Statistica Sinica **17**(2007), 1677-1681

# A NOTE ON SMOOTHED FUNCTIONAL INVERSE REGRESSION

L. Forzani and R. Dennis Cook

*University of Minnesota*

*Abstract:* Estimation in the context of functional data analysis is almost always non-parametric, since the object to be estimated lives in an infinite dimensional space. That is the case for the functional linear model with a real response and a process as covariables. In a recent paper Ferré and Yao state that the estimation of the Effective Dimension Reduction (EDR) subspace via SIR has parametric order. We show that a strong condition is needed for their statement to be true.

*Key words and phrases:* Dimension reduction, functional data analysis, inverse regression.

## 1. Introduction

Functional sliced inverse regression is the generalization of slice inverse regression (SIR; Li (1991)) to the infinite dimensional setting. Functional SIR was introduced by Dauxois, Ferré and Yao (2001) and Ferré and Yao (2003). Those papers show that root-$n$ consistent estimators cannot be expected. Ferré and Yao (2005) claimed a new method of estimation that is root-$n$ consistent. We argue that their result is not true under the conditions that they stated, but may be so when the covariance operator $\Gamma$ of the covariable $X$ is restricted. More specifically, root-$n$ consistency may be achievable when $\Gamma$ has an spectral decomposition with eigenfunctions of the covariance operator $\Gamma_e$ of $E(X|Y)$ or of the orthogonal complement of $\Gamma_e$. The EDR subspace can then be estimated as the span of the eigenfunctions of $\Gamma_e$, and therefore root-$n$ consistency follows from the root-$n$ consistency of principal component analysis for functional data (Dauxois, Pousse and Romain (1982)).

## 2. The Setting in Ferré and Yao (2005)

Let $(X, Y)$ be a random variable that takes values in $L^2[a, b] \times \mathbb{R}$. $X$ is a centered stochastic process with finite fourth moment. Then the covariance operators of $X$ and $E(X|Y)$ exist and are denoted by $\Gamma$ and $\Gamma_e$. $\Gamma$ is a Hilbert-Smith operator that is assumed to be positive definite.

Ferré and Yao (2005) assume the usual linearity condition for sliced inverse regression extended to functional data in the context of the model

$$Y = g(\langle \theta_1, X \rangle, \ldots, \langle \theta_D, X \rangle, \epsilon),$$

where $g$ is a function in $L^2[a,b]$, $\epsilon$ is a centered real random variable, $\theta_1, \ldots, \theta_D$ are $D$ independent functions in $L^2[a,b]$ and $\langle , \rangle$ indicates the usual inner product in $L^2[a,b]$. They called span$(\theta_1, \ldots, \theta_D)$ the Effective Dimension Reduction (EDR) subspace. Then, under their linearity condition the EDR subspace contains the $\Gamma$-orthonormal eigenvectors of $\Gamma^{-1}\Gamma_e$ associated with the positive eigenvalues. If an additional coverage condition is assumed then a basis for the EDR subspace will be the $\Gamma$-orthonormal eigenvectors of $\Gamma^{-1}\Gamma_e$ associated with the $D$ positive eigenvalues. Therefore the goal is to estimate the subspaces generated by those eigenvectors. Since $\Gamma$ is one-to-one and because of the coverage condition, the dimensions of $R(\Gamma_e)$ and $R(\Gamma^{-1}\Gamma_e)$ are both $D$. Here, $R(S)$ denotes the range of an operator $S$, which is the set of functions $S(f)$ with $f$ belonging to the domain $T(S)$ of the operator $S$.

To estimate $\Gamma_e$ it is possible to slice the range of $Y$ (Ferré and Yao (2003)) or to use a kernel approximation (Ferré and Yao (2005)). Under the conditions on the model, $L^2$ consistency and a central limit theorem follow for the estimators of $\Gamma_e$. To approximate $\Gamma$, in general, the sample covariance operator is used and consistency and central limit theorem for the approximation of $\Gamma$ follow (Dauxois, Pousse and Romain (1982)).

