.0000
	1	0.0326	0.0590	0.0336	0.0494	0.3466	0.0744	0.3098	0.07
100	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0386	0.052	0.0388	0.0456	0.1494	0.0676	0.1466	0.0722
200	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0474	0.055	0.0476	0.0514	0.096	0.0646	0.0954	0.0664
500	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0492	0.0514	0.0494	0.0516	0.0656	0.0584	0.0654	0.0584

Table 3. Estimated levels and power in Model Ia for $\alpha = 5\%$

		$\widehat{\Lambda}_1$				$\widehat{\Lambda}_2$		$\widehat{\Lambda}_3$	
n	m	Wood	Resc.	Adj.	CSBoot.	Trad.	CSBoot.	Trad.	CSBoot.
50	0	0.8294	0.9544	0.8274	0.8394	0.9998	0.9980	0.9998	0.9980
	1	0.0386	0.0798	0.0390	0.0582	0.3544	0.0554	0.2820	0.0546
100	0	0.9746	0.9954	0.9732	0.972	1.000	1.0000	1.0000	1.0000
	1	0.0298	0.0536	0.0300	0.039	0.149	0.0582	0.1342	0.0622
200	0	0.9944	0.9996	0.9938	0.9940	1.0000	1.0000	1.000	1.000
	1	0.0348	0.0552	0.0352	0.0384	0.0862	0.064	0.0818	0.065
500	0	0.9992	1.0000	0.9992	0.9992	1.000	1.0000	1.00	1.0000
	1	0.0332	0.0464	0.0340	0.0408	0.061	0.06	0.0604	0.0608

Table 4. Estimated levels and power in Model Ib for $\alpha = 5\%$

		$\widehat{\Lambda}_1$				$\widehat{\Lambda}_2$		$\widehat{\Lambda}_3$	
n	m	Wood	Resc.	Adj.	CSBoot.	Trad.	CSBoot.	Trad.	CSBoot.
50	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.034	0.1072	0.034	0.0378	0.2122	0.0396	0.1394	0.015
100	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.037	0.0904	0.0374	0.0404	0.0986	0.0572	0.0614	0.0284
200	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0484	0.096	0.0488	0.0518	0.0708	0.066	0.056	0.0506
500	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0486	0.0912	0.0486	0.0490	0.0598	0.0664	0.0612	0.0674

Table 5. Estimated levels and power in Model II for $\alpha = 1\%$

		$\widehat{\Lambda}_1$				$\widehat{\Lambda}_2$		$\widehat{\Lambda}_3$	
n	m	Wood	Resc.	Adj.	CSBoot.	Trad.	CSBoot.	Trad.	CSBoot.
50	0	0.9308	0.9884	0.9428	0.9448	1.0000	0.9988	1.0000	0.9988
	1	0.0036	0.0148	0.0050	0.0086	0.1816	0.0148	0.1404	0.0130
100	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0072	0.0122	0.0082	0.0096	0.0536	0.02	0.0496	0.021
200	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0076	0.0114	0.0086	0.0102	0.0252	0.0192	0.0248	0.02
500	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0068	0.0076	0.007	0.0082	0.012	0.011	0.012	0.011

Table 6. Estimated levels and power in Model III for $\alpha = 2\%$

		$\widehat{\Lambda}_1$				$\widehat{\Lambda}_2$		$\widehat{\Lambda}_3$	
n	m	Wood	Resc.	Adj.	CSBoot.	Trad.	CSBoot.	Trad.	CSBoot.
50	0	0.9950	0.9992	0.9962	0.9960	1.0000	0.9966	1.0000	0.9966
	1	0.3750	0.5342	0.3990	0.4676	0.9074	0.5066	0.8344	0.3270
	2	0.0078	0.0156	0.0086	0.0240	0.0620	0.0164	0.0344	0.0136
100	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.9330	0.9556	0.9368	0.9446	0.9952	0.9842	0.9934	0.9806
	2	0.0134	0.0176	0.0138	0.0210	0.0306	0.0228	0.0266	0.0278
200	0	1.000	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	1.000	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000
	2	0.0154	0.0182	0.0158	0.0198	0.025	0.024	0.0244	0.026
500	0	1.0000	1.000	1.0000	1.0000	1.0000	1.000	1.0000	1.0000
	1	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	2	0.0184	0.0194	0.0184	0.02	0.0228	0.0228	0.0228	0.023

