

# A NON PARAMETRIC APPROACH FOR CALIBRATION WITH FUNCTIONAL DATA

Noslen Hernández<sup>1</sup>, Rolando J. Biscay<sup>2</sup>,  
Nathalie Villa-Vialaneix<sup>3,4</sup> and Isneri Talavera<sup>1</sup>

<sup>1</sup> *CENATAV, Havana - Cuba*

<sup>2</sup> *Centro de Investigacin en Matemticas, Guanajuato - Mexico*

<sup>3</sup> *SAMM, Université Paris 1 - France*

and <sup>4</sup> *INRA, UR875 MIA-T, Castanet Tolosan - France*

*Abstract:* A new nonparametric approach for statistical calibration with functional data is studied. The practical motivation comes from calibration problems in chemometrics in which a scalar random variable  $Y$  needs to be predicted from a functional random variable  $X$ . The proposed predictor takes the form of a weighted average of the observed values of  $Y$  in the training data set, where the weights are determined by the conditional probability density of  $X$  given  $Y$ . This functional density, which represents the data generation mechanism in the context of calibration, is so incorporated as a key information into the estimator. The new proposal is computationally simple and easy to implement. Its statistical consistency is proved, and its relevance is shown through simulations and an application to data.

*Key words and phrases:* Calibration, Functional data, Chemometrics, Inverse regression, Gaussian process.

## 1. Introduction

Statistical calibration plays a crucial role in such areas as pharmacology, neuroscience and chemometrics (Osborne (1991); Martens and Naes (1989); Brown (1993); Massart et al. (1997); Lavine and Workman (2002); Walters and Rizzuto (1988)). The calibration problem can be described as follows. An observable random variable  $X$  is related to a variable of interest  $Y$  according to a statistical model specified by a conditional probability density  $f(X|Y)$ . The density of  $Y$  may be imposed by the researcher (controlled or designed experiments) or given by nature (observational or natural experiments). A sample  $\mathcal{D}$  of independent observations  $(x_1, y_1), \dots, (x_n, y_n)$  of  $(X, Y)$  is available (training sample). Given a

new (future) observation  $x_0$  of  $X$  that corresponds to an unknown value  $y_0$  of  $Y$ , the problem is to make statistical inferences about  $y_0$  on the basis of the given statistical model, the data  $\mathcal{D}$  and  $x_0$ .

Our motivation comes from chemometrics, specifically from spectroscopy, where some chemical variable  $Y$  (e.g., concentration of a substance) needs to be predicted from a digitized function  $X$  (e.g., an absorbance spectrum). In this setting, the conditional density  $f(X|Y)$  represents the physical data generation mechanism in which the output spectrum  $X$  is determined by the input chemical concentration  $Y$ , plus some random perturbation mainly due to the measurement procedure. Then, given an observed spectrum  $x_0$  that corresponds to a new substance, one wishes an estimate of its concentration  $y_0$ , based on (past) observations of pairs of spectra and concentrations  $(x_i, y_i)$ .

Hereafter, we restrict ourselves to cases where the variable of interest  $Y$  takes real values (e.g., only the concentration of one substance is considered). In this framework, different calibration setups arise: the space in which the random variable  $X$  takes values can be the real line (univariate calibration), a finite-dimensional space (multivariate calibration), or a functional space (functional calibration); the experimental design can be fixed (the  $Y$  values are not random but set by the researcher) or random ( $Y$  is a random variable as well as  $X$ ); and the nature of the assumed statistical model  $f(X|Y)$  can be linear or nonlinear. A review of the literature on this subject, for both univariate and multivariate calibration, can be found in [Osborne \(1991\)](#).

This paper is concerned with functional calibration, which is useful for dealing with  $X$  measurements corresponding to spectra. In this context, the fact that the spectra are digitized measurements of a continuous phenomenon is directly included in the model by assuming that  $X$  lies in a functional space, such as  $L_2$ . The focus is on random design and nonlinear, in general nonparametric models (we refer the reader to [Cuevas et al. \(2002\)](#) and [Hernández et al. \(2012\)](#) for approaches on functional calibration to linear models under fixed design).

A widely used criterion for calibration in case of random design is mean squared error, which is minimized by the conditional mean  $E(Y|X = x_0)$ . This can be estimated by means of functional regression methods in which the response  $Y$  is a real random variable and the explanatory variable  $X$  is a function,

thus prediction methods that consider the conditional density  $f(Y|X)$  as the regression model. A number of such methods have been proposed. Seminal works focused on linear regression models ([Ramsay and Dalzell \(1991\)](#); [Cardot Ferraty and Sarda \(1999\)](#); [James and Hastie \(2001\)](#); [Ramsay and Silverman \(2005\)](#); [Ferraty and Vieu \(2006\)](#)). More recently, nonlinear functional models have been extensively developed and include nonparametric kernel regression ([Ferraty and Vieu \(2006\)](#)), Functional Inverse Regression ([Ferré and Yao \(2003, 2005\)](#)), neural networks ([Rossi and Conan-Guez \(2005\)](#); [Rossi et al. \(2005\)](#)),  $k$ -nearest neighbors ([Biau et al. \(2010\)](#); [Laloë \(2008\)](#)), Support Vector Regression ([Preda \(2007\)](#); [Hernández et al. \(2007\)](#)).

None of these approaches for predicting  $Y$  makes use of the specific structure of the density  $f(X|Y)$ , which in the calibration context plays the basic role of a physically justified regression model  $X$  *vs*  $Y$ . In the calibration framework, it is natural that the probabilistic assumptions refer to  $f(X|Y)$  that is the data generation mechanism. This is a major specificity of statistical calibration, in contrast with standard prediction problems in regression analysis ([Osborne \(1991\)](#)).

In [Hernández et al. \(2010, 2011\)](#) a new functional calibration approach, which we call Functional Density-Based Inverse Calibration (DBIC), was introduced. This method makes it possible to incorporate knowledge on the density of the regression model  $f(X|Y)$  for the prediction of a scalar variable  $Y$ , on the basis of a functional data  $X$ , taking into consideration the specificities of the prediction problem in the calibration setting. As is common in spectroscopy, this data generation model is assumed conditionally Gaussian. No parametric assumption is required about its mean and covariance functions, which provides remarkable flexibility in applications to capture nonlinear dependencies of  $X$  *vs*  $Y$ . Since the introduced predictor is an estimate of the conditional expectation  $E(Y|X)$ , we regarded it as an inverse calibration method, following customary terminology in the literature on statistical calibration ([Osborne \(1991\)](#)). In [Hernández et al. \(2010, 2011\)](#), preliminary results illustrated the computational feasibility and good behavior of the DBIC method in numerical simulations. However, no theoretical support to such findings has been published so far.

The main aim of the present paper is to provide a theoretical study of the consistency of the DBIC method, as well as to further assess its numerical perfor-

mance in a more elaborate simulation setting. It is organized as follows. Section 2 elaborates the method and proves its consistency. The proofs require the bringing together of results from nonparametric statistics, Functional Data Analysis (FDA) and equivalence of Gaussian measures; details are deferred to the Appendices. Section 3 shows the performance of the functional DBIC approach and provides comparison with functional kernel regression in a simulation study. Section 4 illustrates the method on a benchmark data set.

## 2. Functional Density-Based Inverse Calibration

Let  $(X, Y)$  be a pair of random variables taking values in  $\mathcal{X} \times \mathbb{R}$ , where  $\mathcal{X} = L_2([a, b])$ . Suppose  $n$  independent and identically distributed (i.i.d.) realizations of  $(X, Y)$  are given, denoted by  $(x_i, y_i)_{i=1, \dots, n}$ . The goal is to build a predictor of the value of  $Y$  corresponding to a future observed value of  $X$ . This problem is usually addressed through the estimation of the regression function  $\gamma(x) = E(Y|X = x)$ .

