Statistica Sinica Preprint No: SS-13-242R2	
Title	A non parametric approach for calibration with functional data
Manuscript ID	SS-13-242R2
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.2013.242
Complete List of Authors	Noslen Hernandez, Rolando J. Biscay, Nathalie Villa-Vialaneix and Isneri Talavera
Corresponding Author	Nathalie Villa-Vialaneix
E-mail	nathalie.villa@toulouse.inra.fr
Notice: Accepted version subject to English editing.	

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

A NON PARAMETRIC APPROACH FOR CALIBRATION WITH FUNCTIONAL DATA

Noslen Hernández[1], Rolando J. Biscay[2],
Nathalie Villa-Vialaneix[3,4] and Isneri Talavera[1]

[1] *CENATAV, Havana - Cuba*
[2] *CIMFAV, Universidad de Valparaíso - Chile*
[3] *SAMM, Université Paris 1 - France*
and [4] *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 its application to both simulated and real data.

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

1. Introduction

Statistical calibration plays a crucial role in many areas of technology such 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 $(x_1, y_1), ..., (x_n, y_n)$ 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.

The practical motivation that leads us to study the above problem 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, such 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, it is desired to obtain 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, depending on *a*) the space in which the random variable X takes values: it can be the real line (univariate calibration), a finite-dimensional space (multivariate calibration) or a functional space (functional calibration); *b*) the kind of experimental design: fixed design (the Y values are not random but set by the researcher) or random design (Y is a random variable as well as X); and *c*) the linear or nonlinear nature of the assumed statistical model $f(X|Y)$. A review of the literature on this subject, for both univariate and multivariate calibration, can be found in Osborne (1991).

Furthermore, this paper is concerned with the setting of 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 the assumption that X lies in a functional space, such as L_2. Also, the focus is put on the case of 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 for linear models under fixed design).

A widely used criterion for calibration in case of random design is the 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 has functional nature, i.e., 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 et al., 1999; James and Hastie, 2001) (see also 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), among others.

However, 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. On the contrary, their probabilistic assumptions are made about the densities $f(Y|X)$ and $f(X)$. Notice that, in the calibration framework which is the focus of our interest, it is natural that the probabilistic assumptions refer to this model, since the model $f(X|Y)$ represents the data generation mechanism. In this, the actual response variable is X, and Y is the explanatory variable. Hence, contrary to standard prediction in regression problems, in the calibration setup the variable to be predicted, Y, is not the response variable of the assumed data generation model, $f(X|Y)$, but its explanatory variable. 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, so taking into consideration the aforementioned 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 consistency of the DBIC method, as well as to further assess its numerical performance in a more elaborate simulation setting. The paper is organized as follows. Section 2 elaborates the method and proves its consistency. The proofs require to bring together theoretic results from nonparametric statistics, Functional Data Analysis (FDA) and equivalence of Gaussian measures, whose details are deferred to the Appendixes. Section 3 shows the performance of the functional DBIC approach and provides comparison with functional kernel regression on a simulation study. Section 4 illustrates the method on a real-world benchmark data set.

2. Functional Density-Based Inverse Calibration

Presentation of the method

Let (X, Y) be a pair of random variables taking values in $\mathcal{X} \times \mathbb{R}$, where $(\mathcal{X}, \langle ., . \rangle)$ is the space of square integrable functions from $[a, b]$ to $\mathbb{R}$ (i.e., $\mathcal{X} = L_2([a, b])$). Suppose also that 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, from $(x_i, y_i)_i$, 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 $\gamma(x) = E(Y|X = x)$.

In this paper, a new functional calibration method to estimate $\gamma(X)$ is introduced, which relies on assuming the following regression model:

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

where e is a random process (perturbation or noise), independent of Y, and r is a function from $\mathbb{R}$ into $\mathcal{X}$. Its main practical motivation arises from 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 said model (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). Hence, stating assumptions for this model is more natural than for the inverse distribution $f(Y|X)$, in this specific applied setting. In particular, a simple instance of such 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 to 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 et al. (1985); Melgaard and Haaland (2004) for the importance of nonlinear effects in spectroscopy data).