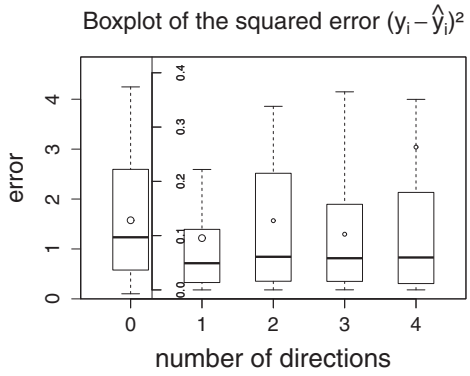

Figure 3. Boxplot of the estimated squared error using the Nadaraya–Watson estimate of the regression with 0 up to 4 directions provided by SIR.

4.3 Case Study: Near Infrared Spectrometry

Datasets resulting from near infrared (NIR) spectrometry are typically high dimensional and it is often necessary to reduce the dimension to analyze such datasets. Since the beginning of chemometrics, NIR spectrometry data have been treated using famous techniques, such as principal component analysis (PCA) or partial least squares (PLS). In the following, we investigate the application of the SIR method to this kind of data.

The data we used were described in Kalivas (1997) and consist of 90 wheat samples with known NIR spectra (explanatory variables) and moisture content (response). The NIR spectra have been reduced from 701 to 141 wavelengths by taking the mean over every five wavelengths. To quantify the level of moisture in the wheat, we try to explain the moisture by the NIR spectra. We randomly split the sample into two subsets. The calibration set contained 50 observations is used to estimate both the central subspace by SIR and the regression function by the Nadaraya–Watson estimate. The validation set contained 40 observations is used to compute the mean squared error (MSE).

Table 7. Results given by the traditional and bootstrap tests for $\widehat{\Lambda}_1$ and $\widehat{\Lambda}_2$ with confidence level $\alpha = 5\%$

	$\widehat{\Lambda}_1$				$\widehat{\Lambda}_2$	
m	Wood	Resc.	Adj.	CSBoot.	Trad.	CSBoot.
0	False	False	False	False	False	False
1	True	True	True	True	False	True
2	True	True	True	True	True	True

Using the calibration set, we run the SIR method with a number of slices equal to 5. The number of directions to keep in the regression is determined by testing the rank of the matrix SIR ($\widehat{C}$ in the previous section) using $\widehat{\Lambda}_1$ and $\widehat{\Lambda}_2$ (the dimension of the predictors was too high to compute reasonably $\widehat{\Lambda}_3$). The results obtained are detailed in Table 7. All the tests based on $\widehat{\Lambda}_1$ give a dimension equal to 1. For $\widehat{\Lambda}_2$, the bootstrap estimate of the dimension is 1 whereas the traditional estimate of the dimension is 2. For each recommended dimension, we compute the associated Nadaraya–Watson estimate of the regression (in each case, the window was chosen minimizing the MSE).

To get an idea of the true dimension, we use the validation set to draw boxplots of the measured error in Figure 3. We compute the error as the squared distance between the true and predicted value of moisture in the validation set.

Since the error does not decrease beyond dimension 1, it seems that the better dimension to used is 1. As a consequence, for this type of datasets the bootstrap of $\widehat{\Lambda}_3$ improves the accuracy of the regression analysis.

We conclude by providing in Figure 4 the plot of the response versus the predictors projected on each first directions of SIR. This gives a simple representation of the data and this also argues in favor of a structural dimension equal to 1. For information, in the same figure, we also provide the plot of the first eigenvector of the matrix SIR.