Here, a functional calibration method to estimate  $\gamma(X)$  is introduced. It relies on assuming:

$$X = r(Y) + e, \tag{2.1}$$

where  $e$  is a random process (perturbation or noise), independent of  $Y$ , and  $r$  is a function from  $\mathbb{R}$  into  $\mathcal{X}$ . Its motivation arises in calibration problems in chemometrics, specifically in spectroscopy, where some chemical variable  $Y$  needs to be predicted from a digitized function  $X$ . In this setting, the conditional mean  $r(y)$  of (2.1) represents the physical data generation mechanism. In this model, according to the physics of molecular spectroscopy, the spectrum  $X$  (recorded by an spectrometer) is determined by the input chemical concentration  $Y$ , and  $e$  is a functional random perturbation mainly due to the measurement procedure (Osborne (1991)). That is why this model, characterized by the conditional density  $f(X|Y)$ , is frequently referred to as the *hard model* or the *physical model* in spectroscopy (Kriesten et al. (2008b,a); Zhou and Cao (2013); Boulet and Roger (2010)). A simple instance of such a hard model in spectroscopy arises in case of an ideal mixture spectrum of pure components that includes a certain quantity of the component of interest  $Y$ , all obeying what is known as the *Lambert-Beer law*. In this case, as a consequence of this law,  $r(y)$  is simply a linear function of the concentration  $y$  (Naes et al. (2002)). The model assumptions in the present

work will allow to cover the remarkable more general hard-type model in which  $r(y)$  involves unknown nonlinearities (see [Chen and Morris \(2009\)](#); [Geladi MacDougall and Martens \(1985\)](#); [Melgaard and Haaland \(2004\)](#) for the importance of nonlinear effects in spectroscopy data).

We assume here that the perturbation  $e$  in [\(2.1\)](#) follows a Gaussian distribution  $P_0$  with zero mean and covariance  $\Gamma$ . This is a common assumption in the context of calibration. In applications,  $y \rightarrow r(y)$  represents an underlying input-output physical system and the perturbation  $e$  is interpreted as due to instrumental noise and possible uncontrolled factors. Popular methods in spectroscopy calibration that are based on multivariate hard models  $f(X|Y)$  usually have underlying Gaussian assumptions, that result in statistical procedures involving only the first two moments of the variables ([Martens and Naes \(1989\)](#); [Kriesten et al. \(2008a\)](#)). Hard models of Gaussian type for  $f(X|Y)$  lead to more complex, non-Gaussian inverse models for  $f(Y|X)$  if nonlinearities are involved in the conditional mean  $r(y)$ . This emphasizes that stating probability assumptions in terms of the hard model  $f(X|Y)$  is physically more meaningful, and easier, than for the inverse model  $f(Y|X)$ .

Under a Gaussian distribution assumption, the conditional distribution  $P(\cdot|y)$  is also Gaussian fully determined by its mean  $r(\cdot) = E(X|Y = \cdot)$ , and its covariance operator  $\Gamma$  (on the space  $\mathcal{X}$ ). There exists an eigenvalue decomposition of  $\Gamma$ ,  $(\varphi_j, \lambda_j)_{j \geq 1}$  such that  $(\lambda_j)_j$  is a decreasing sequence of positive real numbers,  $(\varphi_j)_j$  are orthonormal functions on  $\mathcal{X}$  and  $\Gamma = \sum_j \lambda_j \varphi_j \otimes \varphi_j$  where  $\varphi_j \otimes \varphi_j : h \in \mathcal{X} \rightarrow \langle \varphi_j, h \rangle \varphi_j$ .

If for each  $y \in \mathbb{R}$ ,  $\sum_{j=1}^{\infty} \frac{r_j^2(y)}{\lambda_j} < \infty$ , where  $r_j(y) = \langle r(y), \varphi_j \rangle$  for all  $j \geq 1$  then,  $P(\cdot|y)$  and  $P_0$  are equivalent Gaussian measures, and the density  $f(\cdot|y)$  of  $P(\cdot|y)$  with respect to  $P_0$  has the explicit form:  $f(x|y) = \exp \left\{ \sum_{j=1}^{\infty} \frac{r_j(y)}{\lambda_j} \left( x_j - \frac{r_j(y)}{2} \right) \right\}$ , where  $x_j = \langle x, \varphi_j \rangle$  for all  $j \geq 1$  ([Grenander \(1981\)](#)). Then, if the distribution of  $Y$  has a density  $f_Y(y)$  (with respect to the Lebesgue measure on  $\mathbb{R}$ ), the regression function can be written as  $\gamma(x) = \frac{\int_{\mathbb{R}} f(x|y) f_Y(y) y dy}{f_X(x)}$ , where  $f_X(x) = \int_{\mathbb{R}} f(x|y) f_Y(y) dy$ .

Given an estimate  $\hat{f}(x|y)$  of  $f(x|y)$ , this suggests an (plug-in) estimate of

$\gamma(x)$ :

$$\hat{\gamma}(x) = \frac{\frac{1}{n} \sum_{i=1}^n \hat{f}(x|y_i) y_i}{\hat{f}_X(x)}, \quad (2.2)$$

where  $\hat{f}(x|y)$  is an estimate of the density  $f(x|y)$  of  $P(\cdot|y)$  with respect to the measure  $P_0$ , and  $\hat{f}_X(x) = \frac{1}{n} \sum_{i=1}^n \hat{f}(x|y_i)$  is used to estimate the density  $f_X(x)$  of  $X$ .

An estimate  $\hat{f}(x|y)$  of  $f(x|y)$  can be obtained through the following steps:

1. For each  $t \in [0, 1]$ , compute an estimate  $\hat{r}(\cdot)(t)$  of the function  $r : y \mapsto r(y)(t)$ . This can be carried out through any standard nonparametric method for univariate regression based on the data set  $(y_i, x_i(t))_{i=1, \dots, n}$ . Here a smoothing kernel method, specifically the Nadaraya-Watson kernel estimate of  $r$ ,

$$\hat{r}(y) = \frac{\sum_{i=1}^n K\left(\frac{y_i - y}{h}\right) x_i}{\sum_{i=1}^n K\left(\frac{y_i - y}{h}\right)} = \frac{\hat{m}(y)}{\hat{f}_Y(y)}, \quad (2.3)$$

is used, where  $h$  is the bandwidth parameter,  $K$  an order  $k$  kernel,  $\hat{m}(y) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{y_i - y}{h}\right) x_i$ , and  $\hat{f}_Y(y) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{y_i - y}{h}\right)$ . In this case, the bandwidth  $h$  has a common value for all  $t$ .

2. Obtain estimates  $(\hat{\varphi}_j, \hat{\lambda}_j)_j$  of the eigenfunctions and eigenvalues  $(\varphi_j, \lambda_j)_j$  of the covariance  $\Gamma$  on the basis of the empirical covariance  $\hat{\Gamma}$  of the residuals  $\hat{e}_i = x_i - \hat{r}(y_i)$ ,  $\hat{\Gamma} = \frac{1}{n} \sum_{i=1}^n \hat{e}_i \otimes \hat{e}_i$ . Only the first  $p$  eigenvalues and eigenfunctions are incorporated, where  $p = p(n)$  is an integer smaller than  $n$ . This is a standard functional PCA problem.
3. Estimate  $f(x|y)$  by

$$\hat{f}(x|y) = \exp \left\{ \sum_{j=1}^p \frac{\hat{r}_j(y)}{\hat{\lambda}_j} \left( \hat{x}_j - \frac{\hat{r}_j(y)}{2} \right) \right\}, \quad (2.4)$$

where  $\hat{r}_j(y) = \langle \hat{r}(y), \hat{\varphi}_j \rangle$  and  $\hat{x}_j = \langle x, \hat{\varphi}_j \rangle$  for all  $j \geq 1$ .

Having  $\hat{f}(x|y)$ , substituting (2.4) into (2.2) leads to an estimate  $\hat{\gamma}(x)$  of  $\gamma(x)$ . It is referred to as the functional Density-Based Inverse Calibration (DBIC) because the conditional density  $\hat{f}(X|Y)$  plays a key role in its construction. If  $X$  had been a scalar variable, the proposal reduces to the univariate calibration described in [Lwin and Maritz \(1980\)](#).