We will assume here that the perturbation e in model (2.1) follows a Gaussian distribution P_0 with zero mean and covariance Γ. This is a common assumption in the context of calibration in several applications such as spectroscopy (as explained above). In these applications, $y \to r(y)$ 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, which result in statistical procedures that involve only the first two moments of the variables (i.e., means and covariances, see, e.g., Martens and Naes (1989); Kriesten et al. (2008a)). Also note that 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 further supports that stating probability assumptions in terms of the hard model $f(X|Y)$ is not only physically more meaningful but also easier than for the inverse model $f(Y|X)$ in this peculiar applied settings.

Under this Gaussian distribution assumption, the conditional distribution

$P(\cdot|y)$ is also a Gaussian distribution and is fully determined by its corresponding mean function $r(\cdot) = E(X|Y = \cdot)$, and its covariance operator Γ (not depending on y), which is a symmetric and positive Hilbert-Schmidt operator on the space $\mathcal{X}$. Thus, there exists an eigenvalue decomposition of Γ, $(\varphi_j, \lambda_j)_{j\geq 1}$ 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 $\varphi_j \otimes \varphi_j : h \in \mathcal{X} \to \langle \varphi_j, h\rangle \varphi_j$.

Suppose that the following usual regularity condition holds (Grenander, 1981, p. 271): 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 $j \geq 1$.

Under these assumptions, and the one that 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 $f_X(x) = \int_{\mathbb{R}} f(x|y) f_Y(y)\, dy$.

Hence, given an estimate $\hat{f}(x|y)$ of $f(x|y)$, this suggests the following (plug-in) estimate of $\gamma(x)$:

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

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 $\hat{f}_X(x)$ is defined by $\hat{f}_X(x) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(x|y_i)$ and 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 may be carried out through any standard nonparametric method for univariate regression based on the data set $(y_i, x_i(t))_{i=1,\ldots,n}$. Here a smoothing kernel method, specifically the Nadaraya-Watson kernel estimate of r,

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

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 $(\widehat{\varphi}_j, \hat{\lambda}_j)_j$ of the eigenfunctions and eigenvalues $(\varphi_j, \lambda_j)_j$ of the covariance Γ on the basis of the empirical covariance $\widehat{\Gamma}$ of the residuals $\hat{e}_i = x_i - \hat{r}(y_i)$, that is, $\widehat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n} \hat{e}_i \otimes \hat{e}_i$. Only the first p eigenvalues and eigenfunctions are incorporated, where $p = p(n)$ is an integer smaller than n. Note that this is a standard functional PCA problem.

3. Estimate $f(x|y)$ by

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

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

Once having $\hat{f}(x|y)$, substituting (2.4) into (2.2) leads to an estimate $\hat{\gamma}(x)$ of $\gamma(x)$, which will be referred to as 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 would have reduced to the approach for univariate calibration described in Lwin and Maritz (1980).

Beyond Gaussianity

The DBIC estimate has been explicitly elaborated under the assumption that $f(X|Y)$ is a Gaussian distribution with covariance not depending on y. However, note that this approach is general enough to allows for the extension of DBIC to non-Gaussian distributions. For instance, the method can be extended straightforwardly to the more general setting in which the data x is a diffusion process generated by a 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 $b : \mathbb{R} \times [0,1] \to \mathbb{R}_+$ is a given function. Indeed, under mild conditions (e.g., if the function b is bounded away from zero and infinite, see Liptser and Shiryaev (1977)) the measure $P(\cdot|y)$ of the solution x has a density with respect to the measure $P_0 = P(\cdot|0)$, which is given by $f(x|y) = \exp\left\{\int_0^1 \frac{\dot{r}(y)(t)}{b^2(x(t),t)} dx(t) - \frac{1}{2}\int_0^1 \left|\frac{\dot{r}(y)(t)}{b(x(t),t)}\right|^2 dt\right\}$. 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 (i.e., in case of equations driven by multiplicative noise) 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 constituted by those generated by stochastic differential equations driven by an additive fractional Brownian motion $W^H(t)$. More 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 details).