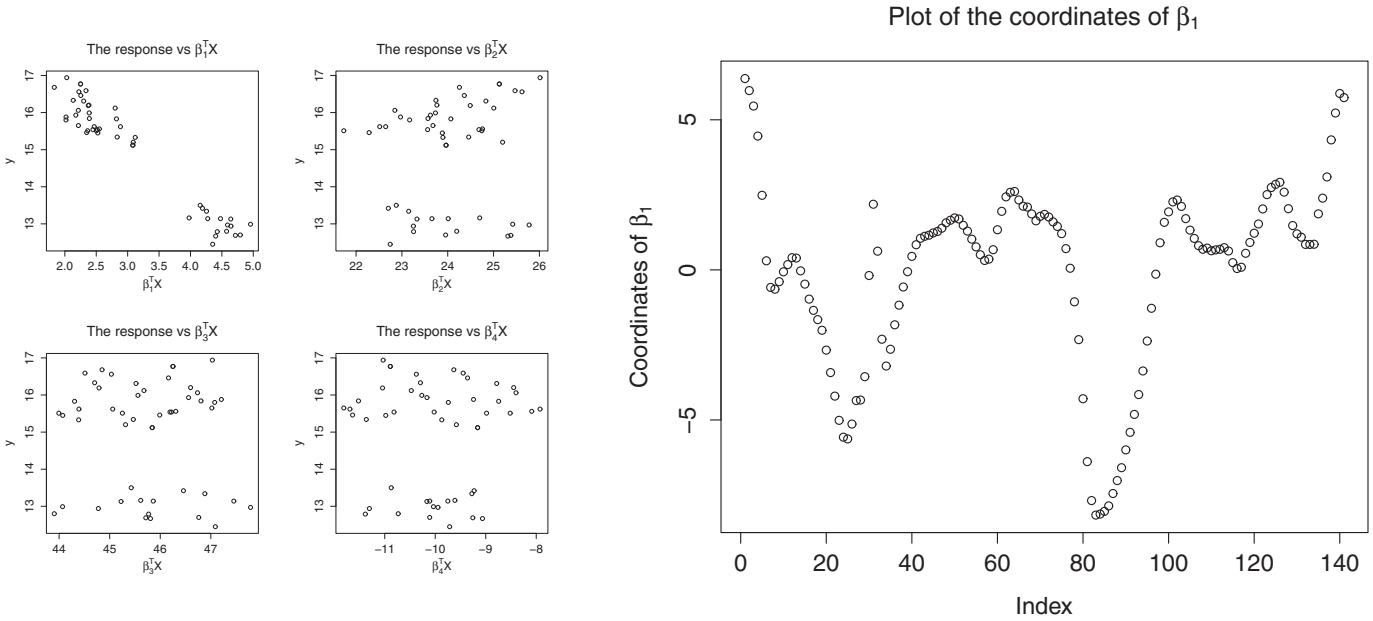

Figure 4. Plot of the data projected on the eigenvectors of the matrix SIR and plot of the principal component of SIR.

5. CONCLUDING REMARKS

Along this study, we found that the main advantages of the CS bootstrap are the following.

1. The CS bootstrap is a powerful alternative to the asymptotic comparison. This argument is even stronger since the asymptotic law can be unknown (or difficult to estimate) or the asymptotic law remains too much different from the statistic law (e.g., large matrix inversion).
2. By Theorem 4, which provides its consistency, the CS bootstrap works under mild assumptions. Essentially, we ask the submanifold to be locally smooth, and we require to be able to bootstrap the unconstrained estimator $\widehat{M}$.
3. Provided that a bootstrap of $\widehat{M}$ is available, the calculation of Λ^* does not involve additional difficulties with respect to $\widehat{\Lambda}$. Indeed, the bootstrap statistic applies the same transformation as the initial statistic (see the box page 161).
4. In the case of rank testing, for each considered statistic, the CS bootstrap improves the accuracy of the traditional test. The simulation study argues for the choice of the statistic $\widehat{\Lambda}_1$ when the sample number is small with respect to the dimension of the matrix.

