Continuous inverse regression

François Portier *

Institut de Statistique, Biostatistique et Sciences Actuarielles (ISBA), Université catholique de Louvain, Belgique

September 3, 2014

Abstract: We provide new theoretical results in the field of inverse regression methods for dimension reduction. Our approach is based on the study of some empirical processes that lie close to a certain dimension reduction subspace, called the central subspace. The study of these processes essentially includes weak convergence results and the consistency of some general bootstrap procedures. While such properties are used to obtain new results about sliced inverse regression, they mainly allow to define a natural family of methods for dimension reduction.

First the estimation methods are shown to have root $n$ rates and the bootstrap is proved to be valid. Second, we describe a family of Cramér-von Mises test statistics that can be used in testing structural properties of the central subspace or the significance of some sets of predictors. We show that the quantiles of those tests could be computed by bootstrap. Most of the existing methods related to inverse regression involve a slicing of the response that is difficult to select in practice. While our approach guarantee a comprehensive estimation, the slicing is no longer needed.

Key words: Dimension reduction; Sliced inverse regression; Cumulative slicing estimation; Test; Weak convergence in $L^\infty(\mathbb{R})$; Bootstrap.

1 Introduction

Dimension reduction is a powerful tool usually employed to synthesise the dependence between two sets of random variables, say $(X,Y)$ where $X \in \mathbb{R}^p$ is called the vector of predictors and $Y \in \mathbb{R}$ is the variable to explain. Dimension reduction can be used to visualize the dependence in high dimensional data [4], as well as to construct accurate estimators of the conditional distribution of $Y$ knowing $X$ or the regression of $X$ on $Y$ [12]. The most common way to model dimension reduction is to assume a certain structure on the conditional distribution of $Y$ given $X$ (see for instance the introduction of [4]). In this paper we assume that the joint distribution of $(X,Y)$ satisfies

$$P(Y \in A|X) = P(Y \in A|\beta_0^T X),$$

for every Borel set $A \in \mathbb{R}$, for some $\beta_0 \in \mathbb{R}^{p \times d_0}$ that has minimal dimension $d_0$. The objective is to estimate the vector $\beta_0$. For identifiability reason [18], we better estimate the subspace generated by $\beta_0$, that is called the central subspace. To this typical semi-parametric problem, many different approaches have been investigated [12], [14], [18], [5]. In this paper we follow

*Electronic address: francois.portier@gmail.com
the idea of inverse regression introduced by Li [18]. Inverse regression might suffer from strong theoretical restriction on $X$ but it often leads to estimators that are very accurate as well as computationally efficient. Inverse regression methods are widely used because they provide a reasonable trade-off between accuracy and complexity.

Throughout the paper, we will assume that the central subspace is unique. This is known to be true as soon as $X$ has a density ([22], Theorem 1). For more clarity in the statements we introduce the standardized predictors $Z = \Sigma^{-1/2}(X - E(X))$ with $\Sigma = \text{var}(X)$. The standardized central subspace, generated by $\Sigma^{1/2}\beta_0$ is noted $E_c$.

Inverse regression is based on the following assumption. We say that $X$ satisfies the linearity condition if

$$E(Z|PZ) = PZ,$$

(LC)

examples include the Gaussian distribution, the uniform distribution on the sphere, or more generally the class of spherical variables [10]. Li noticed in [18] that under (1) and (LC),

$$E(Z|Y) \in E_c,$$

(2)

with probability 1. The author proposed to estimate $E_c$ by estimating the subspace generated by $\text{var}(E(Z|Y))$. The estimation is realized through a slicing of the response $Y$. This approach in addition to be simple and accurate is flexible with respect to (LC) [10]. Other methods based on similar characterisation have been considered as a polynomial approach [28], the minimum discrepancy approach (MDA) [5], or the cumulative inverse regression (CUME) [31].

When facing regression models with a symmetric link function (also called SIR pathology), Li suggested in [18] to use second order moments of the predictors. Based on this idea, some authors have introduced order 2 moments methods as for instance sliced average variance estimation (SAVE) [6], directional regression [17] and order 2 optimal function [22]. These methods require an additional assumption called the constant covariance condition,

$$\text{var}(Z|PZ) = \text{const.},$$

(CCV)

they are based on the result that, under (1), (LC) and (CCV), it holds

$$\text{var}(Z|Y) - I \in E_c,$$

(3)

where $I$ is the unitary matrix.

As a consequence, the current literature have put the focus on the estimation of subspaces that are generated by conditional quantities. A natural issue which arises is to know whether a nonparametric estimation is really necessary. On the one hand, some authors have studied the rates of convergences and the limiting distribution of the SIR estimators as the slicing becomes more strength [15], [30]. A kernel smoothing approach has been considered in [29] and the asymptotics for the SAVE estimator, when the length of each slices goes to 0 is described in [20]. Though the conditional quantities $E(Z|Y)$ or $\text{var}(Z|Y)$ can not be estimated at rates root $n$, these authors shown that the rate root $n$ is in fact available when estimating moments of these quantities, as for instance $\text{var}(E(Z|Y))$. One the other hand, other authors have kept fix the number of slices so that the length of the slices is not forced to go to 0 [5] and [22]. One might argue that since

$$E[Z\psi(Y)] \in E_c,$$

(4)

for any measurable function $\psi$ such that $E[Z\psi(Y)] < \infty$, the whole space $E_c$ can be recovered by considering sufficiently many different functions $\psi$. In this paper, we consider the estimation.
of \( E_c \) when \( \psi \) describe a given class of function. This leads us to an empirical process view of the problem rather than to a local estimation procedure.

The main contribution of the paper is the introduction and the study of two empirical processes that become close to \( E_c \) as the number of observations increases, one is based on the first conditional moments of \( Z \) knowing \( Y \) and the other one rely on the second conditional moments of \( Z \) knowing \( Y \). Let \( \Phi : \mathbb{R} \rightarrow \mathbb{R} \) be a non-decreasing function and denote by \( \Phi^- \) its generalized inverse, given by

\[
\Phi^-(u) = \inf\{t \in \mathbb{R} : \Phi(t) \geq u\},
\]

for every \( u \in \mathbb{R} \). We define the first moment process as

\[
m_\Phi(u) = \mathbb{E}(Z1_{\{Y < \Phi^-(u)\}}),
\]

and the second moment process as

\[
M_\Phi(u) = \mathbb{E}((ZZ^T - I)1_{\{Y < \Phi^-(u)\}}),
\]

for each \( u \in \mathbb{R} \). Clearly, under (\text{LC}) and \text{(CCV)}, \( m_\Phi(u) \in E_c \) and if moreover \text{(CCV)} holds then \( M_\Phi(u) \in E_c \), for every \( u \in \mathbb{R} \).

The fact that \( m_\Phi \) and \( M_\Phi \) are indexed by the class of indicator functions plays a key role in our analysis. First the class of indicators is large enough to ensure an exhaustive characterization of \( E_c \). Second it is well known that indicators enjoy good metric entropy properties and as a consequence, many asymptotic properties might be derived using the theory of empirical process developed for instance in [27]. Such properties include weak convergence of estimators of \( m_\Phi \) and \( M_\Phi \) with root \( n \) rates, and the validity of some general weighted bootstrap procedures.

The function \( \Phi \) is a user-selected function. As in copula modelling, to alleviate the effect of the marginal distribution of \( Y \) in the estimation, it is convenient to “uniformize” the variable \( Y \). This is done by choosing \( \Phi \) equal to the cumulative distribution function (c.d.f.) of \( Y \). This choice involves a little more technicalities in the proof but it leads to some accurate and computably simple rank-based estimator. In [9], the authors studied the weak convergence of the empirical copula process. Following their approach, our theoretical study is based on both the delta-method for stochastic processes and on a “trick” allowing us to consider \( Y \) as uniformly distributed (see Remark 1).

An almost direct consequence of the study of estimators of \( m_\Phi \) and \( M_\Phi \) is the computation of the asymptotic law of SIR and SAVE estimators when the number of slices is fix and each slice contains the same number of observations. This was the case in the original paper by Li [18] (see Remark 4.2) and this is the common way to compute SIR and SAVE. To our knowledge, such results are new in the literature.