In a finite-dimensional context, the estimation of the EDR space does not pose any problem since $\Gamma^{-1}$ is accurately estimated by the inverse of the empirical covariance matrix of $X$. This is not true for functional inverse regression when, as assumed by Ferré and Yao (2005), $\Gamma$ is a Hilbert-Schmidt operator with infinite rank: the inverse is ill-conditioned if the range of $\Gamma$ is not finite dimensional. Regularization of the $\hat{\Gamma}$ can be used to overcome this difficulty. Estimation of $\Gamma_e$ is easier, since $\Gamma_e$ has finite rank. Because of the non-continuity of the inverse of a Hilbert-Smith operator, Ferré and Yao (2003) cannot get a root-$n$ consistent estimator of the EDR subspace. To overcome that difficulty Ferré and Yao (2005, Sec. 4), made the following comment:

> Under our model, $\Gamma^{-1}\Gamma_e$ has finite rank. Then, it has the same eigen subspace associated with positive eigenvalues as $\Gamma_e^{+}\Gamma$, where $\Gamma_e^{+}$ is a generalized inverse of $\Gamma_e$.

They use this comment to justify estimating the EDR subspace from the spectral decomposition of a root-$n$ consistent sample version of $\Gamma_e^{+}\Gamma$. However, the conclusion – $R(\Gamma^{-1}\Gamma_e) = R(\Gamma_e^{+}\Gamma)$ – in Ferré and Yao's comment is not true in the context used by them, but may hold true in a more restricted context. More specifically, additional structure seems necessary to equate $R(\Gamma_e^+\Gamma)$, the space that can be estimated, with $R(\Gamma^{-1}\Gamma_e)$ the space that we wish to know. For clarity and to study the implications of Ferré and Yao's claim we will use

**Condition A:** $R(\Gamma^{-1}\Gamma_e) = R(\Gamma_e^+\Gamma)$.

Condition A is equivalent to Ferré and Yao's claim stated previously. If Condition A were true then it would be possible to estimate the eigenvectors of $\Gamma^{-1}\Gamma_e$ more directly by using the eigenvectors of the operator $\Gamma_e$. In the next section we give justification for these claims, and provide necessary conditions for regressions in which Condition A holds. Since FDA is a relative new area, we do not know if Condition A is generally reasonable in practice. Further study is needed to resolve such issues.

## 3. The Results

We first give counter-examples to show that Condition A is not true in the context used by Ferré and Yao (2005), even in the finite dimensional case. Consider

$$\Gamma = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix} \text{ and } \Gamma_e = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix},$$

then $R(\Gamma^{-1}\Gamma_e) = \mathrm{span}((4,-1)')$ but $R(\Gamma_e^+\Gamma) = \mathrm{span}((1,0)')$ and so $R(\Gamma^{-1}\Gamma_e) \neq R(\Gamma_e^+\Gamma)$.

For the infinite dimensional case we consider $L^2[0,1]$ and any orthonormal basis $\{\phi_i\}_{i=1}^\infty$ of $L^2[0,1]$. We define $f = \sum_{i=1}^\infty a_i\phi_i$ with $a_i \neq 0$ and $\sum_{i=1}^\infty a_i^2 < \infty$. We define $\Gamma$ as the operator in $L^2[0,1]$ with eigenfunctions $\phi_i$ and corresponding eigenvalue $\lambda_i$. We ask that $\lambda_i > 0$ for all $i$ and $\sum_{i=1}^\infty \lambda_i^2 < \infty$. These conditions guarantee that $\Gamma$ is a Hilbert-Smith operator and strictly positive definite. Let $h = \Gamma(f)$; by definition, $h \in T(\Gamma^{-1})$. Now $h \notin \mathrm{span}(f)$. In fact, suppose $h = cf$. Then

$$h = \Gamma(f) = \sum_{i=1}^\infty \lambda_i\langle f,\phi_i\rangle\phi_i = c\sum_{i=1}^\infty \langle f,\phi_i\rangle\phi_i.$$

Now, since $\langle f,\phi_i\rangle = a_i \neq 0$ for all $i$ we have $\lambda_i = c$ for all $i$, contradicting the fact that $\sum_{i=1}^\infty \lambda_i^2 < \infty$.