For these kinds of non Gaussian functional data, DBIC estimators may be carried out through the steps 1)-3). They would only differ in their specific implementation of the approximation $\widehat{f}(x|y)$ that depends on the numerical computation of the stochastic integrals involved. Such straightforward extensions of the DBIC approach are beyond the scope of the present work, which focuses on Gaussian models motivated by spectroscopy data. However, they are worth of future studies and could be of interest in other applied fields.

Consistency

Hereafter, asymptotic properties for the estimators proposed on steps (1)-(3) are given as well as a consistency result for the DBIC estimator $\hat{\gamma}(x)$. Their proofs are given in the Appendixes (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 as well as a rate of convergence for $\hat{r}(y)$ can be obtained under the following 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 Ω_Y;

  (A2) K is an order k kernel with compact support;

  (A3) there exists d_1 and d_2 such that $\sup_{y\in\Omega_Y} \left|f_Y^{(k)}(y)\right| < d_1$ and $\sup_{y\in\Omega_Y} \left\|r^{(k)}(y)\right\| < d_2$;

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

Several remarks about this proposition and its underlying assumptions can be done. First, note that (A1)-(A3) are standard regularity assumptions in the framework of kernel-based density estimation (Rao, 1983). Second, (A5) is satisfied in most calibration settings that are the motivation of the present work; but it can be shown that, 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 δ does so. Third, also note that the estimation of $r(y)$ by a Nadaraya-Watson kernel estimate is not mandatory and that this step (and the corresponding assumptions) could be replaced by any other (1-dimensional nonparametric) 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, Γ, based on the estimated residuals. The consistency of this estimate with $\sqrt{n}$-rate will ensure 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, the following additional assumption is required:

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

Note that this assumption, which serves as a basis for the DBIC method, implies the condition usually assumed on moments: $E\left(\|e\|^4\right) < +\infty$. The following proposition can then be proved.

Proposition 2. *Under assumptions (A1)-(A7) we have:* $\left\|\hat{\Gamma} - \Gamma\right\| = O_P\left(\frac{1}{n^{1/2-2c_1}}\right)$, *where* $\|.\|$ *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 $\hat{\Gamma}$. As previously explained, this result is derived from a Theorem given 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) $\sum_{j=1}^{\infty} \sup_{y\in\Omega_Y} \frac{|r_j(y)|}{\sqrt{\lambda_j}} < \infty$;

(A9) The $(\lambda_j)_j$ are all distinct;

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

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

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

Note that since $(\lambda_j)_j$ is a sequence of positive real number decreasing to 0 when j tends to $+\infty$, Assumption (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 $\sup_{y\in\Omega_Y} \|r(y)\|^2 = \sup_{y\in\Omega_Y} \sum_{j=1}^{\infty} r_j^2(y) \leq \sum_{j=1}^{\infty} \sup_{y\in\Omega_Y} r_j^2(y) < \infty$, Assumption (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). Due to the previous remark, Assumption (A6) is no longer required in this proposition.

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

From the last proposition, the consistency of $\hat{\gamma}(x)$, defined as in Equation (2.2), can be proved. The final Theorem is demonstrated in Appendix :

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

From Theorem 1 and the Lebesgue's dominated convergence theorem, proving that $E\left(\hat{\gamma}(x) - \gamma(x)\right)^2 \to_n 0$ is straightforward. That is, the DBIC estimator also converges in the sense of the quadratic Bayesian risk.

3. A simulation study

In this section, the feasibility and the performance of the nonparametric functional calibration method described in Section 2 is discussed through a sim-

ulation study. In the calibration setting, which is the motivation of the present work as discussed in previous sections, the physical data generation mechanism is specified by the density $f(X|Y)$. Here, training data are generated under various statistical models for this density. Based on this training data set, the DBIC estimator is computed to predict Y values corresponding to new values of the variable X (test data set).