Moreover, the study of estimators of \( m_\Phi \) and \( M_\Phi \) is also needed to derive some results concerning a new dimension reduction method we call continuous inverse regression (CIR). The methods of CIR are integral based methods that approximate \( E_c \) through the range of the matrices

\[
\int m_{id}(t)m_{id}(t)^T d\Phi(t) \quad \text{and} \quad \int M_{id}(t)^2 d\Phi(t)
\]

where \( id \) denotes the function \( t \mapsto t \). Our framework includes the choice \( \Phi = F \). Contrary to most of the existing methods, for instance SIR, MDA and SAVE, a slicing is no longer necessary. For both classes of methods described by [5], respectively CIR1 and CIR2, under mild assumptions we prove the asymptotic normality and obtain the validity of the bootstrap. An integral method similar to CIR1 has been introduced in [31] in the context \( p = o(n^{1/2}) \).
Their simulations highlight the accuracy of integral methods with respect to slicing methods in dimension reduction. Using the empirical process theory, our approach re-discover one of their asymptotic result and go further by showing the validity of the bootstrap.

Finally, our methodology allows to focus on different kind of Cramér-von Mises tests in the field of dimension reduction. First we test the dimension of the model, i.e. whether

\[ d_0 = d \text{ against } d_0 > d, \]

for some \( 1 \leq d \leq p \). Second inspired by [7], testing procedures are developed to assess the no effect of some user-selected sets of predictors, say \( \eta^T Z \) with \( \eta \in \mathbb{R}^{p \times (p-d)} \). More precisely, let \( \beta \in \mathbb{R}^{p \times d} \) be such that \((\beta, \eta)\) has full rank, we test if

\[ P(Y \in A | \beta^T Z) = P(Y \in A | Z). \]

We also handle the case where \( \eta \) is estimated by a given dimension reduction method. This typically leads us to evaluate whether a model is subject to the SIR pathology.

The limiting laws of the considered statistics are fairly hard to estimate so that we provide a valid Bootstrap procedure in order to compute the quantile. The choice of the bootstrap is crucial in testing since the bootstrap statistic needs to behave similarly as the statistic under \( H_0 \) even if \( H_1 \) is realized [11]. To implement the bootstrap, we follow ideas from [23] where the bootstrap is used in testing the rank of a matrix.

The paper is organized as follows. In section 2 we study the asymptotic behaviour of estimators of the processes \( m_{\Phi} \) and \( M_{\Phi} \). In this section, we also consider the case where \( \Phi \) is the c.d.f. of \( Y \) and the behaviour of the bootstrap estimator. Section 3 is dedicated to the use of the processes \( m_{\Phi} \) and \( M_{\Phi} \) in dimension reduction. More precisely, we revisit some theoretical results about slicing estimation (e.g. SIR, MDA, SAVE). Then, we introduce and study the integral methods CIR1 and CIR2 that estimate \( E_c \). Finally, we consider testing procedures about the structure of \( E_c \). Concluding comments are given in Section 4.

## 2 Preliminary results on empirical processes

### 2.1 Definitions

Using the outer integral, the author Hoffman-Jorgensen has defined a notion of weak convergence of random sequences [13]. This authorizes some elements of the considered sequences of being non-measurable with respect to the sigma-field induced by the uniform metric, provided that their limits are measurable. As a result, we equip the space of bounded real-valued functions with the supremum norm \( \| \cdot \|_\infty \). We denote it by \( l^\infty(\mathbb{R}) \). The outer integral of \( S \), a random element in \( l^\infty(\mathbb{R}) \), is defined by

\[ E^*[S] = \inf\{E[T] : T \geq S, T \text{ measurable and } E[T] \text{ is finite}\}. \]

We say that a sequence of random elements \( S_n \) in \( l^\infty(\mathbb{R}) \) converges weakly to a measurable element \( S \in l^\infty(\mathbb{R}) \) if

\[ E^*[f(S_n)] \longrightarrow E[f(S)], \]

for every \( f \) bounded continuous real function defined on \( l^\infty(\mathbb{R}) \), it is denoted by \( S_n \Rightarrow S \). Given an i.i.d. sequence \((Z_i, Y_i)_{1 \leq i \leq n}\) with law \( P \), we denote by \( P_n \) the associated empirical measure. We say that a class of measurable functions \( \mathcal{F} \subset l^\infty(\mathbb{R}) \) is \( P \)-Donsker if

\[ n^{1/2}(P_n - P)f \quad \text{ converges weakly in } l^\infty(\mathbb{R}), \]
where for any signed measure $Q$, $Qf = \int fdQ$. A complete study of the notion of weak convergence and Donsker classes is proposed in [27], and many links with statistical issues are provided in [16]. The following lemma will be useful in the next.

**Lemma 1.** Assume that $\int x^2dP$ is finite, then $\{(x,y) \mapsto x\mathbb{1}_{(-\infty,t]}(y), \ t \in \mathbb{R}\}$ and $\{(x,y) \mapsto x\mathbb{1}_{(-\infty,t]}(y), \ t \in \mathbb{R}\}$ are P-Donsker.

**Proof.** The proof is the same for both classes. Let us consider $G = \{(x,y) \mapsto x\mathbb{1}_{(-\infty,t]}(y), \ t \in \mathbb{R}\}$. First, it is classical that $F = \{1_{(-\infty,t]}, \ t \in \mathbb{R}\}$ is P-Donsker (see for instance [27], Example 2.5.4, page 129). In particular, the covering number of $F$ given in (6), and the fact that the covering number of a single element is 1, the uniform entropy condition is checked and the class $G$ is P-Donsker.

**2.2 Asymptotic behavior when $\Phi$ is known**

Let $(Z_i, Y_i)_{1 \leq i \leq n}$ be an i.i.d. sample drawn from model (1). The variables $Z_i$'s are assumed to be standardized in order to clarify the statements of the results. In practice we must account for the error induced by estimations of the mean and the variance. The empirical processes that estimate the processes $m_\Phi$ and $M_\Phi$ are defined as follows, for every $u \in \mathbb{R}$

$$\hat{m}_\Phi(u) = \frac{1}{n} \sum_{i=1}^{n} Z_i \mathbb{1}_{\{Y_i < \Phi^{-1}(u)\}} \quad \text{and} \quad \hat{M}_\Phi(u) = \frac{1}{n} \sum_{i=1}^{n} (Z_i Z_i^T - I) \mathbb{1}_{\{Y_i < \Phi^{-1}(u)\}}.$$

We introduce the matrix

$$\gamma_1(u,v) = \text{cov} \left(Z \mathbb{1}_{\{Y < \Phi^{-1}(u)\}}, Z \mathbb{1}_{\{Y < \Phi^{-1}(v)\}}\right).$$

**Theorem 2.** Assume that $\mathbb{E}[\|Z\|_2^2]$ is finite and $\Phi$ is a non-decreasing function, then $\sqrt{n}(\hat{m}_\Phi - m_\Phi)$ converges weakly in $l^\infty(\mathbb{R})$ to a tight Gaussian process with zero-mean and covariance function $\gamma_1$.

**Proof.** Each coordinate of the process $\sqrt{n}(\hat{m}_\Phi - m_\Phi)$ can be written as $\sqrt{n}(P_n - P)g$ where by Lemma 1 $g$ lies in a Donsker class. Because tightness is equivalent to tightness of each coordinates, it implies that the process $\sqrt{n}(\hat{m}_\Phi - m_\Phi)$ is tight. The limiting process is then given by the limiting distribution of the finite dimensional laws obtained by the central limit theorem. \qed
By a similar proof we obtain the weak convergence of $\sqrt{n}(\hat{M}_\Phi - M_\Phi)$. To state this we define the operator vec that vectorizes a matrix by stacking its column. We then introduce the matrix
\[
\Gamma_1(u, v) = \text{cov}(\text{vec}(ZZ^T - I) \mathbb{1}_{\{Y < \phi^-(u)\}}, \text{vec}(ZZ^T - I) \mathbb{1}_{\{Y < \phi^-(v)\}}).
\]

**Corollary 3.** Assume that $\mathbb{E}[[|Z_1|^4]]$ is finite and $\phi$ is a non-decreasing function, then $\sqrt{n}(\hat{M}_\Phi - M_\Phi)$ converges weakly in $l^\infty(\mathbb{R})$ to a tight Gaussian process with zero-mean and covariance function $\Gamma_1$.