The DBIC estimate has been explored under the assumption that  $f(X|Y)$  is Gaussian with covariance not depending on  $y$ , but the approach is general enough to allow for extension to non-Gaussian distributions. For instance, the method can be extended to the more general setting in which  $x$  is a diffusion process generated by the stochastic differential equation  $dx(t) = \dot{r}(y)(t) dt + b(x(t), t) dW(t)$ , where  $W(t)$  is a Brownian motion,  $\dot{r}$  denotes the derivative of  $r$  with respect to  $t$ , and  $b : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R}_+$  is a given function. Under mild conditions (e.g., if the function  $b$  is bounded away from zero and infinity, see [Liptser and Shiryaev \(1977\)](#)) the measure  $P(\cdot|y)$  of the solution  $x$  has a density with respect to the measure  $P_0 = P(\cdot|0)$ , given by  $f(x|y) = \exp \left\{ \int_0^1 \frac{\dot{r}(y)(t)}{b(x(t), t)} dx(t) - \frac{1}{2} \int_0^1 \left| \frac{\dot{r}(y)(t)}{b(x(t), t)} \right|^2 dt \right\}$ . Here, the integral with differential  $dx(t)$  is thought of as an Ito integral. If  $b(u, t)$  does not reduce to a function depending only on  $t$ , then the resulting random function  $x$  is not Gaussian.

Another wide class of non-Gaussian random functions for which their distributions have explicitly known densities  $f(x|y)$  with respect to some reference measure is generated by stochastic differential equations driven by an additive fractional Brownian motion  $W^H(t)$ . Specifically, random functions satisfying equations of the type  $x(t) = r(y)(t) + \int_0^t b(s) dW^H(s)$  have distributions  $P(\cdot|y)$  with explicitly known densities  $f(x|y)$  with respect to the measure  $P_0 = P(\cdot|0)$  (see, e.g., [Rao \(2010 chapter 2\)](#)).

For these kinds of non-Gaussian functional data, DBIC estimators can be carried out through the steps 1)-3). They differ only in their specific implementation of the approximation  $\hat{f}(x|y)$  that depends on the numerical computation of the stochastic integrals involved. We do not carry this further, but note that these matters are worth study and may be of interest in other applied fields.

We turn to asymptotic properties of the estimators proposed in steps (1)-(3) and provide a consistency result for the DBIC estimator  $\hat{\gamma}(x)$ . Proofs are given in the Appendices (Section 6). To obtain the consistency of  $\hat{\gamma}(x)$  to  $\gamma(x)$ , the same steps as the ones used for the DBIC estimation are followed.

The first step of the DBIC method is the estimation of the conditional mean  $r(y)$  by a Nadaraya-Watson kernel estimate  $\hat{r}(y)$  as in [\(2.3\)](#). A consistency result and a rate of convergence for  $\hat{r}(y)$  can be obtained under some assumptions:

- (A1)  $f_Y$  has support  $\Omega_Y \subset \mathbb{R}$ , and  $f_Y$  and  $r$  are  $\mathcal{C}^k$ , for a  $k \geq 2$ , on  $\Omega_Y$ .
- (A2)  $K$  is an order  $k$  kernel with compact support.
- (A3) There exists  $d_1$  and  $d_2$  such that  $\sup_{y \in \Omega_Y} |f_Y^{(k)}(y)| < d_1$  and  $\sup_{y \in \Omega_Y} \|r^{(k)}(y)\| < d_2$ .
- (A4)  $h = O(n^{-c_1})$ , where  $\frac{1}{4+2k} < c_1 < \frac{1}{4}$ ;
- (A5) There exists  $b_1 > 0$  such that  $\inf_{y \in \Omega_Y} f_Y(y) \geq b_1$ .
- (A6) There exists  $b_2 > 0$  such that  $\sup_{y \in \Omega_Y} \|r(y)\| \leq b_2$ .

**Proposition 1.** *Under (A1)-(A6),  $\sup_{y \in \Omega_Y} \|\hat{r}(y) - r(y)\| = O_P\left(n^{-c_1 k} + \left(\frac{\log n}{n^{1-2c_1}}\right)^{1/2}\right)$ .*

Assumptions (A1)-(A3) are standard in the framework of kernel-based density estimation (Rao (1983)). Assumption (A5) is satisfied in most calibration settings that motivate the present work; with minor technical modifications of the DBIC estimator, it can be replaced by the weaker assumption that, for any  $\delta > 0$ ,  $\sup_{y \in \Omega_Y, f_Y(y) < \delta} \|r(y)\|$  goes to zero when  $\delta$  does so. The estimation of  $r(y)$  by a Nadaraya-Watson kernel estimate is not necessary, this step (and the corresponding assumptions) can be replaced by any other (1-dimensional non-parametric) method leading to the same kind of convergence rate.

The second step of the DBIC method is the estimation of the covariance operator of the error,  $\Gamma$ , based on the estimated residuals. The consistency of this estimate with  $\sqrt{n}$ -rate ensures the consistency of the corresponding eigen-decomposition, using a result given in Bosq (1991). This convergence is needed in the last step of the DBIC method. To obtain the consistency of the covariance operator estimator, another assumption is required.

- (A7)  $e$  in model (2.1) is a Gaussian process.

This assumption, which serves as a basis for the DBIC method, implies the condition usually assumed on moments:  $E(\|e\|^4) < +\infty$ .

**Proposition 2.** *Under (A1)-(A7),  $\|\hat{\Gamma} - \Gamma\| = O_P\left(\frac{1}{n^{1/2-2c_1}}\right)$ , where  $\|\cdot\|$  denotes the operator norm.*

The last step of the DBIC method is to estimate the conditional density  $f(X|Y)$  from the eigen-decomposition of  $\hat{\Gamma}$ . The result here is derived from a Theorem in [Bosq \(1991\)](#), and the corresponding technical assumptions made therein are thus required. If  $(a_j)_j$  is the sequence defined by  $a_1 = 2\sqrt{2}/(\lambda_1 - \lambda_2)$  and  $a_j = 2\sqrt{2}/\min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})$ , it is assumed that

$$(A8) \quad \sum_{j=1}^{\infty} \sup_{y \in \Omega_Y} \frac{|r_j(y)|}{\sqrt{\lambda_j}} < \infty;$$

$$(A9) \quad \text{the } (\lambda_j)_j \text{ are all distinct};$$

$$(A10) \quad \lim_{n \rightarrow +\infty} p = +\infty;$$

$$(A11) \quad \lim_{n \rightarrow +\infty} \frac{\sum_{j=1}^p a_j}{\lambda_p n^{1/2-2c_1}} = 0;$$

$$(A12) \quad \frac{p}{\lambda_p^2} = O(n^q) \text{ for some } 0 < q < \min(c_1 k, \frac{1}{2} - c_1).$$

Since  $(\lambda_j)_j$  is a sequence of positive numbers decreasing to 0 when  $j$  tends to  $+\infty$ , (A8) implies that  $\sum_{j=1}^{\infty} \sup_{y \in \Omega_Y} |r_j(y)| < \infty$  and, consequently,  $\sum_{j=1}^{\infty} \sup_{y \in \Omega_Y} r_j^2(y) < \infty$ . Since  $\sup_{y \in \Omega_Y} \|r(y)\|^2 = \sup_{y \in \Omega_Y} \sum_{j=1}^{\infty} r_j^2(y) \leq \sum_{j=1}^{\infty} \sup_{y \in \Omega_Y} r_j^2(y) < \infty$ , (A8) implies that  $\sup_{y \in \Omega_Y} \|r(y)\| < \infty$ , which is the Assumption (A6) required in Propositions 1 and 2. Also, (A8) implies that  $\sum_{j=1}^{\infty} \frac{r_j^2(y)}{\lambda_j} = \sum_{j=1}^{\infty} \left( \frac{r_j(y)}{\sqrt{\lambda_j}} \right)^2 < \infty$ , which is the regularity assumption needed for the existence of the conditional density (see Section 2). Then, Assumption (A6) is no longer required for this proposition.

**Proposition 3.** *Under (A1)-(A5) and (A7)-(A12), for any  $x \in \mathcal{X}$ ,  $\sup_{y \in \Omega_Y} \left| \hat{f}(x|y) - f(x|y) \right| = o_P(1)$ .*

From this, the consistency of  $\hat{\gamma}(x)$ , defined at (2.2), can be proved. Proof is the in Appendix.

**Theorem 1.** *Under (A1)-(A5) and (A7)-(A12), for all  $x \in \mathcal{X}$  such that  $f_X(x) > 0$ , we have:  $\lim_{n \rightarrow +\infty} \hat{\gamma}(x) =^P \gamma(x)$ .*

From Theorem 1 and the Lebesgue's Dominated Convergence Theorem, proving that  $E(\hat{\gamma}(x) - \gamma(x))^2 \rightarrow_n 0$  is straightforward. Thus, the DBIC estimator converges in the sense of the quadratic Bayesian risk.