Besides, there exist some natural extensions of the previous work. First, although it is suitable for testing, the form of the objective function we minimize is quiet restrictive. For example, we believe that the CS bootstrap could be extended to M- and Z-estimation. Second, conditions that guarantee

$$\widehat{q}(\alpha) = q_n(\alpha) + o_{\mathbb{P}}(n^{-1/2})$$

have not been provided yet. This would valid theoretically the use of the CS bootstrap with respect to traditional testing. Third, because of the application of the CS bootstrap to rank estimation, the extension of the CS bootstrap to the high-dimensional context, when p varies with n, is an interesting subject for further study (see, e.g., Feng and He 2009 in a more specific problem).

APPENDIX: PROOFS

Proof of Lemma 1

The whole proof is made conditionally on the sample. By definition of $\widehat{\theta}_c$, with high probability, A^* is full rank for n large enough, we have

$$\begin{aligned}\|A^{*1/2}(\theta_c^* - \theta_c)\| &\le \|A^{*1/2}(\theta_c^* - \theta_0^*)\| + \|A^{*1/2}(\theta_0^* - \theta_c)\| \\ &\le 2\|A^{*1/2}(\theta_0^* - \theta_c)\|. \end{aligned} \tag{A1}$$

Then since $\theta_0^* - \widehat{\theta}_c \xrightarrow{\mathbb{P}} 0$, $\widehat{\theta}_c \to \theta_c$, and because $A^* \xrightarrow{\mathbb{P}} A$ is full rank, one gets that $\theta_c^* \xrightarrow{\mathbb{P}} \theta_c$. Therefore, since θ_c is $\mathcal{M}$-nonsingular and referring to Definition 1, we get

$$\operatorname*{argmin}_{\theta\in\mathcal{M}} \|\Gamma^{*1/2}(\theta_0^* - \theta)\| = \operatorname*{argmin}_{g(\theta)=0} \|\Gamma^{*1/2}(\theta_0^* - \theta)\|,$$

with g continuously differentiable on θ_c and $J_g(\theta_c)$ full rank. By assumption on g, at least for n large enough, θ_c^* satisfies the first-order conditions, that are

$$\begin{cases} A^*(\theta_0^* - \theta_c^*) - J_g(\theta_c^*)\lambda_n^* = 0 \\ g(\theta_c^*) = 0 \end{cases},$$

where λ_n^* is the Lagrange multiplier. Using a Taylor expansion of g around $\widehat{\theta}_c$, we get $g(\theta_c^*) = g(\widehat{\theta}_c) + J_g(\widehat{\theta}_c)^T(\theta_c^* - \widehat{\theta}_c) + o_{\mathbb{P}}(\|\theta_c^* - \widehat{\theta}_c\|)$, and with the previous equations we have

$$\begin{pmatrix} A^* & J_g(\theta_c^*) \\ J_g(\widehat{\theta}_c)^T & 0 \end{pmatrix}\begin{pmatrix} \theta_c^* - \widehat{\theta}_c \\ \lambda_n^* \end{pmatrix} = \begin{pmatrix} A^*(\theta_0^* - \widehat{\theta}_c) \\ o_{\mathbb{P}}(\|\theta_c^* - \widehat{\theta}_c\|) \end{pmatrix}.$$

Now by Slutsky's lemma, we get

$$\begin{pmatrix} A & J_g(\theta_c) \\ J_g(\theta_c)^T & 0 \end{pmatrix}\begin{pmatrix} n^{1/2}(\theta_c^* - \widehat{\theta}_c) \\ n^{1/2}\lambda_n^* \end{pmatrix} = n^{1/2}\begin{pmatrix} A(\theta_0^* - \widehat{\theta}_c) \\ 0 \end{pmatrix} + o_{\mathbb{P}}(1),$$

and the conclusion follows by multiplying on the left by the matrix

$$\left(A^{-1} - PA^{-1}, \quad A^{-1}J_g(\theta_c)(J_g(\theta_c)^T A^{-1} J_g(\theta_c))^{-1}\right)$$

with $P = A^{-1}J_g(\theta_c)(J_g(\theta_c)^T A^{-1} J_g(\theta_c))^{-1} J_g(\theta_c)^T$. □