### 2.3 Asymptotic behavior when $\Phi$ is the cdf of $Y$

We focus on the case where $\Phi = F$ the unknown distribution function of $Y$. Since
\[
m_F(u) = \mathbb{E}(Z \mathbb{1}_{\{F(Y) < u\}}) \quad \text{and} \quad M_F(u) = \mathbb{E}((ZZ^T - I) \mathbb{1}_{\{F(Y) < u\}}),
\]
this choice "uniformizes" the variable $Y$ and, as a consequence, vanishes the effect of the distribution of $Y$ on the estimation. Clearly, we can not follow the same path as previously since the estimation of $F$ will certainly affect the limiting process. We introduce the empirical process
\[
\hat{F}(t) = n^{-1} \sum_{i=1}^n \mathbb{1}_{\{Y_i \leq t\}}
\]
defined for each $t \in \mathbb{R}$. Our estimators are plugged-in, that is, $m_F$ and $M_F$ are respectively estimated by $\hat{m}_F$ and $\hat{M}_F$ given by
\[
\hat{m}_F(u) = \frac{1}{n} \sum_{i=1}^n Z_i \mathbb{1}_{\{Y_i < \hat{F}^-(u)\}} \quad \text{and} \quad \hat{M}_F(u) = \frac{1}{n} \sum_{i=1}^n (Z_i Z_i^T - I) \mathbb{1}_{\{Y_i < \hat{F}^-(u)\}}.
\]

**Remark 1.** An important point is the following: if $F$ is continuous, then without loss of generality, the variables $Y_i$ can be assumed to be uniformly distributed on $[0, 1]$. As a consequence the limiting function $F$ can be taken equal to the identity function on $[0, 1]$. To show this, first note that because $\hat{F}$ is a càdlàg function, it is easy to show that for any $(t, u)$,
\[
\{t < \hat{F}^-(u)\} \Leftrightarrow \{\hat{F}(t) < u\}. \quad (7)
\]
Then we have that
\[
\hat{m}_F(u) = \frac{1}{n} \sum_{i=1}^n Z_i \mathbb{1}_{\{\hat{F}(Y_i) < u\}}
\]
and a similar expression holds for $\hat{M}_F$. Then both previous estimators are empirical sums over the $Z_i$'s and the rank statistics $\hat{F}(Y_i)$'s. Second remark that the rank statistics based on the $Y_i$'s are the same as the rank statistics based on the $F(Y)$'s. As a consequence of this two facts, the processes $\hat{m}_F$ and $\hat{M}_F$ can be constructed identically from the sample $(F(Y_i), Z_i)_{1 \leq i \leq n}$. Now since $F(Y_i)$ is uniformly distributed on $[0, 1]$ (because of the continuity of $F$), we can put $F = \text{id}$.

To compute the asymptotic distribution, since $m_F = m_{id} \circ F^-$, we use the Delta method in metric spaces stated in Theorem 3.9.4 of [27]. This approach has been employed for instance in [27], page 389, and in [9], both in the context of the weak convergence of the empirical copula process. As a consequence, convergence results are obtained following this scheme:
1. Use Lemma \[1\] to obtain the weak convergence of the process \((s, t) \mapsto \sqrt{n}(\hat{F}(s), \hat{m}_{id}(t)) - (F(s), m_{id}(t))\)

2. Apply the Delta method to the map \((F, m_{id}) \mapsto m_{id} \circ F^-\)

Because the second step involves the quantile transformation that is not Hadamard differentiable everywhere, the fact that the limit of \(\hat{F}\) can be assumed to be the identity, by Remark \[1\] simplifies the proof. We define the function \(\gamma_2 : [0, 1]^2 \to \mathbb{R}^{(p+1) \times (p+1)}\) given by

\[
\gamma_2(u, v) = \text{cov}\left(\left(1 \over Z\right) 1_{\{F(Y) < u\}}, \left(1 \over Z\right) 1_{\{F(Y) < v\}}\right),
\]

and \(\gamma_3 : [0, 1]^2 \to \mathbb{R}^{p \times p}\) by

\[
\gamma_3(u, v) = (-\partial m_F(u), I) \gamma_2(u, v) (-\partial m_F(v), I)^T.
\]

**Theorem 4.** Assume that \(\mathbb{E}[\|Z_1\|^2]\) is finite and \(F\) is continuous. Then if \(m_{id}\) is continuously differentiable, \(\sqrt{n}(\hat{m}_F - m_F)\) converges weakly in \(l^\infty([0, 1])\) to a tight Gaussian process with zero-mean and covariance function \(\gamma_3\).

**Proof.** Without loss of generality (see Remark \[1\]), in the whole proof we put \(F = \text{id}\). By applying Lemma \[1\] the process \(\sqrt{n}(\hat{G} - G)\), with \(\hat{G}(s, t) = (\hat{F}(s), \hat{m}_{id}(t))\) and \(G(s, t) = (F(s), m_{id}(t))\), converges weakly in \(l^\infty(\mathbb{R}) \times l^\infty(\mathbb{R})\) to a tight Gaussian element. Now since

\[
\hat{m}_F = \psi(\hat{G}) \quad \text{and} \quad m_F = \psi(G),
\]

we can apply Theorem 3.9.4, page 374 in \[27\] which basically says that \(\sqrt{n}(\psi(\hat{G}) - \psi(G))\) is \(P\)-Donsker provided that the map \(\psi\) is Hadamard differentiable. In what follows, we first show that \(\psi\) is Hadamard differentiable, and then we compute the asymptotic variance. Using Lemma 3.9.23, assertion (ii) page 386 in \[27\], the first map of Equation \[13\] reduced to \(f \mapsto f^-\) is Hadamard differentiable at the function \(F\) tangentially to \(C[0, 1]\). Moreover its derivative at \(F = \text{id}\), in the direction \(h_1\) is given by \(-h_1\). Since \(m_{id}\) is Fréchet differentiable, by Lemma 3.9.27, page 388 in \[27\], the second map in Equation \[13\] is Hadamard differentiable at \((F^-, m_{id})\), tangentially to \(C[0, 1]\) (because continuous functions are uniformly continuous on compacts). Its derivative at \((F^-, m_{id})\), in the direction \((h_1, h_2)\), is given by \(h_1 \times \partial m_{id} \circ F^- + h_2 \circ F^-\). By the chain rule, the function \(\psi\) is Hadamard differentiable at \((F, m_{id})\) tangentially to \(C[0, 1]\). At the point \((F, m_{id})\), in the direction \((h_1, h_2)\), its derivative is given by \(-h_1 \times \partial m_{id} \circ F^- + h_2 \circ F^-\). The limiting process has the representation

\[
u \mapsto w_1(\nu - \partial m_F(\nu) \times B(\nu) = (-\partial m_F(\nu), I) \left(B(\nu) \over w_1(\nu)\right),
\]

where \((B, w_1)\) is the Gaussian limiting process of \(\hat{H} : u \mapsto \sqrt{n}(\hat{G} - G) \circ (u, u)\). Its covariance function is computed by applying the central limit theorem that gives

\[
(\hat{H}(u_1), \ldots, \hat{H}(u_K)) \xrightarrow{d} ((B, w_1)(u_1), \ldots, (B, w_1)(u_K)),
\]

where \(\text{vec}((B, w_1)(u_1), ..., (B, w_1)(u_K))\) is a Gaussian vector with mean 0 and covariance matrix having the block decomposition \((\gamma_2(u_k, u_l))_{1 \leq k, l \leq K}\).

\[\square\]
For the order 2 moments process define the function $\Gamma_2 : [0, 1]^2 \rightarrow \mathbb{R}^{(p+1) \times (p+1)}$ given by

$$
\Gamma_2(u, v) = \text{cov} \left( \left( \frac{1}{\text{vec}(ZZ^T - I)} \right)_{1\{F(Y) < u\}}, \left( \frac{1}{\text{vec}(ZZ^T - I)} \right)_{1\{F(Y) < v\}} \right),
$$

and $\Gamma_3 : [0, 1]^2 \rightarrow \mathbb{R}^{p \times p}$ by

$$
\Gamma_3(u, v) = (- \partial \text{vec}(M_F)(u), I) \Gamma_2(u, v) (- \partial \text{vec}(M_F)(v), I)^T.
$$

Replacing $\hat{m}_{\hat{F}}$, $m_F$, $\gamma_2$ and $\gamma_3$ by $\hat{M}_{\hat{F}}$, $M_F$, $\Gamma_2$ and $\Gamma_3$, respectively, and following exactly the same steps as in the proof of Theorem 4 we obtain the following corollary.