### 3. A simulation study

In this section, the feasibility and performance of the nonparametric functional calibration method is discussed through a simulation study. The physical data generation mechanism is specified by the density  $f(X|Y)$ . Here, training data were generated under various models. Based on this training data set, the DBIC estimator was computed to predict  $Y$  values corresponding to new values of the variable  $X$  (test data set).

The data were simulated as follows. The variable  $Y$  was uniform on the interval  $[0, 10]$ . For all models,  $e$  was a Gaussian process independent of  $Y$  with zero mean and covariance  $\Gamma = \sum_{j \geq 1} \frac{1}{j^{(1+0.1)}} v_j \otimes v_j$ , where  $v_{2k-1} = \sqrt{2} \cos(2\pi kt)$  and  $v_{2k} = \sqrt{2} \sin(2\pi kt)$ ,  $k = 1, \dots, 250$ . In this,  $X$  was generated in four settings.

**M1**  $E(X|Y)$  is the linear function of  $Y$  given by  $X = Y(v_1 + v_2 + v_5 + v_{10}) + e$ ;

**M2**  $E(X|Y)$  is a nonlinear function of  $Y$  given by  $X = \sin(Y)v_1 + \log(Y+1)v_5 + e$ ;

**M3**  $E(X|Y)$  is a linear function of  $Y$  given by  $X = Y(q_1 + 5q_2) + e$ , where  $q_1 = 2t^3$  and  $q_2 = t^4$ .

**M4**  $E(X|Y)$  is the nonlinear function of  $Y$  given by  $X = \sin(Y)q_1 + 20 \log(Y + 1)q_2 + e$ .

For each model, training and test samples of size  $n_L = 300$  and  $n_T = 200$ , respectively, were generated. To apply the DBIC method, simulated discretized functions were converted into continuous data (or functional predictors)  $X$  by approximation through 128 B-spline basis functions of order 4.

For the first step, the conditional mean  $r(y)$  was estimated from the training sample by kernel smoothing. For this, it was necessary to tune the bandwidth parameter  $h$ , and was done through a 10-fold cross-validation for minimizing the  $L_2$ -norm between the data and the estimated mean curves in the training sample.

The number  $p$  of eigenfunctions used to estimate  $f(x|y)$  was selected by a 10-fold cross-validation for minimizing the root mean squared error (RMSE) criterion on the training sample. For M1, the cross-validation gives the value  $p = 15$ , which is close to the true one ( $p = 10$ ) according to the model; for

---

All the simulations were done using Matlab<sup>®</sup> and the DBIC method was also implemented for Matlab<sup>®</sup>. Parts of the implementation use the Matlab<sup>®</sup> FDA functions developed by Jim Ramsay and freely available at <http://www.psych.mcgill.ca/faculty/ramsay/software.html>. The DBIC code is available upon request.

Model	DBIC	NWK	DBIC (parametric est. of the mean)
<i>M1</i>	0.23	0.28	0.22
<i>M2</i>	1.71	1.91	X
<i>M3</i>	0.07	0.19	0.02
<i>M4</i>	0.35	0.47	X

Table 3.1: RMSE achieved by DBIC and NWK for the four simulated models

M4 the resulting value was  $p = 47$ . Unlike M1, M4 was not built by using the first eigenfunctions of the covariance operator  $\Gamma$ , hence the need for more eigenfunctions.

Once the estimate  $\hat{\gamma}(x)$  was obtained on the basis of the training set, the performance was assessed according to RMSE was computed:  $\text{RMSE} = \sqrt{\frac{1}{n_T} \sum_{i=1}^{n_T} (y_i - \hat{y}_i)^2}$ , where  $y_i$  denotes the observed value of  $Y$  in the test sample and  $\hat{y}_i$  the corresponding prediction  $\hat{\gamma}(x_i)$ . For comparison, the standard functional nonparametric kernel estimate (NWK) (Ferraty and Vieu (2006)) was computed from the training sample (using a Gaussian kernel and also tuning the bandwidth parameter by 10-fold cross-validation on the training sample) and its predictions on the test set were calculated. Table 3.1 presents the DBIC and NWK RMSE for each of the simulated models. It can be observed that the DBIC outperforms the NWK estimator. The fourth column in the table is the RMSE achieved by DBIC using a parametric estimation of the mean; instead of estimating the mean using kernel smoothing, the mean was estimated by least squares for models M1 and M3. It can be observed that the RMSE resulting from such parametric estimates are smaller than those obtained by kernel smoothing. This illustrates that the DBIC approach has the flexibility to incorporate prior knowledge about the mean, if available, and that this additional information can improve the performance.

A detailed analysis of these experiments is provided in supplemental material.

#### 4. A study of Tecator dataset

DBIC was also tested on the Tecator dataset (Borggaard and Thodberg (1992)) which consists of spectrometric data from the food industry. Each of the 215 observations is the near infrared absorbency spectrum of a meat sam-

ple recorded on a Tecator Infratec Food and Feed Analyzer. Each spectrum was sampled at 100 wavelengths uniformly spaced in the range 850–1050 nm. The composition of each meat sample was determined by analytic chemistry, so percentages of moisture, fat, and protein were associated in this way to each spectrum. We focus on predicting the percentage of fat on the basis of the absorbency spectrum. This problem is more challenging than the ones in Section 3 where the data were generated to fulfill the conditions of the DBIC method. The data set was randomly split 100 times into training and test sets having approximately the same size. Table 4.2 reports the mean of the mean square error (MSE), and its standard deviation, over the 100 splits, for the DBIC and NWK methods.

Model	DBIC	NWK
MSE	2.41 (0.9)	11.01 (3.09)

Table 4.2: Prediction results on Tecator dataset

The results obtained by DBIC are remarkably better than those of NWK. In [Ferraty and Vieu \(2006\)](#), results based on the use of a semi-metric involving second order derivatives (known to be useful for this data set) were reported. Even incorporating this information in the model, a MSE of 3.5 was obtained, still larger than the one obtained by using DBIC without derivative information.

## 5. Conclusion

The functional Density-Based Inverse Calibration (DBIC) can be extended to other sample spaces and distribution families. Two appealing features of the method are its rather mild model assumptions and its computational simplicity, and one can incorporate parametric information on the conditional mean  $E(X|Y)$  of the “inverse” model if available. DBIC can be considered as a promising functional calibration method, particularly appealing for calibration problems in which said “inverse” model  $X$  vs.  $Y$  represents the actual physical mechanism generating the data. It would be interesting to obtain a limit distribution for the estimate in order to derive confidence bounds.

## 6. Appendix

In the appendices,  $m$  is the function defined on  $\Omega_Y$  such that  $r(y) = \frac{m(y)}{f_Y(y)}$  and  $g(x) = \int_{\mathbb{R}} f(x|y)yf_Y(y)dy$ ,  $\hat{g}(x) = \frac{1}{n} \sum_{i=1}^n \hat{f}(x|y_i)y_i$ .

*Proof of Proposition 1*

Two lemmas are needed to obtain Proposition 1. Their proofs can be found in the cited articles.

**Lemma 1.** (*Rao (1983)*) Under (A1)-(A3),  $\sup_{y \in \Omega_Y} \left| \hat{f}_Y(y) - f_Y(y) \right| = O_P \left( h^k + \frac{\sqrt{\log n}}{\sqrt{nh}} \right)$ .

**Lemma 2.** (*Yao (2001)*) Under (A1)-(A3),  $\sup_{y \in \Omega_Y} \|m(y) - \hat{m}(y)\| = O_P \left( h^k + \frac{\sqrt{\log n}}{\sqrt{nh}} \right)$ .

Proof of Proposition 1.