Proof of Theorem 2

The proof is divided in two parts each corresponding to the level and the power of the test. Assume H_0 and define F_n and F_∞, respectively, as the cdf of $\widehat{\Lambda}$ and the weak limit of F_n. Note that we can apply Proposition 1 to get

$$n^{1/2}\begin{pmatrix} \widehat{\theta} - \theta_0 \\ \widehat{\theta}_c - \theta_0 \end{pmatrix} = n^{1/2}\begin{pmatrix} I \\ I - P \end{pmatrix}(\widehat{\theta} - \theta_0) + o_{\mathbb{P}}(1),$$

and Theorem 1 to get conditionally a.s.

$$n^{1/2}\begin{pmatrix} \theta_0^* - \widehat{\theta}_c \\ \theta_c^* - \widehat{\theta}_c \end{pmatrix} = n^{1/2}\begin{pmatrix} I \\ I - P \end{pmatrix}(\theta_0^* - \widehat{\theta}_c) + o_{\mathbb{P}}(1).$$

with P detailed in the statement of Proposition 1. Using (9), (12), and Slutsky's theorem, we have

$$\mathcal{L}_\infty(\Lambda^*|\widehat{P}) = \mathcal{L}_\infty(\widehat{\Lambda}) \quad \text{a. s.}$$

In other words, with probability 1, $\widehat{F}$ converges pointwise to F_∞. As in van der Vaart (1998) chap. 23, Lemma 3, consider Δ the set of discontinuity of F_∞^{-1}. For every $\alpha \in (0, 1)\backslash\Delta$, we have $\widehat{q}(\alpha) \longrightarrow q(\alpha)$ a.s. (see, e.g., van der Vaart 1998, chap. 21). Using Slutsky's theorem, we get $\mathcal{L}_\infty(\widehat{\Lambda} - \widehat{q}(\alpha)) = \mathcal{L}_\infty(\widehat{\Lambda} - q(\alpha))$, accordingly

$$\mathbb{P}(\widehat{\Lambda} \le \widehat{q}(\alpha)) \longrightarrow F_\infty(q(\alpha)) \quad \text{for all } \alpha \in (0, 1)\backslash\Delta.$$

Because F_∞ is continuous $F_\infty(q(\alpha)) = \alpha$. Since F_∞ is nondecreasing, Δ is denumerable, since $\alpha \mapsto \mathbb{P}(\widehat{\Lambda} \le \widehat{q}(\alpha))$ is nondecreasing with continuous limit, the convergence is uniform and so holds for every $\alpha \in (0, 1)$. This concludes the proof for the level. It remains to show that the power of the test goes to 1. Assume H_1 and let $\alpha \in (0, 1)$, the statistic $\widehat{\Lambda}$ goes to infinity in probability and it suffices to show that with probability 1 the bootstrap quantile $\widehat{q}(\alpha)$ remains bounded. This means exactly that conditionally a.s. the sequence Λ^* is tight. Note that conditionally a.s. we have

$$\Lambda^* \le n\|A^{*1/2}(\widehat{\theta}_c - \theta_0^*)\|^2 = \widetilde{\Lambda}^*,$$

where $\widetilde{\Lambda}^*$ converges in distribution by (11), and is therefore tight. □

Proof of Proposition 3

We have

$$\widehat{\Lambda}_1 = \|n^{1/2}\widehat{Q}_1\widehat{M}\widehat{Q}_2\|_F^2 = \|n^{1/2}\operatorname{vec}(\widehat{Q}_1\widehat{M}\widehat{Q}_2)\|^2.$$