**Corollary 5.** Assume that $\mathbb{E}[\|Z_1\|^4]$ is finite and $F$ is continuous. Then if $\text{vec}(M_{id})$ is continuously differentiable, $\sqrt{n}(\hat{M}_F - M_F)$ converges weakly in $l^\infty([0, 1])$ to a tight Gaussian process with zero-mean and covariance function $\Gamma_3$.

### 2.4 The Bootstrap

Regarding the dependence of the limiting covariance processes in $\partial m_{id}$ and $\partial \text{vec}(M_{id})$ and also the possibly high-dimensionality of the Gaussian limits, the asymptotic variance is hard to estimate. As a consequence, for making inference, it seems necessary to develop a bootstrap strategy. Efron [8] introduce the original bootstrap that consists in a sampling with equi-probability and replacement of the original sample. In [24], the authors considered a more general resampling plan based on weights $w_{i,n}$, $i = 1, \ldots, n$ that verified

(B1) The sequence $(w_{i,n})$ is exchangeable, i.e. for every permutation $(\pi_1, \ldots, \pi_n)$ of $(1, \ldots, n)$, $(w_{i,n})$ has the same law as $(w_{\pi_i, n})$. Moreover $w_{i,n} \geq 0$ and $\sum_{i=1}^{n} w_{i,n} = n$.

(B2) Denote by $S$ the survival function of $w_{1,n}$, we have

$$
\int S(u)^{1/2} du < K \quad \text{and} \quad \lim_{A \to +\infty} \lim_{n \to +\infty} \sup_{t \geq A} t^2 S(t) = 0.
$$

(B3) $n^{-1} \sum_{i=1}^{n} (w_{i,n} - 1)^2 \xrightarrow{p} 1.$

For examples of such weights, we refer to [24]. Our bootstrap estimators are constructed as follows. Define, for every $t \in \mathbb{R}$, every $u \in \mathbb{R}$,

$$
\hat{F}^*(t) = n^{-1} \sum_{i=1}^{n} w_{i,n} \mathbbm{1}_{\{Y_i \leq t\}}
$$

$$
\hat{m}_F^*(u) = \frac{1}{n} \sum_{i=1}^{n} w_{i,n} Z_i \mathbbm{1}_{\{Y_i < \Phi^{-1}(u)\}},
$$

then the bootstrap of $\hat{m}_F$ (resp. $\hat{m}_{\hat{F}}$) is made by $\hat{m}_F^*$ (resp. $\hat{m}_{\hat{F}}^*$). The following theorem essentially says that the bootstrap of the considered stochastic processes works.

**Theorem 6.** Under (B1) to (B3), assume that $\mathbb{E}[\|Z_1\|^2]$ is finite and $\Phi$ is non-decreasing function, then conditionally on the sample,

$$
\frac{n^{1/2}(\hat{m}_F^* - m_F)}{\|m_F|} \text{ has the same weak limit as } \frac{n^{1/2}(\hat{m}_F - m_F)}{\|m_F|}.
$$

If moreover $F$ is continuous and $m_F$ is continuously differentiable, then conditionally on the sample,

$$
\frac{n^{1/2}(\hat{m}_{\hat{F}}^* - \hat{m}_{\hat{F}})}{\|m_F|} \text{ has the same weak limit as } \frac{n^{1/2}(\hat{m}_{\hat{F}} - m_F)}{\|m_F|}.
$$
Proof. The first statement is a direct consequence of Lemma 1 and Theorem 2.1 in [24]. For the second statement, we first apply the trick detailed in Remark 1 to the bootstrap estimator. Indeed it is easy to see that \( \hat{M}_{F*} \) can be constructed as well from the sample \((F(Y_i), Z_i)_{1 \leq i \leq n}\), so that the limit of \( \hat{F}^* \) can be assumed to be the identity function on \([0,1]\). By applying again Lemma 1 with Theorem 2.1 in Paestgrad and Wellner, the process \( \sqrt{n}(\hat{G}^* - \hat{G}) \), with \( \hat{G}^*(s,t) = (\hat{F}^*(s), \hat{m}_{id}(t)) \), has the same limiting distribution as \( \sqrt{n}(\hat{G} - G) \) (defined in the proof of Theorem 4), that converges weakly in \( l^\infty(R) \times l^\infty(R) \) to a tight Gaussian element. Then we can invoke the Delta-method for the bootstrap stated in Theorem 3.9.11, page 378 in [27].

Define for every \( u \in \mathbb{R} \),

\[
\hat{M}_\Phi(u) = \frac{1}{n} \sum_{i=1}^{n} w_{i,n}(Z_iZ_i^T - I) \mathbb{1}_{\{Y_i < \Phi^{-1}(u)\}},
\]

we obtain this corollary, whose proof is the same as the proof of Theorem 4.

**Corollary 7.** Under (H1) to (H3), assume that \( \mathbb{E}[\|Z_1\|^4] \) is finite and \( \Phi \) is non-decreasing function, then conditionally on the sample,

\[
\sqrt{n}(\hat{M}_\Phi - \hat{M}_\Phi) \text{ has the same weak limit as } \sqrt{n}(\hat{M}_\Phi - M_\Phi).
\]

If moreover \( F \) is continuous and vec\((M_{id})\) is continuously differentiable, then conditionally on the sample,

\[
\sqrt{n}(\hat{M}_{F*} - \hat{M}_F) \text{ has the same weak limit as } \sqrt{n}(\hat{M}_{F} - M_F).
\]

### 3 Dimension reduction with continuous inverse regression

In this section we basically use our results about the behaviour of the process \( m_\Phi \) and \( M_\Phi \) in order to first raise a new point about SIR and second to develop CIR (estimation methods as well as testing procedure).

#### 3.1 Revisiting sliced inverse regression

In many studies dealing with sliced inverse regression, the number of slices \( H \) remains fixed and the number of observations falling in each slice is equi-distributed: \( k_n = \lfloor n/H \rfloor \) for the first \( H - 1 \) slices, and \( n - (H - 1)\lfloor n/H \rfloor \) for the last slice (see Remark 4.2 in [18]). Many authors considered the partition of \( Y \) as fix ([3] or [22]) while in such a case, the partition does depend on the sample and it changes with \( n \). In fact such an arrangement of the data induces an additional source of randomness. This is related to the rank transformation \( \hat{F}(Y_i)'s \) that order the sample according to the \( Y_i' \)s. For this reason the work done in the previous section offers interesting conclusions for this context. In the next, we focus on the first order moment methods, leading to SIR, but the same analysis can be extended to second order methods, leading to SAVE.

Following the methodology proposed in [3] (that includes the SIR estimator and many of its variants), the asymptotics of the estimation of \( E_c \) rely on the asymptotics of the matrix

\[
\hat{M} = (\hat{C}_1, \ldots, \hat{C}_H)
\]

\[
\hat{C}_h = \hat{m}_{\hat{F}}(i_n(h)) - \hat{m}_{\hat{F}}(i_n(h - 1)), \quad (10)
\]

where \( i_n(h) = h\lfloor n/H \rfloor /n \), for \( h = 0, \ldots, H - 1 \), and \( i_n(H) = 1 \). Putting \( i(h) = h/H \), we define

\[
M = (C_1, \ldots, C_H)
\]

\[
C_h = m_F(i(h)) - m_F(i(h - 1)). \quad (11)
\]
Clearly, the asymptotic behaviour of $\tilde{M}$ can be derived through the asymptotic behaviour of the vector 

$$(\tilde{m}_F(i_n(0)), \ldots, \tilde{m}_F(i_n(H))),$$

which can indeed be described using Theorem 4. We have the decomposition

$$\tilde{m}_F(i_n(h)) - m_F(i(h)) = ((\tilde{m}_F(i_n(h)) - m_F(i_n(h))) - (\tilde{m}_F(i(h)) - m_F(i(h))) + m_F(i_n(h)) - m_F(i(h)) + \tilde{m}_F(i(h)) - m_F(i(h)).$$

By the asymptotic tightness of the process $\sqrt{n}(\tilde{m}_F - m_F)$, since $|i_n(h) - i(h)| \leq 1/n$ we have that the first term equals $o_P(n^{-1/2})$. For the second term, since $m_F$ is continuously differentiable, we have that

$$|m_F(i_n(h)) - m_F(i(h))| \leq \text{const.}/n,$$

which suffices to show that it is also asymptotically neglectable. Since this is true for each $h$, we have shown that

$$n^{1/2}((\tilde{m}_F(i_n(0)), \tilde{m}_F(i_n(1)), \ldots, \tilde{m}_F(i_n(H))) - (m_F(i(0)), m_F(i(1)), \ldots, m_F(i(H))))$$

has the same asymptotic distribution than

$$n^{1/2}((\tilde{m}_F(i(0)), \tilde{m}_F(i(1)), \ldots, \tilde{m}_F(i(H))) - (m_F(i(0)), m_F(i(1)), \ldots, m_F(i(H))))),$$

which converges to a Gaussian distribution, by Theorem 4. Hence because of (10) and (11), we have shown that $\sqrt{n}(\tilde{M} - M)$ converges to a certain Gaussian random vector. The derivation of its asymptotic law represents heavy computations that we believe not necessary regarding the validity of the bootstrap procedure defined in Section 2.4.