For any  $y \in \Omega_Y$ ,  $\|r(y) - \hat{r}(y)\| = \left\| \frac{r(y)}{\hat{f}_Y(y)} \left( \hat{f}_Y(y) - f_Y(y) \right) + \frac{1}{\hat{f}_Y(y)} (m(y) - \hat{m}(y)) \right\|$  which, by (A6), leads to

$$\begin{aligned} \sup_{y \in \Omega_Y} \|r(y) - \hat{r}(y)\| &\leq \frac{b_2}{\inf_{y \in \Omega_Y} |\hat{f}_Y(y)|} \sup_{y \in \Omega_Y} \left| \hat{f}_Y(y) - f_Y(y) \right| \\ &\quad + \frac{1}{\inf_{y \in \Omega_Y} |\hat{f}_Y(y)|} \sup_{y \in \Omega_Y} \|m(y) - \hat{m}(y)\|. \end{aligned}$$

From Lemma 1 and (A5), it follows that  $\frac{1}{\inf_{y \in \Omega_Y} |\hat{f}_Y(y)|} = \frac{1}{\inf_{y \in \Omega_Y} f_Y(y)} + o_P(1) \leq \frac{1}{b_1} + o_P(1)$ . From this and Lemma 1, Lemma 2, and (A4), we obtain  $\sup_{y \in \Omega_Y} \|r(y) - \hat{r}(y)\| = O_P \left( n^{-c_1 k} + \left( \frac{\log n}{n^{1-2c_1}} \right)^{1/2} \right)$ .  $\square$

**Lemma 3.** (*Cardot Ferraty and Sarda (1999)*) If  $Z$  is a random variable in a Hilbert space with covariance operator  $\Gamma_Z$  and  $E(\|Z\|^4) < +\infty$  then  $E(\|\Gamma_Z - \Gamma_Z^n\|^2) \leq \frac{E(\|Z\|^4)}{n}$ , where  $\Gamma_Z^n = \frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z}) \otimes (Z_i - \bar{Z})$ , and  $Z_i$  are independent and identically distributed (i.i.d) like  $Z$ .

Proof of Proposition 2.

By definition of the estimator  $\hat{\Gamma}$ , we have

$$\begin{aligned} \hat{\Gamma} &= \frac{1}{n} \sum_{i=1}^n \hat{e}_i \otimes \hat{e}_i = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{r}(y_i)) \otimes (x_i - \hat{r}(y_i)) \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - r(y_i) + r(y_i) - \hat{r}(y_i)) \otimes (x_i - r(y_i) + r(y_i) - \hat{r}(y_i)). \end{aligned}$$

This can be expressed as  $\hat{\Gamma} = \Gamma_n + T_1 + T_1^* + T_2$ , where

$$\begin{aligned}\Gamma_n &= \frac{1}{n} \sum_{i=1}^n e_i \otimes e_i = \frac{1}{n} \sum_{i=1}^n (x_i - r(y_i)) \otimes (x_i - r(y_i)), \\ T_1 &= \frac{1}{n} \sum_{i=1}^n (x_i - r(y_i)) \otimes (r(y_i) - \hat{r}(y_i)), \\ T_2 &= \frac{1}{n} \sum_{i=1}^n (r(y_i) - \hat{r}(y_i)) \otimes (r(y_i) - \hat{r}(y_i)),\end{aligned}$$

and  $T_1^*$  is the self-adjoint operator of  $T_1$ . Then,

$$\|\Gamma - \hat{\Gamma}\| \leq \|\Gamma - \Gamma_n\| + 2\|T_1\| + \|T_2\|. \quad (6.1)$$

Each term of the right side of this inequality is addressed separately.

From (A7) and Lemma 3 we obtain directly that

$$\|\Gamma - \Gamma_n\| = O_P(1/\sqrt{n}). \quad (6.2)$$

By definition of  $T_1$ , we have

$$\begin{aligned}n^{1/2-2c_1} \|T_1\| &\leq \frac{1}{n} \sum_{i=1}^n \|e_i\| n^{1/2-2c_1} \|r(y_i) - \hat{r}(y_i)\| \\ &\leq \frac{1}{n} \sum_{i=1}^n \|e_i\| \times n^{1/2-2c_1} \sup_{y \in \Omega_Y} \|r(y) - \hat{r}(y)\|.\end{aligned}$$

Thus, for any  $c > 0$ , using the Cauchy-Schwartz and Markov inequalities, we have

$$\begin{aligned}P\left(n^{1/2-2c_1} \|T_1\| > c\right) &\leq \left\{P\left(\frac{1}{n} \sum_{i=1}^n \|e_i\| > \sqrt{c}\right)\right\}^{\frac{1}{2}} \times \\ &\quad \left\{P\left(n^{1/2-2c_1} \sup_{y \in \Omega_Y} \|r(y) - \hat{r}(y)\| > \sqrt{c}\right)\right\}^{\frac{1}{2}} \\ &\leq \left\{\frac{E(\|e\|)}{\sqrt{c}}\right\}^{\frac{1}{2}} \left\{P\left(n^{1/2-2c_1} \sup_{y \in \Omega_Y} \|r(y) - \hat{r}(y)\| > \sqrt{c}\right)\right\}^{\frac{1}{2}}\end{aligned}$$

As  $E\|e\|^2 < +\infty$ , and with Proposition 1,  $\sup_{y \in \Omega_Y} \|r(y) - \hat{r}(y)\| = O_P\left(n^{-c_1 k} + \left(\frac{\log n}{n^{1-2c_1}}\right)^{1/2}\right)$ , we have that  $n^{1/2-2c_1} \sup_{y \in \Omega_Y} \|r(y) - \hat{r}(y)\| =$

$O_P\left(n^{1/2-c_1(2+k)} + \frac{\sqrt{\log n}}{n^{c_1}}\right)$  which is  $o_P(1)$  by (A4). Hence,  
 $\lim_{n \rightarrow +\infty} P\left(n^{1/2-2c_1} \|T_1\| > c\right) = 0$  and

$$\|T_1\| = o_P\left(\frac{1}{n^{1/2-2c_1}}\right). \quad (6.3)$$

By definition of  $T_2$ , we have

$$\begin{aligned} \sqrt{n} \|T_2\| &\leq \sqrt{n} \frac{1}{n} \sum_{i=1}^n \|r(y_i) - \hat{r}(y_i)\|^2 \\ &\leq \sqrt{n} \sup_{y \in \Omega_Y} \|r(y) - \hat{r}(y)\|^2. \end{aligned}$$

By Proposition 1, this gives  $\sqrt{n} \|T_2\| = O_P\left(\left(n^{-c_1 k + 1/4} + \left(\frac{\log n}{n^{1/2-2c_1}}\right)^{1/2}\right)^2\right)$   
but, by (A4),  $1/2 - 2c_1 > 0$ , and also, since  $k \geq 2$ ,  $c_1 > \frac{1}{4+2k} \geq \frac{1}{4k}$ , and so  $-kc_1 + 1/4 < 0$ . Then  $\sqrt{n} \|T_2\| = o_P(1)$  and thus

$$\|T_2\| = o_P(1/\sqrt{n}). \quad (6.4)$$

Using (6.2), (6.3) and (6.4) in (6.1),  $\|\Gamma - \hat{\Gamma}\| = O_P\left(\frac{1}{n^{1/2-2c_1}}\right)$ .  $\square$

**Proof of Proposition 3.**

The proof of Proposition 3 requires the use of a lemma whose proof follow that of Bosq (1991) for the particular case in which  $\tilde{\Delta}$  is an empirical covariance operator associated with the covariance operator  $\Delta$ .

**Lemma 4.** *Let  $\Delta$  and  $\tilde{\Delta}$  be two linear self-adjoint and compact operators defined in a Hilbert space, with  $(\nu_j, \phi_j)_{j \in \mathbb{N}}$ ,  $(\tilde{\nu}_j, \tilde{\phi}_j)_{j \in \mathbb{N}}$  being the respective decreasing sequence of eigenvalues and sequence of orthonormal eigenvectors. Then, for all  $j \in \mathbb{N}$ ,*

$$i) |\nu_j - \tilde{\nu}_j| \leq \|\Delta - \tilde{\Delta}\|;$$

$$ii) \|\phi_j - \tilde{\phi}_j\| \leq a_j \|\Delta - \tilde{\Delta}\| \text{ where } a_j = \begin{cases} \frac{2\sqrt{2}}{\nu_1 - \nu_2} & \text{if } j = 1 \\ \frac{2\sqrt{2}}{\min(\nu_{j-1} - \nu_j, \nu_j - \nu_{j+1})} & \text{if } j \geq 2 \end{cases}.$$

**Proof of Proposition 3.**

For any  $y \in \Omega_Y$ , let

$$\begin{aligned} \mathcal{E}(y) &= \left| \ln \hat{f}(x|y) - \ln f(x|y) \right| \\ &= \left| \sum_{j=1}^p \frac{\hat{r}_j(y)}{\hat{\lambda}_j} \left( \hat{x}_j - \frac{\hat{r}_j(y)}{2} \right) - \sum_{j \geq 1} \frac{r_j(y)}{\lambda_j} \left( x_j - \frac{r_j(y)}{2} \right) \right|. \end{aligned}$$

Then,  $\mathcal{E}(y) \leq \mathcal{E}_1(y) + \mathcal{E}_2(y)$  where  $\mathcal{E}_1(y) = \left| \sum_{j=p+1}^{+\infty} \frac{r_j(y)}{\lambda_j} \left( x_j - \frac{r_j(y)}{2} \right) \right|$  and  $\mathcal{E}_2(y) = \left| \sum_{j=1}^p \left[ \frac{\hat{r}_j(y)}{\hat{\lambda}_j} \left( \hat{x}_j - \frac{\hat{r}_j(y)}{2} \right) - \frac{r_j(y)}{\lambda_j} \left( x_j - \frac{r_j(y)}{2} \right) \right] \right|$ .