By the Delta method and because H_0 is realized, we can apply convergence results about eigenprojectors to both matrices $\widehat{M}^T\widehat{M}$ and $\widehat{M}\widehat{M}^T$ to obtain the $\sqrt{n}$-convergence for $\widehat{Q}_1$ and $\widehat{Q}_2$. Then we write

$$\begin{aligned} n^{1/2}\widehat{Q}_1\widehat{M}\widehat{Q}_2 &= n^{1/2}\widehat{Q}_1(\widehat{M} - M)\widehat{Q}_2 + n^{1/2}(\widehat{Q}_1 - Q_1)M(\widehat{Q}_2 - Q_2) \\ &= n^{1/2}Q_1(\widehat{M} - M)Q_2 + O_{\mathbb{P}}(n^{-1/2}), \end{aligned}$$

which suffices to obtained the first statement of the theorem. For the second statement, the symmetric matrix $(Q_2 \otimes Q_1)\Gamma(Q_2 \otimes Q_1)$ is estimated consistently by $(\widehat{Q}_2 \otimes \widehat{Q}_1)\widehat{\Gamma}(\widehat{Q}_2 \otimes \widehat{Q}_1)$ and so are its eigenvalues. □

Proof of Proposition 5

We can notice that $\sqrt{n}\widehat{Q}_1\widehat{M}\widehat{Q}_2$ has the same asymptotic law than $\sqrt{n}Q_1(\widehat{M} - M)Q_2$ whose asymptotic variance is consistently estimated by $[(\widehat{Q}_2 \otimes \widehat{Q}_1)\widehat{\Gamma}(\widehat{Q}_2 \otimes \widehat{Q}_1)]^+$ (see the proof of Proposition 3). □

Proof of Proposition 8

Recall that $\widehat{K}_i = (X_i - \overline{X})(\Psi_i - \overline{\Psi})$, $K_i^* = w_i(\widehat{K}_i - \overline{\widehat{K}})$, and define $K_i = (X_i - \mathbb{E}[X])(\Psi_i - \mathbb{E}[\Psi])$. First note that, by Slutsky's theorem, $\sqrt{n}\ \overline{K^*}$ has the same asymptotic law than $n^{-1/2}\sum_{i=1}^n w_i(\widehat{K}_i - \mathbb{E}[K])$. Then we can develop

$$
\begin{aligned}
&n^{-1/2}\sum_{i=1}^n w_i(\widehat{K}_i - \mathbb{E}[K]) \\
&= n^{-1/2}\sum_{i=1}^n w_i((X_i - \mathbb{E}[X])(\Psi_i - \overline{\Psi})^T - \mathbb{E}[K]) \\
&\quad + (\mathbb{E}[X] - \overline{X})n^{-1/2}\sum_{i=1}^n w_i(\Psi_i - \overline{\Psi})^T \\
&= n^{-1/2}\sum_{i=1}^n w_i(K_i - \mathbb{E}[K]) + n^{-1/2}\sum_{i=1}^n w_i(X_i - \mathbb{E}[X])(\mathbb{E}[\Psi] - \overline{\Psi})^T \\
&\quad + (\mathbb{E}[X] - \overline{X})n^{-1/2}\sum_{i=1}^n w_i(\Psi_i - \overline{\Psi})^T.
\end{aligned}
$$

Checking a Lindeberg condition as below to ensure the weak convergence of $n^{-1/2}\sum_{i=1}^n w_i(X_i - \mathbb{E}[X])$ and $n^{-1/2}\sum_{i=1}^n w_i(\Psi_i - \overline{\Psi})^T$, and using the Slutsky's theorem we get conditionally a.s.

$$
n^{1/2}\ \overline{K^*} = n^{-1/2}\sum_{i=1}^n w_i(K_i - \mathbb{E}[K]) + O_{\mathbb{P}}(n^{-1/2}).
$$

