Focusing on structural assumptions in regression on functional variable.

DELSOL, Laurent
Université d'Orléans, MAPMO
Bâtiment de mathématiques - Rue de Chartres
B.P. 6759 -45067 Orléans cedex 2 FRANCE
E-mail: laurent.delsol@univ-orleans.fr

FERRATY, Frédéric
Université Paul Sabatier, Institut de Mathématiques de Toulouse.
118 route de Narbonne, 31062 TOULOUSE Cedex 9 FRANCE E-mail: ferraty@cict.fr

VIEU, Philippe
Université Paul Sabatier, Institut de Mathématiques de Toulouse.
118 route de Narbonne, 31062 TOULOUSE Cedex 9 FRANCE E-mail: vieu@cict.fr

Abstract

Functional statistics aims to provide relevant tools to deal with samples of curves (or more generally infinite dimensional variables). This field of modern statistics presents theoretical challenges and opens a wide scope of potential real world applications. Various well known multivariate models and methods have been extended to take into account the infinite dimension of the variables. A lot of consideration has been given to the study of the regression of a real-valued response on a functional variable. Many estimation procedures, often based on structural assumptions on the regression operator, have been considered. This work concerns an other type of issue: the construction of structural testing procedures in such regression models. Structural tests may be relevant by themselves to check the validity of some a priori model or some theoretical assumptions. They are also relevant complementary tools to estimation methods to check the validity of some structural assumptions used to construct the estimator or test if some heuristic hypothesis coming from an estimation of the regression operator holds. However, the literature devoted to such tests is reduced to a small number of papers. A theoretical background has been recently proposed to consider a wide scope of structural tests. The aim of this talk is to describe the practical use of some of these tests, discuss some recent improvements and talk about interesting prospects. The test statistic is computed from an estimator specific to the null (structural) hypothesis and uses recent advances in kernel smoothing for functional data. Then, several bootstrap methods are proposed to compute the threshold value. Simulation studies and applications are finally presented to illustrate the interest of the proposed testing procedures.

Keywords: regression, functional variable, structural test, bootstrap, spectrometry

1. Introduction

A great variety of real world issues involve functional phenomena which may be represented as curves or more complex objects. They may for instance come from the observation of phenomenon over time or more generally its evolution when the context of the study changes (e.g. growth curves, sound records, spectrometric curves, electrocardiograms, images). It is nowadays common to deal with a large amount of discretized observations of a given functional phenomenon that actually gives a relevant

understanding its dynamic and regularity. Classical multivariate statistical tools may be irrelevant in that context to take benefit from the underlying functional structure of these observations.
Recent advances in functional statistics offer a large panel of alternative methods to deal with functional variables (i.e. variables taking values in an infinite dimensional space) which become popular in real world studies. A general overview on functional statistics may be fond in Ramsay and Silverman (1997, 2002, 2005), Bosq (2000), Ferraty and Vieu (2006), and more recently Ferraty and Romain (2010). This talk focuses on the study of regression models involving a functional covariate:

$$Y = r(\mathcal{X}) + \epsilon,$$

where Y is a real valued random variable, $\mathcal{X}$ is a random variable taking values in a semi-metric space $(\mathcal{E}, d)$ and $\mathbb{E}[\epsilon|\mathcal{X}] = 0$.
A lot of work has already been done on the estimation of the regression operator r through various versions of this model corresponding to structural assumptions on r. The most famous example is certainly the functional linear model introduced by Ramsay and Dalzell (1991):

$$Y = \alpha_0 + < \alpha, \mathcal{X} >_{\mathbb{L}^2([0;1])} + \epsilon, \ (\alpha_0, \alpha) \in \mathbb{R} \times \mathbb{L}^2([0;1]).$$

This model has received a lot of attention and is still a topical issue what is illustrated through the contributions of Cardot *et al.* (1999,2000,2007), Ramsay and Silverman (1997, 2005), Preda and Saporta (2005), Hall and Cai (2006), Crambes *et al.* (2009), or Ferraty and Romain (2010, Chapter 2) among others.
Several other examples of models based on a given structure of r have been considered as appears in the work of Sood *et al.* (2009) on a multivariate additive model based on the first coefficients of a functional P.C.A., Ait Saidi *et al.* (2008) about the functional single index model, or Aneiros-Perez and Vieu (2009) on the partial linear model. And it is likely other "structural modelisations" will be considered in the future (functional additive models, partially functional models, ...).
On the other hand, nonparametric models in which only the regularity (Hölder) of r with respect to the semi-metric d is assumed, have been considered by Ferraty and Vieu (2000). Many references on recent contributions on this topic are given in Ferraty *et al.* (2002), Masry (2005), Ferraty and Vieu (2006), Delsol (2007,2009) together with Ferraty and Romain (2011, Chapters 1, 4, and 5).