**Remark 2.** First, it must be stressed that the asymptotic law of $\tilde{M}$ we found is very different than the usual asymptotics considered in the literature, either with fix slices ([5], [22]) or with a number of slice that goes to infinity ([15], [30], [20]). This is evident when the length of the slices goes to 0 because it is another quantity that is estimated. When the slices are fixed, the asymptotic covariance between two columns of $\tilde{M}$ is null. This is no longer true when the $Y_i$’s are ordered. Second, while existing bootstrap methods for SIR ([1], [22]) consider the partition as fix, their resampling plan is not as general as the one that is defined in Section 2.4.

### 3.2 Integral methods

#### 3.2.1 Order 1 continuous inverse regression

We define the class CIR1 that characterizes $E_c$ through the space generated by the matrix

$$A_\Phi = \int m_{id}(t)m_{id}(t)^T d\Phi(t),$$

where $\Phi$ is a given probability measure. Since the variable $\Phi^{-1}(U)$ has distribution $\Phi$, whenever $U$ is uniformly distributed on $[0, 1]$ (see for instance [26, page 305]), we have the formula

$$A_\Phi = \int m_\Phi(u)m_\Phi(u)^T du,$$
Our estimator of $A_\Phi$ is given by

$$\hat{A}_\Phi = \int_0^1 \hat{m}_\Phi(u)\hat{m}_\Phi(u)^T du,$$

it is easy to compute since a quick calculation gives

$$\hat{A}_\Phi = n^{-2} \sum_{i,j} Z_i Z_j^T (1 - \max(\Phi(Y_i), \Phi(Y_j))).$$

(12)

An interesting member of CIR1 is the method corresponding to $\Phi = F$ whose estimator is given replacing $\Phi$ by $\hat{F}$ in the above formula. This simply gives an estimator based on the ranks.

To obtain the weak convergence of $\sqrt{n}(\hat{A}_\Phi - A_\Phi)$, whether $\Phi$ is estimated or not, our approach consists in the two following steps:

(A) Weak convergence of the process $\sqrt{n}(\hat{m}_\Phi - m_\Phi)$ in $l^\infty([0,1])$.

(B) Continuous mapping theorem apply to some integral map.

In the previous sections, we have already obtained the first step, so that the proof of the following theorem consists in the second step.

**Theorem 8.** Assume that $E[\|Z_1\|^2]$ is finite and $\Phi$ is a distribution function, then

$$n^{1/2}(\hat{A}_\Phi - A_\Phi) \text{ has Gaussian limit } \int_0^1 w_1(u)m_\Phi(u)^T + m_\Phi(u)w_1(u)^T du,$$

where $w_1$ is the Gaussian process with covariance $\gamma_1$. If moreover $F$ is continuous and $m_{id}$ is continuously differentiable, then

$$n^{1/2}(\hat{A}_F - A_F) \text{ has Gaussian limit } \int_0^1 w_3(u)m_F(u)^T + m_F(u)w_3(u)^T du,$$

where $w_3$ is the Gaussian process with covariance $\gamma_3$.

**Proof.** We make the proof for the first assertion, the second one can be treated similarly. We have

$$(\hat{m}_\Phi\hat{m}_\Phi^T - m_\Phi m_\Phi^T) = (\hat{m}_\Phi - m_\Phi)m_\Phi^T + m_\Phi(\hat{m}_\Phi - m_\Phi)^T + (\hat{m}_\Phi - m_\Phi)(\hat{m}_\Phi - m_\Phi)^T.$$ 

Then invoking Theorem [23] by the weak convergence of $\sqrt{n}(\hat{m}_\Phi - m_\Phi)$ to the Gaussian process $w_1$ and the Delta-method, $\sqrt{n}(\hat{m}_\Phi\hat{m}_\Phi^T - m_\Phi m_\Phi^T)$ converges weakly to $w_1m_{id}^T + m_{id}w_1^T$. By the continuous mapping theorem stated for instance in [27], Theorem 1.3.6, page 20, we obtain that

$$n^{1/2}(\hat{A}_\Phi - A_\Phi) \overset{d}{\to} \int_0^1 w_1(u)m_\Phi(u)^T + m_\Phi(u)w_1(u)^T du.$$

For the following statement about the bootstrap, we define the matrix

$$\hat{A}_\Phi^* = \int_0^1 \hat{m}_\Phi^*(u)\hat{m}_\Phi^*(u)^T du.$$
Theorem 9. Under (H1) to (H3), assume that $E[|Z_1|^2]$ is finite and $\Phi$ is a distribution function, then conditionally on the sample,

$$n^{1/2}(\hat{A}_\Phi^* - \hat{A}_\Phi) \text{ has the same weak limit as } n^{1/2}(\hat{A}_\Phi - A_\Phi).$$

If moreover $F$ is continuous and $m_{id}$ is continuously differentiable, then conditionally on the sample,

$$n^{1/2}(\hat{A}_F^* - \hat{A}_F) \text{ has the same weak limit as } n^{1/2}(\hat{A}_F - A_F).$$

Proof. The proof is the same as the proof of Theorem S in the probability space conditional to the $(Y_i, Z_i)$’s.

3.2.2 Order 2 continuous inverse regression

Following the previous section we introduce the class CIR2 based on the space generated by the matrix

$$B_\Phi = \int M_{id}(t)^2 d\Phi(t),$$

where $\Phi$ is a given probability measure. Our estimator is given by

$$\hat{B}_\Phi = \int \hat{M}_\Phi(u)^2 d(u),$$

and we have the formula

$$\hat{B}_\Phi = n^{-2} \sum_{i,j} b_{ij} (1 - \max(\Phi(Y_i), \Phi(Y_j))),$$

with $b_{ij} = Z_i^T Z_j Z_i^T - Z_i Z_j^T + I$.

Corollary 10. Assume that $E[|Z_1|^4]$ is finite and $\Phi$ is a distribution function, then

$$n^{1/2}(\hat{B}_\Phi - B_\Phi) \text{ has Gaussian limit } \int_0^1 W_1(u) M_\Phi(u) + M_\Phi(u) W_1(u) du,$$

where $W_1$ is the Gaussian process with covariance $\Gamma_1$. If moreover, $F$ is continuous and vec$(M_{id})$ is continuously differentiable, then

$$n^{1/2}(\hat{B}_F - B_F) \text{ has Gaussian limit } \int_0^1 W_3(u) M_F(u) + M_F(u) W_3(u) du,$$

where $W_3$ is the Gaussian process with covariance $\Gamma_3$.

For the bootstrap, define

$$
\hat{B}_\Phi^* = \int_0^1 \hat{M}_\Phi(u)^2 du.
$$

Corollary 11. Under (H1) to (H3), assume that $E[|Z_1|^4]$ is finite and $\Phi$ is a distribution function, then conditionally on the sample,

$$n^{1/2}(\hat{B}_\Phi^* - \hat{B}_\Phi) \text{ has the same weak limit as } n^{1/2}(\hat{B}_\Phi^* - B_\Phi).$$

If moreover $F$ is continuous and $m_{id}$ is continuously differentiable, then conditionally on the sample,

$$n^{1/2}(\hat{B}_F^* - \hat{B}_F) \text{ has the same weak limit as } n^{1/2}(\hat{B}_F^* - B_F).$$
3.3 Characterization of the central subspace

By comparing the spaces generated by the methods CIR1 and SIR, we consider the difference between continuous methods and slicing methods. Our arguments apply also to the comparison between CIR2 and SAVE.