We can apply the multidimensional version of the Lindeberg's central limit theorem (see, e.g., Bhattacharya and Rao 1976, Corollary 18.2), provided that

$$
\frac{1}{n}\sum_{i=1}^n \mathbb{E}[\|\widehat{V}^{-1/2}w_i\xi_i\|^2 \mathbb{1}_{\{\|\widehat{V}^{-1/2}w_i\xi_i\| > \nu n^{1/2}\}} | \widehat{P}] \xrightarrow{\text{a.s.}} 0,
$$

where $\xi_i = \mathrm{vec}(K_i - \mathbb{E}[K])$ and $\widehat{V} = \frac{1}{n}\sum_{i=1}^n (\xi_i - \overline{\xi})(\xi_i - \overline{\xi})^T$. The above convergence is a consequence of the Lebesgue domination theorem, which ensures that each term of the sum goes to 0, afterward we can conclude by the Cesaro's lemma. Thus, we have proved that conditionally a.s.

$$
n^{-1/2}\widehat{V}^{-1/2}\sum_{i=1}^n w_i\xi_i \xrightarrow{\text{d}} \mathcal{N}(0, I),
$$

and it remains to note that $\widehat{V} \xrightarrow{\text{a.s.}} V$ the variance of the limit in law of $\sqrt{n}(\widehat{C} - C)$ provided that K has a finite order 2 moment. For the second convergence, we note that conditionally a.s.

$$
V^* - \widehat{V} = \frac{1}{n}\sum_{i=1}^n (w_i^2 - 1)\xi_i\xi_i^T + o_{\mathbb{P}}(1),
$$

then by noting v_i a coordinate of $\xi_i\xi_i^T$, we calculate

$$
\mathbb{E}\left[\left(n^{-1}\sum_{i=1}^n (w_i^2 - 1)v_i\right)^2\right] = n^{-2}\mathbb{E}\left[(w_i^2 - 1)^2\right]\sum_{i=1}^n v_i^2,
$$

which goes to 0 a.s. provided that K has a finite order 4 moment. We conclude by using the Markov inequality to get that $V^* \xrightarrow{\mathbb{P}} \widehat{V}$ conditionally a.s. □

[Received January 2013. Revised September 2013.]

REFERENCES

Arcones, M. A., and Giné, E. (1992), "On the Bootstrap of *M*-Estimators and Other Statistical Functionals," in *Exploring the Limits of Bootstrap (East Lansing, MI, 1990), Wiley Series in Probability and Statistics*, eds. R. LePage and L. Billard, New York: Wiley, pp. 13–47. [162]

Barrios, M. P., and Velilla, S. (2007), "A Bootstrap Method for Assessing the Dimension of a General Regression Problem," *Statistics and Probability Letters*, 77, 247–255. [165]

Bentler, M. P., and Xie, J. (2000), "Corrections to Test Statistics in Principal Hessian Directions," *Statistics and Probability Letters*, 47, 381–389. [166]

Bhattacharya, R. N., and Rao, R. R. (1976), *Normal Approximation and Asymptotic Expansions (Wiley Series in Probability and Mathematical Statistics)*, New York-London-Sydney: Wiley. [171]

Boos, D. D. (1990), "On Generalized Score Tests," *The American Statistician*, 46, 327–333. [161,163]

Bura, E., and Cook, R. D. (2001), "Extending Sliced Inverse Regression: The Weighted Chi-Squared Test," *Journal of the American Statistical Association*, 96, 996–1003. [164,165,166]

Bura, E., and Yang, J. (2011), "Dimension Estimation in Sufficient Dimension Reduction: A Unifying Approach," *Journal of Multivariate Analysis*, 102, 130–142. [160,164]

Cook, R. D., and Ni, L. (2005), "Sufficient Dimension Reduction via Inverse Regression: A Minimum Discrepancy Approach," *Journal of the American Statistical Association*, 100, 410–428. [160,165]