2. Structural tests

2.1 A general way to construct a test statistic

As discussed in the previous paragraph, a lot of work as been done on the estimation of the regression operator r. This work focuses on a different issue and proposes statistical tools for the construction of testing procedures allowing to check if r has a given structure (e.g. constant, linear, multivariate, . . .). Such testing procedures are interesting by themselves to test the validity of an a priori assumption on the structure of the regression model. They are also complementary tools to estimation methods. They may be used as a preliminary step to check the validity of the structural assumption used to construct an estimator and may be relevant to test some structural assumption made from the estimation of r. To the best of our knowledge, the literature on this kind of problem is restricted to Cardot *et al.* (2003,2004), Müller and Stadtmüller (2005) in the specific case of a linear model, Gadiaga and Ignaccolo (2005) on no effect tests based on projection methods, and Chiou and Müller (2007) on an heuristic goodness of fit test. Hence it seems no general theoretical background has been proposed to test the validity of the different modelizations discussed in the introduction part. In the remainder of this note $\mathcal{R}$ stand for a family of square integrable operators and w a weight function. Our aim is to present and discuss in this work a general methodology allowing to test the null hypothesis:

$$\mathcal{H}_0 : \{\exists r_0 \in \mathcal{R}, P(r(\mathcal{X}) = r_0(\mathcal{X})) = 1\}$$

under local alternatives of the form

$$\mathcal{H}_{1,n} : \{ \inf_{r_0 \in \mathcal{R}} \|r_n - r_0\|_{\mathbb{L}^2(wdP_{\mathcal{X}})} \geq \eta_n \}.$$

Extending the ideas of Härdle and Mammen (1993), we construct our test statistic from an estimator $\hat{r}$ adapted to the structural model (corresponding to the null hypothesis, i.e. induced by $\mathcal{R}$) we want to test and functional kernel smoothing tools (K denotes the kernel):

$$T_n = \int (\sum_{i=1}^{n} (Y_i - \hat{r}(\mathcal{X}_i)) K \left(\frac{d(\mathcal{X}_i, x)}{h_n} \right))^2 w(x) dP_{\mathcal{X}}(x).$$

For technical reasons, we assume the estimator $\hat{r}$ is constructed on a sample D_1 independent from $D = (\mathcal{X}, Y_i)_{1 \leq i \leq n}$. A theoretical result in Delsol *et al.* (2011) states under general assumptions the asymptotic normality of T_n under the null hypothesis and its divergence under the local alternatives. This result opens a large scope of potential applications of this kind of test statistic. Here are few examples:

- test of an a priori model: $\mathcal{R} = \{r_0\}$, $\hat{r} = r_0$.
- no effect test: $\mathcal{R} = \{r : \exists C \in \mathbb{R}, r \equiv C\}$, $\hat{r} = \overline{Y}_n$.
- test of a multivariate effect: $\mathcal{R} = \{r : r = g \circ V, V : \mathcal{E} \to \mathbb{R}^p$ known, $g : \mathbb{R}^p \to \mathbb{R}\}$, $\hat{r}$ multivariate kernel estimator constructed from $(Y_i, V(\mathcal{X}_i))_{1 \leq i \leq n}$.
- linearity test: $\mathcal{R} = \{r : r = \alpha_0 + < \alpha, . >, (\alpha_0, \alpha) \in \mathbb{R} \times \mathbb{L}^2[0; 1]\}$, $\hat{r}$ functional spline estimator (see Crambes *et al.* 2009).
- test of a functional single index model: $\mathcal{R} = \{r : r = g(< \alpha, . >), \alpha \in \mathcal{E}, g : \mathbb{R} \to \mathbb{R}\}$, $\hat{r}$ estimator proposed in Ait Saidi *et al.* (2008).

Other situations may also be considered whenever it is possible to provide an estimator $\hat{r}$ satisfying some conditions.

2.2 Bootstrap methods to get the threshold

The practical use of our test statistic requires the computation of the threshold value. One could propose to get it from the asymptotic distribution. However, the estimation of dominant bias and variance terms is not easy, that is why we prefer to use bootstrap procedures. The main idea is to generate, from the original sample, B samples for which the null hypothesis approximately holds. Then, compute on each of these samples the tests statistic and take as threshold the $1 - \alpha$ empirical quantile of the values we have obtained.
We propose the following bootstrap procedure in which steps 2-4 are made separately on samples $D : (\mathcal{X}_i, Y_i)_{1 \leq i \leq n}$ and $D_1 : (\mathcal{X}_i, Y_i)_{n+1 \leq i \leq N}$. In the following lines $\hat{r}_K$ stands for the functional kernel estimator of the regression operator r computed from the whole dataset.