The Karhunen-Loeve expansion of  $x$  has coordinates  $x_j = r_j(y) + \sqrt{\lambda_j} \xi_j$ , where the  $\xi_j$  are independent standard normal. Then,

$$\begin{aligned} \sup_{y \in \Omega_Y} \mathcal{E}_1(y) &= \sup_{y \in \Omega_Y} \left| \sum_{j=p+1}^{+\infty} \frac{r_j(y)}{\lambda_j} \sqrt{\lambda_j} \xi_j + \sum_{j=p+1}^{+\infty} \frac{r_j^2(y)}{2\lambda_j} \right| \\ &\leq \sum_{j=p+1}^{+\infty} \sup_{y \in \Omega_Y} \frac{|r_j(y)|}{\sqrt{\lambda_j}} |\xi_j| + \sum_{j=p+1}^{+\infty} \sup_{y \in \Omega_Y} \frac{|r_j(y)|^2}{2\lambda_j}. \end{aligned} \quad (6.5)$$

(A8) implies that  $\sum_{j=1}^{+\infty} \sup_{y \in \Omega_Y} \left( \frac{|r_j(y)|}{\sqrt{\lambda_j}} \right)^2 < +\infty$ , hence

$$\sum_{j=p+1}^{+\infty} \sup_{y \in \Omega_Y} \frac{|r_j(y)|^2}{\lambda_j} \xrightarrow{p \rightarrow +\infty} 0. \quad (6.6)$$

$E \left( \sum_{j=1}^{+\infty} \sup_{y \in \Omega_Y} \frac{|r_j(y)|}{\sqrt{\lambda_j}} |\xi_j| \right) = \sum_{j=1}^{+\infty} \sup_{y \in \Omega_Y} \frac{|r_j(y)|}{\sqrt{\lambda_j}} < +\infty$  by (A8), which implies that  $P_0 - a.s.$ ,  $\sum_{j=1}^{+\infty} \sup_{y \in \Omega_Y} \frac{|r_j(y)|}{\sqrt{\lambda_j}} |\xi_j| < +\infty$  and then

$$P_0 - a.s., \quad \sum_{j=p+1}^{+\infty} \sup_{y \in \Omega_Y} \frac{|r_j(y)|}{\sqrt{\lambda_j}} |\xi_j| \xrightarrow{p \rightarrow +\infty} 0. \quad (6.7)$$

Putting (6.6) and (6.7) into (6.5) leads to  $P_0 - a.s.$ ,  $\sup_{y \in \Omega_Y} \mathcal{E}_1(y) \xrightarrow{p \rightarrow +\infty} 0$ .  $\mathcal{E}_2$  can be divided into four parts:  $\mathcal{E}_2(y) \leq A(y) + B(y) + C(y) + D(y)$ , where

$$A(y) = \left| \sum_{j=1}^p \frac{r_j(y)}{\lambda_j} (x_j - \hat{x}_j) \right|.$$

$$B(y) = \left| \sum_{j=1}^p \frac{r_j(y)}{2\lambda_j} (r_j(y) - \hat{r}_j(y)) \right|.$$

$$C(y) = \left| \sum_{j=1}^p \frac{\hat{x}_j - \hat{r}_j(y)/2}{\lambda_j} (r_j(y) - \hat{r}_j(y)) \right|.$$

$$D(y) = \left| \sum_{j=1}^p \left( \frac{1}{\lambda_j} - \frac{1}{\hat{\lambda}_j} \right) \hat{r}_j(y) \left( \hat{x}_j - \frac{\hat{r}_j(y)}{2} \right) \right|.$$

From  $\sum_j r_j^2(y) = \|r(y)\|^2$  it follows that  $|r_j(y)| \leq \|r(y)\|$  for all  $j$ . Then, since  $\lambda_j \geq \lambda_p$  for  $j = 1, \dots, p$  and  $|x_j - \hat{x}_j| = |\langle x, \varphi_j - \hat{\varphi}_j \rangle| \leq \|x\| \|\varphi_j - \hat{\varphi}_j\|$  we obtain, for any  $y \in \Omega_Y$ ,

$$\begin{aligned} A(y) &\leq \frac{\|r(y)\|}{\lambda_p} \|x\| \sum_{j=1}^p \|\varphi_j - \hat{\varphi}_j\| \\ &\leq \|r(y)\| \|x\| n^{1/2-2c_1} \|\Gamma - \hat{\Gamma}\| \frac{\sum_{j=1}^p a_j}{n^{1/2-2c_1} \lambda_p}, \end{aligned}$$

where the last inequality follows from Lemma 4. Since  $n^{1/2-2c_1} \|\Gamma - \hat{\Gamma}\| = O_P(1)$  by Proposition 2,  $\frac{\sum_{j=1}^p a_j}{n^{1/2-2c_1} \lambda_p} \xrightarrow{n \rightarrow +\infty} 0$  by (A11), and  $\sup_{y \in \Omega_Y} \|r(y)\| < \infty$  by (A8), we have that  $\sup_{y \in \Omega_Y} A(y) = o_P(1)$ .

By the same arguments as those used for  $A$ , we have  $B(y) \leq \frac{\|r(y)\|}{2\lambda_p} \sum_{j=1}^p |r_j(y) - \hat{r}_j(y)|$ . Moreover, we have that, for any  $j$  and any  $y \in \Omega_Y$ ,  $|r_j(y) - \hat{r}_j(y)| \leq \|r(y)\| \|\varphi_j - \hat{\varphi}_j\| + \|\hat{\varphi}_j\| \|r(y) - \hat{r}(y)\|$ . Thus, applying Lemma 4 and Proposition 1 we obtain, for any  $j$ ,

$$\begin{aligned} \sup_{y \in \Omega_Y} |r_j(y) - \hat{r}_j(y)| &\leq \sup_{y \in \Omega_Y} \|r(y)\| a_j \|\Gamma - \hat{\Gamma}\| \\ &\quad + O_P \left( n^{-c_1 k} + \left( \frac{\log n}{n^{1-2c_1}} \right)^{1/2} \right). \end{aligned}$$

Then,

$$\begin{aligned} \sup_{y \in \Omega_Y} B(y) &\leq \frac{1}{2} \left( \sup_{y \in \Omega_Y} \|r(y)\| \right)^2 n^{1/2-2c_1} \|\Gamma - \hat{\Gamma}\| \frac{\sum_{j=1}^p a_j}{\lambda_p n^{1/2-2c_1}} \\ &\quad + \frac{1}{2} \sup_{y \in \Omega_Y} \|r(y)\| \frac{p}{\lambda_p} O_P \left( n^{-c_1 k} + \left( \frac{\log n}{n^{1-2c_1}} \right)^{1/2} \right) \end{aligned}$$

where the first term is  $o_P(1)$  due to (A11) and Proposition 2, and the second term is  $O_P \left( \frac{1}{n^{c_1 k - q}} + \frac{(\log n)^{1/2}}{n^{1/2 - c_1 - q}} \right)$  by (A12) (because, taking into consideration that  $\lambda_j$  is a decreasing sequence, (A12) implies that  $p/\lambda_p = O(n^q)$ ). Since by Assumption (A12),  $c_1 k - q > 0$  and  $1/2 - c_1 - q > 0$ , the second term in the last inequality is also  $o_P(1)$ . Thus,  $\sup_{y \in \Omega_Y} B(y) = o_P(1)$ .