The data were simulated in the following way: values for the real random variable Y were drawn from a uniform distribution in the interval $[0, 10]$. e is a Gaussian process independent of Y with zero mean and covariance operator $\Gamma = \sum_{j\geq 1} \frac{1}{j^{(1+0.1)}} v_j \otimes v_j$, where $(v_i)_{i\geq 1}$ is the trigonometric basis of $\mathcal{X} = L_2([0, 1])$ (i.e., $v_{2k-1} = \sqrt{2}\cos(2\pi kt)$, and $v_{2k} = \sqrt{2}\sin(2\pi kt)$). More precisely, for all models, e was simulated by using a truncation of Γ, $\Gamma(s, t) \simeq \sum_{j=1}^{q} \frac{1}{j^{(1+0.1)}} v_j(t)v_j(s)$ by setting $q = 500$. Then, X was generated by four different models or settings including linear and nonlinear ones.

M1 a model where $E(X|Y)$ is a linear function of Y expressed on the error eigenfunction basis: $X = Yv_1 + Yv_2 + Yv_5 + Yv_{10} + e$;

M2 a model where $E(X|Y)$ is a nonlinear function of Y expressed on the error eigenfunction basis: $X = \sin(Y)v_1 + \log(Y + 1)v_5 + e$;

M3 a model where $E(X|Y)$ is a linear function of Y expressed not on the error eigenfunction basis but on polynomials: $X = Yq_1 + 5Yq_2 + e$, where $q_1 = 2t^3$ and $q_2 = t^4$. Note that such polynomials have coefficients in the Fourier basis that decay faster than $\frac{1}{j^3}$, and so assumption (A8) is fulfilled;

M4 a model where $E(X|Y)$ is a nonlinear function of Y expressed using the aforementioned polynomials: $X = \sin(Y)q_1 + 20\log(Y + 1)q_2 + e$.

From these 4 models, a training and a test samples with sizes $n_L = 300$ and $n_T = 200$, respectively, were generated. To apply the DBIC method, simulated discretized functions were approximated by continuous functions using a functional basis expansion. Specifically, the discrete data were converted into

All the simulations were done using Matlab© and the DBIC method was also implemented for Matlab©. Parts of the implementation use the Matlab© 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.

continuous data (or functional predictors) X by approximation through 128 B-spline basis functions of order 4.

The DBIC method was used according to the 3 steps described in Section 2. For the first step, the conditional mean $r(y)$ was estimated from the training sample by kernel smoothing (such as in Equation (2.3)). For this, it was necessary to tune the bandwidth parameter h. This 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. That is, the training sample $(x_1, y_1), ..., (x_{n_L}, y_{n_L})$ was randomly partitioned into 10 blocks or folds of approximately the same size, and $h_{opt} = \arg\min_{h \in H} \frac{1}{n_L} \sum_{i=1}^{n_L} \|\hat{x}_i^{(h)} - x_i\|_{L_2}^2$ where H is the search interval for possible values of h, and $\hat{x}_i^{(h)}$ is the estimate of the mean $r(y_i)$ using a kernel smoothing with parameter h and the data not belonging to the fold in which (x_i, y_i) is.

Another hyperparameter involved in the estimation of the regression function $\hat{\gamma}(x)$ is the number p of eigenfunctions (Equation (2.4)) used to estimate $f(x|y)$. This hyperparameter was selected also by a 10-fold cross-validation for minimizing the root mean squared error (RMSE) criterion on the training sample. Specifically, $p_{opt} = \arg\min_p \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\hat{y}_i^{(p)} - y_i\right)^2}$, where $\hat{y}_i^{(p)}$ is DBIC prediction of y_i using the conditional density $\hat{f}^{(p)}$ calculated with p eigenfunctions and the data not belonging to the fold in which y_i is, fold(i). That is, $\hat{y}_i^{(p)} = \frac{\sum_{j \notin \text{fold}(i)} \hat{f}^{(p)}(x_i|y_j) y_j}{\sum_{j \notin \text{fold}(i)} \hat{f}^{(p)}(x_i|y_j)}$.