Cook, R. D., and Weisberg, S. (1991), Discussion of "Sliced Inverse Regression for Dimension Reduction," by Ker-Chau Li, *Journal of the American Statistical Association*, 86, 28–33. [165]

Cragg, J. G., and Donald, S. G. (1996), "On the Asymptotic Properties of LDU-Based Tests of the Rank of a Matrix," *Journal of the American Statistical Association*, 91, 1301–1309. [160]

——— (1997), "Inferring the Rank of a Matrix," *Journal of Econometrics*, 76, 223–250. [160,165]

Eaton, M. L., and Tyler, D. (1994), "The Asymptotic Distribution of Singular Values With Applications to Canonical Correlations and Correspondence Analysis," *Journal of Multivariate Analysis*, 50, 238–264. [164]

Feng, X., and He, X. (2009), "Inference on Low-Rank Data Matrices With Applications to Microarray Data," *Annals of Applied Statistics*, 3, 1634–1654. [170]

Gill, L., and Lewbel, A. (1992), "Testing the Rank and Definiteness of Estimated Matrices With Applications to Factor, State-Space and ARMA Models," *Journal of the American Statistical Association*, 87, 766–776. [165]

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion (Springer Series in Statistics)*, New York: Springer-Verlag. [161,163,166]

Hall, P., and Presnell, B. (1999), "Intentionally Biased Bootstrap Methods," *Journal of the Royal Statistical Society,* Series B, 61, 143–158. [162]

Hall, P., and Wilson, S. R. (1991), "Two Guidelines for Bootstrap Hypothesis Testing," *Biometrics*, 47, 757–762. [161]

Hu, F., and Kalbfleisch, J. D. (2000), "The Estimating Function Bootstrap" (with discussion and rejoinder by the authors), *Canadian Journal of Statistics*, 28, 449–499. [162]

Kalivas, J. H. (1997), "Two Data Sets of Near Infrared Spectra," *Chemometrics and Intelligent Laboratory Systems*, 37, 255–259. [169]

Kleibergen, F., and Paap, R. (2006), "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97–126. [160]

Lee, J. M. (2003), *Introduction to Smooth Manifolds (Graduate Texts in Mathematics)* (Vol. 218), ed. V. P. Godambe, New York: Springer-Verlag. [162,164]

Lele, S. (1991), "Resampling Using Estimating Equations," in *Estimating functions (Oxford Statistical Science Series)* (Vol. 7), New York: Oxford University Press, pp. 295–304. [162]

Li, B., and Wang, S. (2007), "On Directional Regression for Dimension Reduction," *Journal of the American Statistical Association*, 102, 997–1008. [165]

Li, K.-C. (1991), "Sliced Inverse Regression for Dimension Reduction," *Journal of the American Statistical Association*, 86, 316–342. [160,165,167]

——— (1992), "On Principal Hessian Directions for Data Visualization and Dimension Reduction: Another Application of Stein's Lemma," *Journal of the American Statistical Association*, 87, 1025–1039. [166]

Portier, F., and Delyon, B. (2013), "Optimal Transformation: A New Approach for Covering the Central Subspace," *Journal of Multivariate Analysis*, 115, 84–107. [165]

Robin, J.-M., and Smith, R. J. (2000), "Tests of Rank," *Econometric Theory*, 16, 151–175. [160]

van der Vaart, A. W. (1998), *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)* (Vol. 3), Cambridge: Cambridge University Press. [170]

Wellner, J. A., and Zhan, Y. (1996), "Bootstraping z-Estimators," Technical Report 308, University of Washington. [162]

Wood, A. T. A. (1989), "An f Approximation to the Distribution of a Linear Combination of Chi-Squared Variables," *Communications in Statistics—Simulation and Computation*, 18, 1439–1456. [166]

Ye, Z., and Weiss, R. E. (2003), "Using the Bootstrap to Select One of a New Class of Dimension Reduction Methods," *Journal of the American Statistical Association*, 98, 968–979. [166]