From  $|\hat{x}_j| \leq \|x\|$  and  $|\hat{r}_j(y)| \leq \|\hat{r}(y)\|$ , we have that

$$C(y) \leq \frac{\|x\| + \|\hat{r}(y)\|}{\lambda_p} \sum_{j=1}^p |r_j(y) - \hat{r}_j(y)|. \quad \text{Thus,} \quad \sup_{y \in \Omega_Y} C(y) \leq$$

$(\|x\| + \sup_{y \in \Omega_Y} \|\hat{r}(y)\|) \frac{1}{\lambda_p} \sup_{y \in \Omega_Y} \sum_{j=1}^p |r_j(y) - \hat{r}_j(y)|$ . As with  $B$ , it can be shown that  $\frac{1}{\lambda_p} \sup_{y \in \Omega_Y} \sum_{j=1}^p |r_j(y) - \hat{r}_j(y)| = o_P(1)$ . Moreover,  $\sup_{y \in \Omega_Y} \|\hat{r}(y)\| \leq \sup_{y \in \Omega_Y} \|r(y)\| + \sup_{y \in \Omega_Y} \|r(y) - \hat{r}(y)\| = O_P(1)$  by Proposition 1 and (A8). Putting all this together gives  $\sup_{y \in \Omega_Y} C(y) = o_P(1)$ .

From the same arguments as for  $C$ , and Lemma 4, we have that

$$\begin{aligned} D(y) &\leq \|\hat{r}(y)\| \left( \|x\| + \frac{\|\hat{r}(y)\|}{2} \right) \sum_{j=1}^p \left| \frac{1}{\lambda_j} - \frac{1}{\hat{\lambda}_j} \right| \\ &\leq \left( \|\hat{r}(y)\| \|x\| + \frac{\|\hat{r}(y)\|^2}{2} \right) \frac{p \|\Gamma - \hat{\Gamma}\|}{\lambda_p \hat{\lambda}_p}. \end{aligned}$$

Now by using  $\hat{\lambda}_p \geq |\lambda_p - |\lambda_p - \hat{\lambda}_p||$  we have that  $\frac{p \|\Gamma - \hat{\Gamma}\|}{\lambda_p \hat{\lambda}_p} \leq \frac{p \|\Gamma - \hat{\Gamma}\|}{|\lambda_p^2 - \lambda_p |\lambda_p - \hat{\lambda}_p|} = \frac{\frac{p}{\lambda_p^2} \|\Gamma - \hat{\Gamma}\|}{\left| 1 - \frac{|\lambda_p - \hat{\lambda}_p|}{\lambda_p} \right|}$ . Using the expansion  $\frac{1}{1-x} = \sum_{j=0}^{\infty} x^j$  for all  $|x| < 1$  together with Lemma 4,

$$\begin{aligned} \frac{\frac{p}{\lambda_p^2} \|\Gamma - \hat{\Gamma}\|}{\left| 1 - \frac{|\lambda_p - \hat{\lambda}_p|}{\lambda_p} \right|} &= \frac{p}{\lambda_p^2} \|\Gamma - \hat{\Gamma}\| \left( 1 + \frac{|\lambda_p - \hat{\lambda}_p|}{\lambda_p} + o\left(\frac{|\lambda_p - \hat{\lambda}_p|}{\lambda_p}\right) \right) \\ &= \frac{p}{\lambda_p^2} O_P\left(\frac{1}{\sqrt{n}}\right) + \frac{p}{\lambda_p^3} O_P\left(\frac{1}{n}\right) + \frac{p}{\lambda_p^2} O_P\left(\frac{1}{\sqrt{n}}\right) o_P\left(\frac{1}{\sqrt{n}\lambda_p}\right). \end{aligned}$$

Here, the first term is  $o_P(1)$  by (A4) and (A12) (taking into consideration that the last one implies  $q - 1/2 < 0$ ). The second and third terms are equivalent to  $O_P\left(\frac{p}{\lambda_p^2 \sqrt{n}} \times \frac{1}{\lambda_p \sqrt{n}}\right)$  which is also  $o_P(1)$  due to (A11) and (A12).

Hence,  $\sup_{y \in \Omega_Y} D(y) \leq \left( \sup_{y \in \Omega_Y} \|\hat{r}(y)\| \|x\| + \frac{1}{2} \left( \sup_{y \in \Omega_Y} \|\hat{r}(y)\| \right)^2 \right) o_P(1)$ . In demonstrating the convergence of  $C$ , we showed that  $\sup_{y \in \Omega_Y} \|\hat{r}(y)\| = O_P(1)$ , so it can be concluded that  $\sup_{y \in \Omega_Y} D(y) = o_P(1)$ .

Now,  $\sup_{y \in \Omega_Y} \mathcal{E}(y) = o_P(1)$ . On the other hand, for any  $\eta > 0$ ,

$$\begin{aligned} P\left(\sup_{y \in \Omega_Y} |f(x|y) - \hat{f}(x|y)| > \eta\right) &\leq P\left(\sup_{y \in \Omega_Y} |f(x|y) - \hat{f}(x|y)| > \eta, \sup_{y \in \Omega_Y} \mathcal{E}(y) \leq \frac{1}{2}\right) \\ &\quad + P\left(\sup_{y \in \Omega_Y} \mathcal{E}(y) > \frac{1}{2}\right) \\ &\leq P\left(\sup_{y \in \Omega_Y} \mathcal{E}(y) e^{1/2} \sup_{y \in \Omega_Y} f(x|y) > \eta\right) \\ &\quad + P\left(\sup_{y \in \Omega_Y} \mathcal{E}(y) > \frac{1}{2}\right), \end{aligned}$$

where  $\sup_{y \in \Omega_Y} f(x|y)$  is finite by (A8). The right hand side of the last inequality goes to zero as  $n$  increases, which concludes the proof.  $\square$

**Proposition 4.** *Under (A1)-(A5) and (A7)-(A12), for any  $x \in \mathcal{X}$ ,  $|f_X(x) - \hat{f}_X(x)| = o_P(1)$ , and  $|\frac{1}{n} \sum_{i=1}^n \hat{f}(x|y_i) y_i - \int_{\mathbb{R}} f(x|y) y f_Y(y) dy| = o_P(1)$ .*

*Proof of Proposition 4.*

For any  $x \in \mathcal{X}$ ,  $|f_X(x) - f_X(x)| \leq |f_X(x) - \frac{1}{n} \sum_{i=1}^n f(x|y_i)| + |\frac{1}{n} \sum_{i=1}^n f(x|y_i) - f_X(x)|$ . By Proposition 3,  $|\hat{f}_X(x) - \frac{1}{n} \sum_{i=1}^n f(x|y_i)| \leq \sup_{y \in \Omega_Y} |f(x|y) - \hat{f}(x|y)| = o_P(1)$ . (A8) ensures that, for all  $x \in \mathcal{X}$ ,  $f_X(x)$  is finite. Hence, by the Law of Large Numbers,  $\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n f(x|y_i) =^{as} E_Y(f(x|Y)) = f_X(x)$ . These two arguments complete the first part of the proof.

The second part is similar. For any  $x \in \mathcal{X}$ ,

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^n \hat{f}(x|y_i) y_i - \int_{\mathbb{R}} f(x|y) y f_Y(y) dy \right| &\leq \left| \frac{1}{n} \sum_{i=1}^n \hat{f}(x|y_i) y_i - \frac{1}{n} \sum_{i=1}^n f(x|y_i) y_i \right| \\ &\quad + \left| \frac{1}{n} \sum_{i=1}^n f(x|y_i) y_i - \int_{\mathbb{R}} f(x|y) y f_Y(y) dy \right|. \end{aligned}$$

Here, the first part of the right side of this inequality is bounded by  $\sup_{y \in \Omega_Y} |f(x|y) - \hat{f}(x|y)| \times \frac{1}{n} \sum_i y_i$ . As  $E(Y) < \infty$ , we have that  $\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_i y_i =^{as} E(Y)$ . By Proposition 3,  $\sup_{y \in \Omega_Y} |f(x|y) - \hat{f}(x|y)| = o_P(1)$ . Thus,  $|\frac{1}{n} \sum_{i=1}^n \hat{f}(x|y_i) y_i - \frac{1}{n} \sum_{i=1}^n f(x|y_i) y_i| = o_P(1)$ . The second part of the right side of the inequality converges to 0 almost surely by the Law of Large Numbers under the fact that  $E(Y) < \infty$ .  $\square$

Proof of Theorem 1.