For model M1 the cross-validation gives the value $p = 15$, which is close to the true one ($p = 10$) according to the model. For model M4 the resulting value was $p = 47$, which is larger. Unlike M1, M4 was not built by using the first eigenfunctions of the covariance operator Γ in the expression of $E(X|Y)$, hence the need for more eigenfunctions.

Once the estimate $\hat{\gamma}(x)$ is obtained on the basis of the training set, the performance of the DBIC approach was assessed by predicting the y values on the test sample. More precisely, the RMSE was computed: 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 $\hat{\gamma}(x_i)$. In order to have a reference to compare with, the standard functional nonparametric kernel estimate (NWK)

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

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

(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 RMSE and the NWK RMSE for each of the simulated models. It can be observed that DBIC performs well in all models and outperforms the NWK estimator. The fourth column in the table is the RMSE achieved by DBIC but using a parametric estimation of the mean: instead of estimating the mean using kernel smoothing, the mean was estimated by linear regression (least squares estimates) for models M1 and M3 in which the means are linear functions of Y. It can be observed that the RMSE resulting from such a parametric estimates are smaller that 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 (with several graphics showing the different estimation results) is provided in supplemental material.

4. A study of Tecator dataset

DBIC was also tested on a benchmark functional data set: the Tecator dataset (Borggaard and Thodberg, 1992). It consists of spectrometric data from the food industry. Each of the 215 observations is the near infrared absorbency spectrum of a meat sample recorded on a Tecator Infratec Food and Feed Analyzer. Each spectrum is sampled at 100 wavelengths uniformly spaced in the range 850–1050 nm. The composition of each meat sample is determined by analytic chemistry, so percentages of moisture, fat and protein are associated in this way to each spectrum. We will focus on predicting the percentage of fat on

the basis of the absorbency spectrum. This problem is more challenging than the ones presented in Section 3 where the data were generated to fulfill exactly the conditions of the DBIC method. The whole 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, both for 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 on Tecator by DBIC are remarkably better than the ones by NWK in the sense of MSE. In Ferraty and Vieu (2006), results based on the use of a semi-metric involving the second order derivatives (which is known to be useful for this data set) were reported. But even incorporating this information in the model, a MSE of 3.5 is obtained, which is still larger than the one obtained by using DBIC without derivative information.

5. Conclusion

A new functional nonparametric calibration approach has been introduced motivated by the calibration problem in spectrometrics. The new method, named functional Density-Based Inverse Calibration (DBIC), was fully described for the sample space $\mathcal{X} = L_2([a, b])$ under a Gaussian assumption for the conditional law $P(\cdot|Y)$ but it can be extended to other sample spaces and distribution families. Two appealing features of the new method are its rather mild model assumptions and its computational simplicity. Furthermore, it allows us to incorporate parametric information on the conditional mean $E(X|Y)$ of the "inverse" model if available. The DBIC method is consistent under quite general assumptions, and the simulation study has shown that it performs well for both linear and nonlinear models. Thus, 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. However it would be interesting to obtain a limit distribution for the estimate in order to derive confidence bounds.

6. Appendix

In the following appendixes, the additional notations defined below will be used:

- m is the function defined on Ω_Y such that: $r(y) = \frac{m(y)}{f_Y(y)}$;
- $g(x) = \int_{\mathbb{R}} f(x|y) y f_Y(y) dy$ and $\hat{g}(x) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(x|y_i) y_i$;

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 assumptions (A1)-(A3), we have:* $\sup_{y\in\Omega_Y} \left|\hat{f}_Y(y) - f_Y(y)\right| = O_P\left(h^k + \frac{\sqrt{\log n}}{\sqrt{nh}}\right)$.

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

Proof of Proposition 1:

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

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

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

Proof of Proposition 2

The proof of Proposition 2 also requires the use of an additional lemma whose proof can be found in the cited article.