Define the operator $\Gamma_e$ to be the identity operator in $\mathrm{span}(h)$ and 0 in $\mathrm{span}(h)^\perp$. Here given a set $B \subset L^2[0,1]$, let us denote by $B^\perp$ its orthogonal complement using the usual inner product in $L^2[a,b]$. The generalized inverse of $\Gamma_e$ coincides with $\Gamma_e$. Now, $R(\Gamma^{-1}\Gamma_e) =$ $\mathrm{span}(f)$ and $R(\Gamma_e^+\Gamma) = \mathrm{span}(h)$ and, from the fact that $h \notin \mathrm{span}(f)$, we get $R(\Gamma^{-1}\Gamma_e) \neq R(\Gamma_e^+\Gamma)$.

The next three lemmas give implications of Condition A.

**Lemma 1.** *If Condition A holds then* $R(\Gamma_e) = R(\Gamma^{-1}\Gamma_e)$.

**Proof.** The closure of the set $B \subset L^2[a,b]$, denoted by $\bar{B}$, will be the smallest closed set (using the topology defined through the usual inner product) containing $B$. For an operator $S$ from $L^2[a,b]$ into itself, let $S^*$ denote its adjoint operator, again using the usual inner product.

Let $\{\beta_1, \ldots, \beta_D\}$ denote the $D$ eigenfunctions, with eigenvalues nonzero, of $\Gamma_e^+\Gamma$. If Condition A is true then

$$\operatorname{span}(\beta_1, \ldots, \beta_D) = R(\Gamma^{-1}\Gamma_e) = R(\Gamma_e^+\Gamma) \subset R(\Gamma_e^+).$$

By definition of generalized inverse (Groetsch (1977)) we have

$$R(\Gamma_e^+) = N(\Gamma_e)^\perp = \overline{R(\Gamma_e^*)} = \overline{R(\Gamma_e)} = R(\Gamma_e),$$

where we use the fact that $\Gamma_e$ is self-adjoint and the fact that $R(\Gamma_e)$ has dimension $D$ and therefore is closed. Since $R(\Gamma_e)$ has dimension $D$, the result follows.

Lemma 1 shows that we can construct $\operatorname{span}(\beta_1, \ldots, \beta_D)$ from the $D$ eigenfunctions of $\Gamma_e$ associated with nonzero eigenvalues. From Daxouis, Pousse and Romain (1982), the eigenvectors of the approximate $\Gamma_e^n$ converge to the eigenvectors of $\Gamma_e$ at the root-$n$ rate ($\Gamma_e^n$ and $\Gamma_e$ have finite rank $D$ and therefore they are compact operators). Therefore we can approximate $\operatorname{span}(\beta_1, \ldots, \beta_D)$ at the same rate. Let us note that the $D$ eigenfunctions of $\Gamma_e$ need not be $\Gamma$-orthonormals.

**Lemma 2.** *Under Condition A we have* $R(\Gamma\Gamma_e) \subset R(\Gamma_e)$.

**Proof.** Since $\Gamma$ is one to one, $\overline{R(\Gamma)} = L^2[a,b]$. On the other hand, by hypothesis, $R(\Gamma_e) \subset T(\Gamma^{-1})$. From the definition of the inverse of an operator (Groetsch (1977)) we have that $\Gamma\Gamma^{-1} = \mathcal{I}_d$ in $T(\Gamma^{-1})$, where $\mathcal{I}_d$ indicates the identity operator. Now, let us take $v \in R(\Gamma\Gamma_e)$. Then $v = \Gamma\Gamma_e w$ for some $w \in L^2[a,b]$, and therefore $\Gamma^{-1}v = \Gamma_e w = \Gamma^{-1}\Gamma_e h$ for some $h \in L^2[a,b]$ (this last equality follows from Lemma 1). Since $\Gamma^{-1}$ is one to one (in its domain) we get $v = \Gamma_e h \in R(\Gamma_e)$.