The space estimated by SIR is

$$E_{\text{SIR}}^H = \Sigma^{-1} \text{span}(C_1, \ldots, C_H),$$

where

$$C_h = E[Z_1 1_{\{Y_1 \in I(h)\}}],$$

and the $I(h)$’s form a partition of the range of the variable $Y$. Theorem 3 in [22] implies that, for $H$ sufficiently large $E_{\text{SIR}}^H = E_c$. This result is important because it ensures that when $H$ increases, SIR eventually estimates the whole subspace. Nevertheless, this is not sufficient to guarantee a complete estimation of $E_c$ since in practice, we do not know how to choose $H$. The space estimated by CIR1 is

$$E_{\text{CIR1}} = \Sigma^{-1} \text{span}(m_{id}(t), t \in \mathbb{R}).$$

Since for any $h$, we have that $\mu_h = m_{id}(t_h) - m_{id}(t_{h-1})$ for some $t_h, t_{h-1}$, it follows that $E_{\text{SIR}}^H \subset E_{\text{CIR1}}$, with equality when $H$ is sufficiently large. As a consequence, the main advantage of using CIR1 rather than SIR is that as $n$ increase, the whole space $E_c = E_{\text{CIR1}} = E_{\text{SIR}}^\infty$ is reached, without the need of choosing the smoothing parameter $H$.

3.4 Cramér-von Mises tests

Because the integral methods introduced in Section 3.2 produce an exhaustive estimation of $E_c$ (see Section 3.3), it is suitable to use the underlying matrices $A_\Phi$ and $B_\Phi$ in order to test certain structural properties of $E_c$. In the following, we introduce three tests for dimension reduction: we test the dimension of $E_c$, the no effect of a set of predictors and the contribution of a given method. For clarity, each test is introduced with respect to only one methods: CIR1 or CIR2 (with either $\Phi$ or $F$). The adaptation to the other situation can be made easily by the reader. At the end of the section, under a unified approach, we show that all the tests are consistent and that the bootstrap is valid to compute their quantiles.

All the test statistics that we introduce are of Cramér-von Mises type, i.e. of the form

$$\int_0^1 \|\hat{f}(u)\|^2 du,$$

where $\hat{f}$ is a certain process that belongs to $l^\infty([0,1])$. In our precise situation, they lead to closed-form statistics, this does not happen for Kolmogorov type statistics.

3.4.1 Testing dimensionality

In order to determine the dimension $d_0$ of $E_c$, it is usual to test whether $d_0$ equals a given number, say $d$. More precisely we test

$$H_0 : d_0 = d \quad \text{against} \quad H_1 : d_0 > d,$$

for $d = 0$, if rejected we put $d := d + 1$, until the first acceptance. The situation is summarized in [2] and [23]. In the following we focus on the most common test statistic, based on the sum of the eigenvalues of $\hat{A}_\Phi$ or $\hat{B}_\Phi$. Define the statistic

$$\hat{\Lambda}_1 = n \sum_{k=d+1}^p \hat{\lambda}_k,$$
where the $\hat{\lambda}_k$'s are the eigenvalues of the matrix $\hat{A}_\Phi$ or $\hat{B}_\Phi$, arranged in decreasing order. For instance for $\hat{A}_\Phi$, we have the formula

$$\hat{\Lambda}_1 = n \text{tr}(\hat{Q}\hat{A}_\Phi\hat{Q}) = \int_0^1 \|n^{1/2}\hat{Q}\hat{m}_\Phi(u)\|^2 du,$$

where $\hat{Q}$ equals the eigenprojector on the eigenspace associated to the $p - d$ smallest eigenvalues of $\hat{A}_\Phi$.

### 3.4.2 Testing predictor contribution

Following [7], we develop tests of no effect, on the variable to explain $Y$, of a selected group of predictor, say $\eta^T Z$ where $\eta \in \mathbb{R}^{p \times (p - d)}$ is such that $\eta^T \eta = I$. We define $\beta$ as the counterpart of $\eta$ in $\mathbb{R}^p$, i.e. $(\beta, \eta) \in \mathbb{R}^{p \times p}$ is an orthogonal matrix. We say that $\eta^T Z$ has no effect on $Y$ if

$$P(Y \in A | \beta^T Z) = P(Y \in A | Z),$$

for any Borel set $A$. By [7], Proposition 1, this is equivalent to $\eta \in E_c^\perp$. As a consequence, we introduce the hypotheses

$$H_0 : \eta \in E_c^\perp \quad \text{against} \quad H_1 : \eta \notin E_c^\perp,$$

(16)

In the following we focus only on $A_F$ for the sake of brevity but the same approach works replacing $A$ by $B$ and $F$ by $\Phi$. Under some coverage condition, that basically says that $E_c$ is spanned by the inverse regression curve, the previous hypotheses are equivalent to

$$H_0 : \eta^T A_F \eta = 0 \quad \text{against} \quad H_1 : \eta^T A_F \eta \neq 0.$$

Then a good statistic for testing $H_0$ is

$$\hat{\Lambda}_2 = n \text{tr}(\hat{\eta}^T \hat{A}_F \hat{\eta}) = n \int_0^1 \|\hat{\eta}^T \hat{m}_F(u)\|_F^2 du,$$

where $\| \cdot \|_F$ stands for the Frobenius norm.

### 3.4.3 Testing method contribution

Consider a given method whose estimated basis is noted $\hat{\beta} \in \mathbb{R}^{p \times d}$ and let us assume that $\beta \in \mathbb{R}^{p \times d}$ is such that $\sqrt{n}(\hat{\beta} - \beta)$ converges to a Gaussian vector. We want to test whether the method misses a direction, i.e.

$$H_0 : \eta \in E_c^\perp \quad \text{against} \quad H_1 : \eta \notin E_c^\perp,$$

(17)

where $(\beta, \eta) \in \mathbb{R}^{p \times p}$ is an orthogonal matrix. Based on CIR2, let $\hat{\eta}$ be such that $(\hat{\beta}, \hat{\eta})$ is orthogonal, we define the statistic

$$\hat{\Lambda}_3 = \text{tr}(\hat{\eta}^T \hat{B}_\Phi \hat{\eta}) = n \int_0^1 \|\hat{\eta}^T \hat{M}_\Phi(u)\|_F^2 du$$

where the $b_{ij}$'s are defined in Section 3.2.2. We have in mind two typical applications. First we aim at testing the so called SIR pathology or more precisely whether the order 1 moment based methods fail in recovering the whole subspace (see [22] for more details about this pathology). For that purpose, $\beta$ might be for instance the estimated basis by the SIR method. Clearly if the model is subject to the order 1 pathology, the test shall reject $H_0$. Second the later procedure can be applied to select the estimated directions for the method called the order 2 optimal function method developed in [22]. This method alleviates the assumption CCV and produces accurate estimates but classical eigenvalue-based selection of the direction fails. The initial test of independence developed in [22] was too strong. It seems better to apply the above test with $\hat{\beta}$ being the estimated basis of the order 2 optimal function method.
3.4.4 Consistency of the tests

Theoretically, a test is said to be consistent if, as \( n \) increase, the level converges to the nominal level and the power goes to 1. As it will stressed out, every of the tests considered previously is consistent. Practically one needs to compute the quantiles of the asymptotic law of the statistic. In our case, those quantiles are difficult to estimate and this could diminish the accuracy of the test [23]. As a consequence, we recommend to use a bootstrap strategy for computing these quantiles.