Lemma 3. *[Cardot et al. (1999)] If Z is a random variable in a Hilbert space with covariance operator Γ_Z and $E\left(\|Z\|^4\right) < +\infty$ then $E\left(\|\Gamma_Z - \Gamma_Z^n\|^2\right) \leq \frac{E\left(\|Z\|^4\right)}{n}$, where $\Gamma_Z^n = \frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z}) \otimes (Z_i - \overline{Z})$, 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:

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

which 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 operators of T_1. Then,

$$\left\|\Gamma-\hat{\Gamma}\right\| \leq \|\Gamma-\Gamma_n\| + 2\|T_1\| + \|T_2\|. \tag{6.1}$$

Each part of the right term of this inequality is addressed separately:

- From Assumption (A7) and Lemma 3 we obtain directly:

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

- 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 Cauchy-Schwartz and Markov inequalities, we

have:

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

As $E\|e\|^2 < +\infty$ and by Proposition 1, $\sup_{y\in\Omega_Y}\|r(y)-\hat{r}(y)\| = O_P\left(n^{-c_1k}+\left(\frac{\log n}{n^{1-2c_1}}\right)^{1/2}\right)$, 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 Assumption (A4). Hence, $\lim_{n\to+\infty}P\left(n^{1/2-2c_1}\|T_1\|>c\right)=0$ and

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

- By definition of T_2, we have:

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

which, by Proposition 1, gives $\sqrt{n}\|T_2\| = O_P\left(\left(n^{-c_1k+1/4}+\left(\frac{\log n}{n^{1/2-2c_1}}\right)^{1/2}\right)^2\right)$, but by, Assumption (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

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

Putting the results reached in Equations (6.2), (6.3) and (6.4) into Equation (6.1), the conclusion holds: $\left\|\Gamma-\hat{\Gamma}\right\| = O_P\left(\frac{1}{n^{1/2-2c_1}}\right)$.□

Proof of Proposition 3

The proof of Proposition 3 requires the use of the following lemma whose proof can be carried out in a similar way to the one given in Bosq (1991) for the

particular case in which $\tilde{\Delta}$ is an empirical covariance operator associated with the covariance operator Δ.

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

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

ii) $\left\|\phi_j - \tilde{\phi}_j\right\| \leq a_j \left\|\Delta - \tilde{\Delta}\right\|$ *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 $y \in \Omega_Y$, denote $\mathcal{E}(y)$ the difference

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

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

Convergence of $\mathcal{E}_1$:

The Karhunen-Loeve expansion of x has coordinates $x_j = r_j(y) + \sqrt{\lambda_j}\xi_j$, where $\xi_j \sim \mathcal{N}(0,1)$ and independent. Then,

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

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

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

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

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

Finally, putting Equations (6.6) and (6.7) into (6.5) leads to $P_0-a.s.,\quad \sup_{y\in\Omega_Y}\mathcal{E}_1(y)\xrightarrow{p\to+\infty}0$.

Convergence of $\mathcal{E}_2$:

$\mathcal{E}_2$ can be divided into 4 parts: $\mathcal{E}_2(y)\le A(y)+B(y)+C(y)+D(y)$ where

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

Convergence of A

From $\sum_j r_j^2(y)=\|r(y)\|^2$ it follows that $|r_j(y)|\le\|r(y)\|$ for all j. Putting together this inequality and the facts that $\lambda_j\ge\lambda_p$ for $j=1,\dots,p$ and $|x_j-\hat{x}_j|=|\langle x,\varphi_j-\widehat{\varphi}_j\rangle|\le\|x\|\|\varphi_j-\widehat{\varphi}_j\|$ we obtain, for any $y\in\Omega_Y$,

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

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

Convergence of B

By the same arguments as those used for A, we have $B(y)\le\frac{\|r(y)\|}{2\lambda_p}\sum_{j=1}^{p}|r_j(y)-\hat{r}_j(y)|$. Moreover, we have that, for any j and any $y\in\Omega_Y$,

$|r_j(y) - \hat{r}_j(y)| \leq \|r(y)\|\|\varphi_j - \widehat{\varphi}_j\| + \|\widehat{\varphi}_j\|\|r(y) - \hat{r}(y)\|$, Thus, applying Lemma 4 and Proposition 1 we obtain, for any j,

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

Then,

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

where the first term is $o_P(1)$ due to Assumption (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 Assumption (A12) (because, taking into consideration that λ_j is a decreasing sequence, (A12) implicates 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 $o_P(1)$. Then, $\sup_{y\in\Omega_Y} B(y) = o_P(1)$.