For any  $x \in \mathcal{X}$  such that  $f_X(x) > 0$ ,

$$\begin{aligned} |\hat{\gamma}(x) - \gamma(x)| &= \left| \frac{\hat{g}(x)f_X(x) - g(x)\hat{f}_X(x)}{f_X(x)\hat{f}_X(x)} \right| \\ &\leq \frac{1}{|f_X(x)|} |g(x) - \hat{g}(x)| + \left| \frac{\hat{g}(x)}{f_X(x)\hat{f}_X(x)} \right| |f_X(x) - \hat{f}_X(x)|. \end{aligned}$$

Since,  $|\hat{g}(x)| \leq \|\hat{g}(x) - g(x)\| + |g(x)|$  and  $\hat{f}_X(x) \geq |f_X(x) - |\hat{f}_X(x) - f_X(x)||$ ,

$$\begin{aligned} |\hat{\gamma}(x) - \gamma(x)| &\leq \frac{|g(x) - \hat{g}(x)|}{f_X(x)} + \\ &\quad \frac{\|\hat{g}(x) - g(x)\| + |g(x)|}{|f_X(x) - |\hat{f}_X(x) - f_X(x)||} \frac{1}{f_X(x)} |f_X(x) - \hat{f}_X(x)|. \end{aligned}$$

Since  $f_X(x)$  is finite (A8) and positive, the first term in the right side of this inequality is  $o_P(1)$  by Proposition 4. By Proposition 4, the second term is  $\frac{|g(x)|}{f_X(x)^2} o_P(1)$ , which is trivially  $o_P(1)$  since  $g(x)$  is finite for all  $x$ . Hence we can conclude that  $|\hat{\gamma}(x) - \gamma(x)| = o_P(1)$ , which completes the proof.  $\square$

### Acknowledgment

This work is partially supported by the project CONICYT ACT 1112.

### References

- Biau, G., Cérou, F., and Guyader, A. (2010). Rates of convergence of the functional k-nearest neighbor estimate. *IEEE Transactions on Information Theory*, **56**, 2034–2040.
- Borggaard, C. and Thodberg, H. (1992). Optimal minimal neural interpretation of spectra. *Analytical Chemistry*, **64**, 545–551.
- Bosq, D. (1991). Modelization, Non-Parametric Estimation and Prediction for Continuous Time Processes, *ASI Series*. **335**, 509–529. NATO.
- Boulet, J. and Roger, J. (2010). Improvement of direct calibration in spectroscopy. *Analytica Chimica Acta*. **66**, 130–136.
- Brown, P. (1993). *Measurement, Regression and Calibration*. Oxford.
- Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. *Statistics and Probability Letters*. **45**, 11–22.
- Chen, Z. and Morris, J. (2009). Process analytical technology and compensating for nonlinear effects in process spectroscopic data for improved process monitoring and control. *Biotechnology Journal*. **4**, 610–619.
- Cuevas, A., Febrero, M., and Fraiman, R. (2002). Linear functional regression: the case of fixed design and functional response. *The Canadian Journal of Statistics*. **30**, 285–300.
- Ferraty, F. and Vieu, P. (2006). *NonParametric Functional Data Analysis*. Springer.
- Ferré, L. and Yao, A. (2003). Functional sliced inverse regression analysis. *Statistics*. **37**, 475–488.

- Ferré, L. and Yao, A. (2005). Smoothed functional inverse regression. *Statistica Sinica*. **15**, 665–683.
- Geladi, P., MacDougall, D., and Martens, H. (1985). Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy*. **39**, 491–500.
- Grenander, U. (1981). *Abstract Inference*. Berlin.
- Hernández, N., Biscay, R., and Talavera, I. (2007). Progress in Pattern Recognition, Image Analysis and Applications (*Proceedings of 12th Iberoamericann Congress on Pattern Recognition*, 564–573. Springer, Berlin/Heidelberg.
- Hernández, N., Biscay, R., and Talavera, T. (2012). A non-Bayesian predictive approach for statistical calibration. *Journal of Statistical Computation and Simulation*. **82**, 529–545.
- Hernández, N., Biscay, R., Villa-Vialaneix, N., and Talavera, I. (2011). A simulation study of functional density-based inverse regression. *Revista Investigacion Operacional*. **32**, 146–159.
- Hernández, N., Biscay, R., Villa-Vialaneix, N., and Talavera-Bustamante, I. (2010). A functional density-based nonparametric approach for statistical calibration. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Bloch, I. and Cesar, R., editors, 450–457. Springer.
- James, G. and Hastie, T. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society, B*. **63**, 533–550.
- Kriesten, E., Alsmeyer, F., Bardow, A., and Marquardt, W. (2008a). Fully automated indirect hard modeling of mixture spectra. *Chemometrics and Intelligent Laboratory Systems*. **91**, 181–193.
- Kriesten, E., Mayer, D., Alsmeyer, F., Minnich, C., Greiner, L., and Marquardt, W. (2008b). Identification of unknown pure component spectra by indirect hard modeling. *Chemometrics and Intelligent Laboratory Systems*. **93**, 108–119.

- Laloë, T. (2008). A k-nearest neighbor approach for functional regression. *Statistics and Probability Letters*, **78**, 1189–1193.
- Lavine, B. and Workman, J. (2002). Fundamental reviews: chemometrics. *Analytical Chemistry*, **74**, 2763–2770.
- Liptser, R. and Shiryaev, A. (1977). *Statistics of Random Processes I. General Theory*. Springer, New York.
- Lwin, T. and Maritz, J. (1980). A note on the problem of statistical calibration. *Journal of the Royal Statistical Society C*, **29**, 135–141.
- Martens, H. and Naes, T. (1989). *Multivariate Calibration*. Chichester.
- Massart, D., Vandeginste, B., Buydens, L., Jong, S., Lewi, P., and Smeyers-Verbeke, J. (1997). *Handbook of Chemometrics and Qualimetrics: Part B*. The Netherlands.
- Melgaard, D. and Haaland, D. (2004). Effects of nonlinearities and uncorrelated or correlated errors in realistic simulated data on the prediction abilities of augmented classical least squares and partial least squares. *Applied Spectroscopy*, **58**, 1065–1073.
- Naes, T., Isaksson, T., Fearn, T., and Davies, T. (2002). *A User-Friendly Guide to Multivariate Calibration and Classification*. NIR Publications, Chichester, UK.
- Osborne, C. (1991). Statistical calibration: a review. *International Statistical Review*, **59**, 309–336.
- Preda, C. (2007). Regression models for functional data by reproducing kernel Hilbert spaces methods. *Journal of Statistical Planning and Inference*, **137**, 829–840.
- Ramsay, J. and Dalzell, C. (1991). Some tools for functional data analysis (with discussion). *Journal of the Royal Statistical Society B*, **53**, 539–572.
- Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer New York.

- Rao, B. (1983). *NonParametric Functional Estimation*. Academic Press, Orlando, Florida, US.
- Rao, P. (2010). *Statistical Inference for Fractional Diffusion Processes*. Wiley.
- Rossi, F. and Conan-Guez, B. (2005). Functional multi-layer perceptron: a nonlinear tool for functional data analysis. *Neural Networks*. **18**, 45–60.
- Rossi, F., Delannay, N., Conan-Guez, B., and Verleysen, M. (2005). Representation of functional data in neural networks. *Neurocomputing*. **64**, 183–210.
- Walters, F. and Rizzuto, G. (1988). The calibration problem in statistics and its application to chemistry. *Analytical Letters*. **21**, 2069–2076.
- Yao, A. (2001). *Un modèle semi-paramétrique pour variables fonctionnelles : la régression inverse fonctionnelle*. PhD thesis, Université Toulouse III, France.
- Zhou, Y. and Cao, H. (2013). An augmented classical least squares method for quantitative raman spectral analysis against component information loss. *The Scientific World Journal*, 306937.

Advanced Technologies Application Center, CENATAV, Havana, Cuba

E-mail: (nhernandez@cenatav.co.cu,italavera@cenatav.co.cu)

Centro de Investigacin en Matemáticas, Guanajuato, Mexico

E-mail: (rolando.biscay@cimat.mx)

SAMM, Université Paris 1 & INRA, UR875 MIA-T, Castanet-Tolosan, France

E-mail: (nathalie.villa@toulouse.inra.fr)