In the next, we provide the consistency and the bootstrap validity for the three testing procedures introduced in the previous section. In fact, all the tests (15), (16) and (17), with either CIR1 or CIR2, can be described under a unified framework. Denoting by \( \hat{\mu} \) the statistical process of interest. It might be equal to \( \hat{m}_\Phi, \hat{m}_{\hat{F}} \) or \( \hat{M}_\Phi, \hat{M}_{\hat{F}} \) depending which method, CIR1 or CIR2 is considered. The statistics \( \hat{\Lambda}_k \) for \( k = 1, 2, 3 \), can be described as follows

\[
\hat{\Lambda}_k = n \int_0^1 \| \hat{Q}_k \hat{\mu}(u) \|^2 \hat{F} \, du,
\]

with \( \hat{Q}_1 \) the eigenprojector associated to the \( p - d \) smallest eigenvalues of \( \int_0^1 \hat{\mu}(u)\hat{\mu}(u)^T \hat{F} \, du \), \( \hat{Q}_2 = \eta \eta^T \) with \( \eta \in \mathbb{R}^{p \times (p-d)} \) a basis, and \( \hat{Q}_3 \) is the orthogonal projector on the orthogonal complement of the estimated space of a given method. To guarantee the consistency of the tests, we introduce the following assumptions. A discussion is postponed latter.

(A1) Let \( \hat{\mu} : [0, 1] \to \mathbb{R}^{p \times q} \) and \( \mu : [0, 1] \to \mathbb{R}^{p \times q} \) with \( q \geq 1 \), be processes such that

\[
n^{1/2}(\hat{\mu} - \mu) \text{ converges in } l^\infty([0, 1]) \text{ to a Gaussian vector.}
\]

(A2) Assume that \( \text{span}(\mu(u), u \in [0, 1]) = E_c \).

(A3) Let \( \hat{\beta} \) be the estimated basis of a given method. Let \( \hat{P}_3 = I - \hat{Q}_3 = \hat{\beta} \hat{\beta}^T \) be such that

\[
\hat{P}_3 - P_3 = \mathbb{P}_n \varphi + o_p(n^{-1/2}),
\]

for some \( P_3 \) being an orthogonal projector.

(A4) Let \( \hat{\mu}^* : [0, 1] \to \mathbb{R}^{p \times q} \) be a process and \( \hat{P}_3^* \) be an orthogonal projector such that, conditionally to the sample,

\[
n^{1/2} \left( \frac{\hat{\mu}^* - \hat{\mu}}{\hat{P}_3^* - \hat{P}_3} \right) \text{ has the same asymptotic law as } n^{1/2} \left( \frac{\hat{\mu} - \mu}{\hat{P}_3 - P_3} \right).
\]

The previous set of assumptions has to be understood as follows. Assumption (A1) is precisely the statement of Theorems 2 and 4 for CIR1, and Corollaries 3 and 5 for CIR2. The reader might refer to the mentioned theorems to get conditions that ensure (A1). Assumption (A2) is a coverage condition that has been used by several authors [5], [22]. This coverage condition is legitimate by Section 3.3, where the exhaustive property of the CIR framework is demonstrated. Assumption (A3) deals with the dimension reduction method that is the object of the testing procedure (17). This could be any method of the literature, as for instance SIR, IRE and SAVE, for which such a result is available. When testing (15) or (16), the assumptions about \( \hat{P} \) are no longer needed. Assumption (A4) is needed for bootstrap inference.

We have the following proposition.

**Proposition 12.** Assume (A1) and (A2), then testing (15) and (16) with respectively \( \hat{\Lambda}_1 \) and \( \hat{\Lambda}_2 \) is consistent. If moreover (A3) holds, then testing (17) with \( \hat{\Lambda}_3 \) is consistent.
Proof. The statistics $\hat{\Lambda}_k$’s have the same form up to the choice of $\hat{Q}_k$. The quantity $\hat{Q}_2$ is fix, $\hat{Q}_3$ is random but its behaviour is described by (A3), $\hat{Q}_1$ is random and depends on the process $\hat{\mu}$. As a consequence, the most difficult case is testing (15) with $\hat{\Lambda}_1$. In the proof we focus on this case, the other cases can be derived following the same steps.

We remark that $\hat{\Lambda}_1$ is a continuous transformation of the process $n^{1/2}\hat{Q}_1\hat{\mu}$ that equals, under $H_0$,

$$n^{1/2}Q_1(\hat{\mu} - \mu) + n^{1/2}(\hat{Q}_1 - Q_1)\mu + n^{1/2}(\hat{Q}_1 - Q_1)(\hat{\mu} - \mu). \tag{18}$$

By the weak convergence of $n^{1/2}(\hat{\mu} - \mu)$ (A11), we follow the proof of Theorem 8 to get the weak convergence of $n^{1/2}\int_0^1 \hat{\mu}(u)\hat{\mu}(u)^T - \mu(u)\mu(u)^T du$,

then by [25], under $H_0$, $n^{1/2}(\hat{Q}_1 - Q_1)$ converges to a Gaussian distribution. As a consequence, under $H_0$, the last term of (A3) vanishes asymptotically. Then using (A3), $\hat{\Lambda}_1$ converges in distribution. Under $H_1$, by (A2) some of the eigenvalues of the matrix $A_0$ goes to a positive number so that $\hat{\Lambda}_1$ goes to infinity.

Bootstrap testing requires particular attention to make sure that the bootstrap estimator mimics the hypothesis $H_0$ even when $H_1$ is realized ([11], [23]). Concerning $\hat{\Lambda}_1$, [23] shows that the quantiles can be computed using the technique of the constraining bootstrap. Following their approach, we define the bootstrap statistics $\hat{\Lambda}_k$’s by the formula

$$\hat{\Lambda}_k = n \int_0^1 \|\hat{\mu}_k\|_2^2 du \text{ for } k = 1, 2, 3,$$

with $\hat{Q}_1^*$ the eigenprojector associated to the $p - d$ smallest eigenvalue of $\int_0^1 \hat{\mu}_1(u)\hat{\mu}_1(u)^T du$, $\hat{Q}_2^* = I - \beta \beta^T$, $\hat{Q}_3^* = I - \hat{P}_3^*$ and for every $u \in [0, 1]$,

$$\hat{\mu}_k^*(u) = (I - \hat{Q}_k)\hat{\mu}(u) + (\hat{\mu}^*(u) - \hat{\mu}(u)).$$

The later formula is the cornerstone of the bootstrap procedure. It ensures that the bootstrap process $\hat{\mu}_k^*$ lies close to a subspace of dimension $d$. As a result the bootstrap process has a $H_0$-likelihood behaviour that ensures a high power for the tests.

**Proposition 13.** Assume (A7), (A2) and (A3) then testing (15) and (16) with respectively $\hat{\Lambda}_1$, $\hat{\Lambda}_2$ and calculation of the quantiles with $\hat{\Lambda}_1^*$, $\hat{\Lambda}_2^*$ respectively, is consistent. If moreover (A3) holds, then testing (17) with $\hat{\Lambda}_3$ and calculation of the quantile with $\hat{\Lambda}_3^*$, is consistent.

**Proof.** First we focus on $\hat{\Lambda}_1$ which represents the most difficult situation. We rely on Proposition 2 in [23]. Since one has

$$\hat{\Lambda}_1 = n \sum_{k=d+1}^p \hat{\lambda}_k \quad \text{and} \quad \hat{\Lambda}_1^* = n \sum_{k=d+1}^p \hat{\lambda}_k^*,$$

where the $\hat{\lambda}_k$’s (resp. the $\hat{\lambda}_k^*$’s) are the $p - d$ smallest eigenvalue of the matrix $\int_0^1 \hat{\mu}(u)\hat{\mu}(u)^T du$ (resp. $\int_0^1 \hat{\mu}_1(u)\hat{\mu}_1(u)^T du$), we only need to check Equation (13) of [23]. This is done by writing

$$\int_0^1 \hat{\mu}_1^*(u)\hat{\mu}_1^*(u)^T du = \hat{P} \int_0^1 \hat{\mu}(u)\hat{\mu}(u)^T du \hat{P} + \int_0^1 (\hat{\mu}^*(u) - \hat{\mu}(u))\hat{\mu}(u)^T du \hat{P} + \hat{P} \int_0^1 \hat{\mu}(u)(\hat{\mu}^*(u) - \hat{\mu}(u))^T du + \int_0^1 (\hat{\mu}^*(u) - \hat{\mu}(u))(\hat{\mu}^*(u) - \hat{\mu}(u))^T du,$$
noticing that the last term is $o_p(n^{-1/2})$ and the middle term has the same asymptotic law than
\[ \int_0^1 \mu(u)\hat{\mu}(u)^2 du - \int_0^1 \mu(u)(\mu(u))^2 du \text{ under } H_0. \]