Convergence of C

From $|\hat{x}_j| \leq \|x\|$ and $|\hat{r}_j(y)| \leq \|\hat{r}(y)\|$, we have that $C(y) \leq \frac{\|x\| + \|\hat{r}(y)\|}{\lambda_p}\sum_{j=1}^p |r_j(y) - \hat{r}_j(y)|$. Thus, $\sup_{y\in\Omega_Y} C(y) \leq \left(\|x\| + \sup_{y\in\Omega_Y}\|\hat{r}(y)\|\right)\frac{1}{\lambda_p}\sup_{y\in\Omega_Y}\sum_{j=1}^p |r_j(y) - \hat{r}_j(y)|$. As for 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 Assumption (A8). Putting all this together leads to $\sup_{y\in\Omega_Y} C(y) = o_P(1)$.

Convergence of D

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 $\hat{\lambda}_p \geq \left|\lambda_p - |\lambda_p - \hat{\lambda}_p|\right|$ we have that $\frac{p\|\Gamma-\hat{\Gamma}\|}{\lambda_p\hat{\lambda}_p} \leq \frac{p\|\Gamma-\hat{\Gamma}\|}{\left|\lambda_p^2 - \lambda_p|\lambda_p - \hat{\lambda}_p|\right|} =$

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

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

in which

- the first term is $o_P(1)$ by Assumptions (A4) and (A12) (taking into consideration that the last one implies $q - 1/2 < 0$);
- the second term and the 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 Assumptions (A11) and (A12).

Hence, we finally obtain: $\sup_{y\in\Omega_Y} D(y) \leq \left(\left(\sup_{y\in\Omega_Y}\|\hat{r}(y)\|\right)\|x\| + \frac{1}{2}\left(\sup_{y\in\Omega_Y}\|\hat{r}(y)\|\right)^2\right)o_P(1)$. 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 $\sup_{y\in\Omega_Y} D(y) = o_P(1)$.

Conclusion

Hence, $\sup_{y\in\Omega_Y}\mathcal{E}(y) = o_P(1)$. 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 Assumption (A8) (see proof of Proposition 3). The right hand side of the last inequality goes to zero as n increases, which concludes the proof.□

Proof of Theorem 1

Before proving Theorem 1, the following proposition is demonstrated:

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

Proof of Proposition 4:

For any $x \in \mathcal{X}$, $\left|\hat{f}_X(x) - f_X(x)\right| \leq \left|\hat{f}_X(x) - \frac{1}{n}\sum_{i=1}^n f(x|y_i)\right| + \left|\frac{1}{n}\sum_{i=1}^n f(x|y_i) - f_X(x)\right|$. Furthermore,

- by Proposition 3, $\left|\hat{f}_X(x) - \frac{1}{n}\sum_{i=1}^n f(x|y_i)\right| \leq \sup_{y\in\Omega_Y}\left|f(x|y) - \hat{f}(x|y)\right| = o_P(1)$;

- Assumption (A8) ensures that, for all $x \in \mathcal{X}$, $f_X(x)$ is finite. Hence, by the law of large numbers, $\lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^n f(x|y_i) =^{as} E_Y\left(f(x|Y)\right) = f_X(x)$.

These two arguments complete the first part of the proof.