In mathematical terms, $R(\Gamma\Gamma_e) \subset R(\Gamma_e)$ implies that $R(\Gamma_e)$ is an invariant subspace of the operator $\Gamma$ (see Conway (1990, p.39)). That, in turn, implies that $\Gamma$ has a spectral decomposition with eigenfunctions that live in $R(\Gamma_e)$ or its orthogonal complement, as indicated by the following lemma, the finite dimensional form of which was stated by Cook, Li and Chiaromonte (2006).

**Lemma 3.** *Suppose Condition A is true. Then* $\Gamma$ *has a spectral decomposition with eigenfunctions on* $R(\Gamma_e)$ *or* $R(\Gamma_e)^\perp$.

**Proof.** Let $v$ be an eigenvector of $\Gamma$ associated to the eigenvalue $\lambda > 0$. Since $R(\Gamma_e)$ is closed (for being finite dimensional), $v = u + w$ with $u \in R(\Gamma_e)$ and $w \in R(\Gamma_e)^\perp$. Since from Lemma 2, $\Gamma u \in R(\Gamma_e)$ and $\Gamma w \in R(\Gamma_e)^\perp$ we have that $u$ and $w$ are also eigenvectors of $\Gamma$ if both $u$ and $w$ are different from zero. Otherwise $v$ belongs to $R(\Gamma_e)$ or $R(\Gamma_e)^{\perp}$.

Now, let $\{v_i\}_{i=1}^{\infty}$ be a spectral decomposition of $\Gamma$. We can assure that there is a enumerable quantity of them since $\Gamma$ is compact in $L^2[0,1]$. From what we said above $v_i = u_i + w_i$ with $u_i$ and $w_i$ eigenvectors in $R(\Gamma_e)$ and $R(\Gamma_e)^{\perp}$, respectively. Now, we consider $\{u_i : u_i \neq 0\}$ and $\{w_i : w_i \neq 0\}$. Clearly they form a spectral decomposition of $\Gamma$ with eigenfunctions on $R(\Gamma_e)$ or $R(\Gamma_e)^{\perp}$.

### Acknowledgements

This work was supported in part by grant DMS-0405360 from the U.S. National Science Foundation.

## References

Conway, J. B. (1990). *A Course in Functional Analysis.* 2nd edition. Springer, New York.

Cook, R. D., Li, B. and Chiaromonte, F. (2006). Reductive Multivariate Linear Models. Preprint.

Dauxois, J., Ferré, L and Yao, A. (2001). Un modèle semi-paramétrique pour variables aléatoires. *C. R. Acad Paris* **333**, 947-952.

Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivariate Anal.* **12**, 136-154.

Ferré, L. and Yao, A. F. (2003). Functional sliced inverse regression analysis. *Statistics* **37**, 475-488.

Ferré, L. and Yao, A. F. (2005). Smoothed functional inverse regression. *Statist. Sinica* **15**, 665-683.

Groetsch, C. W. (1977). *Generalized Inverses of Linear Operators.* Marcel Dekker, Inc. New York.

Li, K. C. (1991). Sliced inverse regression for dimension reduction (with discussion). *J. Amer. Statist. Assoc.* **86**, 316-342.

School of Statistics, University of Minnesota, 1994 Buford Ave, St. Paul, MN 55108, U.S.A.

E-mail: lilianaforzani@gmail.com

School of Statistics, University of Minnesota, 1994 Buford Ave, St. Paul, MN 55108, U.S.A.

E-mail: dennis@stat.umn.edu

(Received September 2006; accepted February 2007)