The proofs for the tests associated to $\hat{A}_2$ and $\hat{A}_3$ are similar. We only give the proof for $\hat{A}_3$. First we show that the asymptotic law of $\hat{A}_3$ and $\hat{A}_3^*$ are the same under $H_0$, then we show that $\hat{A}_3^*$ is tight under $H_1$. Clearly, $\hat{A}_3$ (resp. $\hat{A}_3^*$) is a continuous transformation of the process $n^{1/2}\hat{Q}_3\hat{\mu}$ (resp. $n^{1/2}\hat{Q}_3\hat{\mu}^*$). Under $H_0$ since for each $u \in [0,1]$, $Q_3 \perp \mu(u)$, we have
\[
n^{1/2}\hat{Q}_3\hat{\mu}(u) = n^{1/2}Q_3(\hat{\mu}(u) - \mu(u)) + n^{1/2}(\hat{Q}_3 - Q_3)\mu(u) + n^{1/2}(\hat{Q}_3 - Q_3)(\hat{\mu}(u) - \mu(u)),
\]
and
\[
n^{1/2}\hat{Q}_3\hat{\mu}^*(u) = n^{1/2}\hat{Q}_3\hat{\mu}_3^*(u) + n^{1/2}(\hat{Q}_3^* - \hat{Q}_3)\hat{\mu}_3^*(u)
= n^{1/2}\hat{Q}_3(\hat{\mu}^*(u) - \hat{\mu}(u)) + n^{1/2}(\hat{Q}_3^* - \hat{Q}_3)(I - \hat{Q}_k)\hat{\mu}(u)
+ n^{1/2}(\hat{Q}_3^* - \hat{Q}_3)(\hat{\mu}^*(u) - \hat{\mu}(u)).
\]
In both equations, the latter term vanishes asymptotically. By Assumption (A[3]) both quantities have the same limiting law. Under $H_1$, we remark that for every $u \in [0,1]$,
\[
\|n^{1/2}\hat{Q}_3\hat{\mu}_3^*(u)\|_F \leq \|n^{1/2}\hat{Q}_3(\hat{\mu}^*(u) - \hat{\mu}(u))\|_F + \|n^{1/2}\hat{Q}_3\hat{P}_{\hat{\mu}_3}(u)\|_F
\leq \|n^{1/2}Q^*(u - \mu(u))\|_F + \|n^{1/2}(\hat{Q}_3^* - \hat{Q}_3)\hat{P}_{\hat{\mu}_3}(u)\|_F.
\]
The first term converges weakly. The second term is tight. \square

4 Conclusion

We have provided a new approach for inverse regression based on empirical processes. The empirical process theory led us to theoretical result that where unknown in this field. For instance, we described the asymptotic distribution of the SIR estimator when the number of slices remains fix but the number of observation within each slice is equi-distributed. We also provided the asymptotic normality of the estimation methods CIR1 and CIR2, and the consistency of some Cram\'er-von Mises testing procedures. In both case, the bootstrap is shown to be a valid tool for making inference.

The framework we develop in the paper is linked with the class of indicator functions. This choice was convenient since the metric entropy properties of this class are widely known, but also because of the natural link it induced with the famous method SIR. Nevertheless, the approach developed in this paper can be extended to other classes of functions than indicators. Indeed, for the order 1 moment based method, one can consider the vectors
\[
\mathbb{E}[X\psi(Y)],
\]
with $\psi_t$ a family that separates the points (e.g. Fourrier, wavelet, kernel,...).

Another subject for further study might be the study of right-censored data with CIR. Suppose we observe
\[
\min(Y, C) \text{ and } 1_{\{Y \leq C\}}, \text{ where } Y \perp C|X.
\]
In such a censored data context, variations of SIR have been studied for instance in [19] and [21]. It requires a smoothing procedure in order to take into account the effect of the censure. So far a slicing of the variable $Y$ is still needed. Even if the smoothing looks necessary, the slicing might be overcome using an empirical process approach similar to CIR.

17
Acknowledgement. The author would like to thank Bernard Delyon for helpful comments and advices on this article.

References

[1] M. P. Barrios and S. Velilla. A bootstrap method for assessing the dimension of a general regression problem. Statist. Probab. Lett., 77(3):247–255, 2007.

[2] E. Bura and J. Yang. Dimension estimation in sufficient dimension reduction: a unifying approach. J. Multivariate Anal., 102(1):130–142, 2011.

[3] R. D. Cook. Regression graphics. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, 1998.

[4] R. D. Cook and B. Li. Dimension reduction for conditional mean in regression. Ann. Statist., 30(2):455–474, 2002.

[5] R. D. Cook and L. Ni. Sufficient dimension reduction via inverse regression: a minimum discrepancy approach. J. Amer. Statist. Assoc., 100(470):410–428, 2005.

[6] R. D. Cook and S. Weisberg. Discussion of “sliced inverse regression for dimension reduction”. J. Amer. Statist. Assoc., pages 28–33, 1991.

[7] R. Dennis Cook. Testing predictor contributions in sufficient dimension reduction. Ann. Statist., 32(3):1062–1092, 2004.

[8] B. Efron. Bootstrap methods: another look at the jackknife. Ann. Statist., 7(1):1–26, 1979.

[9] J. Fermanian, D. Radulovic, and M. Wegkamp. Weak convergence of empirical copula processes. Bernoulli, 10(5):847–860, 2004.

[10] P. Hall and K. Li. On almost linearity of low-dimensional projections from high-dimensional data. Ann. Statist., 21(2):867–889, 1993.

[11] P. Hall and S. R. Wilson. Two guidelines for bootstrap hypothesis testing. Biometrics, 47(2):757–762, 1991.

[12] W. Härdle and T. M. Stoker. Investigating smooth multiple regression by the method of average derivatives. J. Amer. Statist. Assoc., 84(408):986–995, 1989.

[13] J. Hoffmann-Jorgensen. Stochastic processes on Polish spaces, volume 39 of Various Publications Series (Aarhus). Aarhus Universitet, Matematisk Institut, Aarhus, 1991.

[14] M. Hristache, A. Juditsky, and V. Spokoiny. Direct estimation of the index coefficient in a single-index model. Ann. Statist., 29(3):595–623, 2001.

[15] T. Hsing and R. J. Carroll. An asymptotic theory for sliced inverse regression. The Annals of Statistics, 20(2):1040–1061, 1992.

[16] M. R. Kosorok. Introduction to empirical processes and semiparametric inference. Springer Series in Statistics. Springer, New York, 2008.

[17] B. Li and S. Wang. On directional regression for dimension reduction. J. Amer. Statist. Assoc., 102(479):997–1008, 2007.
[18] K. Li. Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.*, 86(414):316–342, 1991.

[19] K. Li, J. Wang, and C. Chen. Dimension reduction for censored regression data. *Ann. Statist.*, 27(1):1–23, 1999.

[20] Y. Li and L. Zhu. Asymptotics for sliced average variance estimation. *Ann. Statist.*, 35(1):41–69, 2007.

[21] N. V. Nadkarni, Y. Zhao, and M. R. Kosorok. Inverse regression estimation for censored data. *Journal of the American Statistical Association*, 106(493), 2011.

[22] F. Portier and B. Delyon. Optimal transformation: a new approach for covering the central subspace. *J. Multivariate Anal.*, 115:84–107, 2013.

[23] F. Portier and B. Delyon. Bootstrap Testing of the Rank of a Matrix via Least-Squared Constrained Estimation. *J. Amer. Statist. Assoc.*, 109(505):160–172, 2014.

[24] J. Præstgaard and J. A. Wellner. Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.*, 21(4):2053–2086, 1993.

[25] D. E. Tyler. Asymptotic inference for eigenvectors. *Ann. Statist.*, 9(4):725–736, 1981.

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

[27] A. W. van der Vaart and J. A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.

[28] X. Yin and R. D. Cook. Dimension reduction for the conditional $k$th moment in regression. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64(2):159–175, 2002.

[29] L. Zhu and K. Fang. Asymptotics for kernel estimate of sliced inverse regression. *Ann. Statist.*, 24(3):1053–1068, 1996.

[30] L. Zhu and K. W. Ng. Asymptotics of sliced inverse regression. *Statist. Sinica*, 5(2):727–736, 1995.

[31] L. Zhu, L. Zhu, and Z. Feng. Dimension reduction in regressions through cumulative slicing estimation. *J. Amer. Statist. Assoc.*, 105(492):1455–1466, 2010.