The second part is demonstrated in a similar way: for any $x \in \mathcal{X}$,

$$\left|\frac{1}{n}\sum_{i=1}^n \hat{f}(x|y_i)y_i - \int_{\mathbb{R}} f(x|y)yf_Y(y)dy\right| \leq \left|\frac{1}{n}\sum_{i=1}^n \hat{f}(x|y_i)y_i - \frac{1}{n}\sum_{i=1}^n f(x|y_i)y_i\right| + \left|\frac{1}{n}\sum_{i=1}^n f(x|y_i)y_i - \int_{\mathbb{R}} f(x|y)yf_Y(y)dy\right|,$$

where

- the first part of the right hand side of this inequality is bounded by $\sup_{y\in\Omega_Y}\left|f(x|y) - \hat{f}(x|y)\right| \times \frac{1}{n}\sum_i y_i$. As $E(Y) < \infty$, we have that $\lim_{n\to+\infty}\frac{1}{n}\sum_i y_i =^{as} E(Y)$. By Proposition 3, $\sup_{y\in\Omega_Y}\left|f(x|y) - \hat{f}(x|y)\right| = o_P(1)$. Thus, $\left|\frac{1}{n}\sum_{i=1}^n \hat{f}(x|y_i)y_i - \frac{1}{n}\sum_{i=1}^n f(x|y_i)y_i\right| = o_P(1)$.

- the second part of the right hand side of the previous inequality converges to 0 almost surely by the law of large numbers under the fact that $E(Y) < \infty$.□

Proof of Theorem 1:

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

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

Furthermore,

$$\begin{aligned} |\hat{g}(x)| &\leq ||\hat{g}(x) - g(x)| + g(x)|, \\ \hat{f}_X(x) &\geq \left| f_X(x) - |\hat{f}_X(x) - f_X(x)| \right|. \end{aligned}$$

Thus,

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

Since $f_X(x)$ is finite (Assumption (A8)) and positive, the first term in the right hand side of the above inequality is $o_P(1)$ by Proposition 4. Also 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. □

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*, volume 335 of *ASI Series*, pages 509–529. NATO.

Boulet, J. and Roger, J. (2010). Improvement of direct calibration in spectroscopy. *Analytica Chimica Acta*, **66**(2), 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**(5); 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**(6), 475–488.

Ferré, L. and Yao, A. (2005). Smoothed functional inverse regression. *Statistica Sinica*, **15**(3), 665–683.

Geladi, P., MacDougall, D., and Martens, H. (1985). Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy*, **39**(3), 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, CIARP 2007, Valparaiso, Chile, November 13-16, 2007)*, volume 4756 of *Lecture Notes in Computer Science*, chapter Support vector regression methods for functional data, pages 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**(4), 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**(2), 146–159.

Hernández, N., Biscay, R., Villa-Vialaneix, N., and Talavera-Bustamante, I. (2010). A functional density-based nonparametric approach for statistical calibration. In Bloch, I. and Cesar, R., editors, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications - 15th Iberoamerican Congress on Pattern Recognition (CIARP 2010)*, volume 6419 of *Lecture Notes in Computer Science*, pages 450–457, Sao Paulo, Brazil. Springer.

James, G. and Hastie, T. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society, Series 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**(2), 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**(2), 108–119.

Laloë, T. (2008). A k-nearest neighbor approach for functional regression. *Statistics and Probability Letters*, **78**(10), 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 Series 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(9)**, 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**(3), 829–840.

Ramsay, J. and Dalzell, C. (1991). Some tools for functional data analysis (with discussion). *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, **53**, 539–572.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. New York, 2nd edition edition.

Rao, B. (1983). *NonParametric Functional Estimation*. Academic Press, Orlando, Florida, US.

Rao, P. (2010). *Statistical Inference for Fractional Diffusion Processes*. John Wiley & Sons.

Rossi, F. and Conan-Guez, B. (2005). Functional multi-layer perceptron: a nonlinear tool for functional data anlysis. *Neural Networks*, **18**(1), 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-paramérique 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)

CIMFAV, Universidad de Valparaíso, Valparaíso, Chile

E-mail: (rolando.biscay@uv.cl)

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

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