Bootstrap procedure:
Pre-treatment:

1. $\hat{\epsilon}_i = Y_i - \hat{r}_K (X_i)$
2. $\tilde{\epsilon}_i = \hat{\epsilon}_i - \overline{\hat{\epsilon}}$

Repeat B times steps 3-5:

3. Generate residuals (3 different methods NB, SNB or WB)
   - NB • $(\epsilon_i^b)_{1\leq i\leq n}$ drawn with replacement from $(\tilde{\epsilon}_i)_{1\leq i\leq n}$
   - SNB • $(\epsilon_i^b)_{1\leq i\leq n}$ generated from a "smoothed" version $\tilde{F}_n$ of the empirical cumulative distribution function of $(\tilde{\epsilon}_i)_{1\leq i\leq n}$ $(\epsilon_i^b = \tilde{F}_n^{-1}(U_i), U_i \sim \mathcal{U}(0,1))$
   - WB • $(\epsilon_i^b) = \tilde{\epsilon}_i V_i$ where $V_i \sim P_W$ fulfills some moment assumptions: $E[V_i] = 0$, $E\left[V_i^2\right] = 1$ et $E\left[V_i^3\right] = 1$.
4. Generate bootstrap responses "corresponding" to $\mathcal{H}_0$

$$Y_i^b = \hat{r}(X_i) + \epsilon_i^b$$

5. Compute the test statistic T_n^b from the bootstrap sample $(\mathcal{X}_i, Y_i^b)_{1\leq i\leq N}$

Compute the threshold value

6. For a test of level α, take as threshold the $1-\alpha$ quantile of the sample $(T_n^b)_{1\leq b\leq B}$.

Three examples of distributions P_W given in Mammen (1993) are considered. The different methods used to generate bootstrap residuals globally lead to similar results but some of them perform slightly better in terms of level or power. From the results obtained in simulation studies, it seems relevant to use wild bootstrap methods (WB) which lead to more powerful tests and are by nature more robust to the heteroscedasticity of the residuals.

Finally, the integral with respect to $P_{\mathcal{X}}$ which appears in T_n's definition may be approximated by Monte Carlo on a third subsample D_2 independent from D and D_1.

3. Discussion and recent advances

Let us first discuss shortly the impact of the semi-metric d in our testing procedures. Assume d actually take into account only some characteristics (e.g. derivatives, projections, ...) $\tilde{\mathcal{X}}$ of the explanatory curve $\mathcal{X}$. Because of its definition the test statistic T_n only depends on these characteristics. Hence the null and alternative hypothesis considered are actually made on the regression model

$$Y = r_d(\tilde{\mathcal{X}}) + \epsilon_d,$$

with $\mathbb{E}[\epsilon_d|\tilde{\mathcal{X}}] = 0$. Consequently, the use of a semi-metric based on first functional PCA scores will only be able to test assumptions on the regression model corresponding to these first scores and when a semi-metric based on derivatives is used structural assumptions concern the effect of the derivatives. The general method described above is a first attempt in the construction of general structural testing procedures in regression on functional variable (see Delsol, 2008, and Delsol *et al.*, 2011 for a more detailed discussion). The use of these tests on spectrometric data provide relevant informations on the structure of the link between the spectrometric curve and the chemical content of a product. Such tools may be also useful in procedures that aim to extract informative features from the explanatory curve. However it seems relevant to try to improve our approach and propose other test statistics that does not require to split our sample into three subsamples what may cause troubles in practice. To this end, we are now considering the following test statistic:

$$T_{2,n} = \sum_{i\neq j}(Y_i - \hat{r}(\mathcal{X}_i))(Y_j - \hat{r}(\mathcal{X}_j))K\left(\frac{d(\mathcal{X}_i,\mathcal{X}_j)}{h_n}\right)w(\mathcal{X}_i)w(\mathcal{X}_j)$$

The theoretical study of this new test statistic is in progress. However, in the case of no effect tests, it seems $T_{2,n}$ have the same kind of asymptotic properties than T_n. Moreover, the new statistic $T_{2,n}$

Table 1: Empirical rejection probabilities with the third wild bootstrap method, $h = 4$ and $B = 100$

Model	Y^1 and T_2	Y^3 and T_2	Y^6 and T_2
Empirical rejection probability for $\alpha = 0.05$	0.0555	0.9961	1
Empirical rejection probability for $\alpha = 0.01$	0.0141	0.9515	1
Empirical rejection probability for $\alpha = 0,1$	0.1062	0.9986	1
Model	Y^1 and T	Y^3 and T	Y^6 and T
Empirical rejection probability for $\alpha = 0.05$	0.0486	0.5765	0.9894
Empirical rejection probability for $\alpha = 0.01$	0.0085	0.2427	0.8438
Empirical rejection probability for $\alpha = 0.1$	0.1002	0.763	0.9987

seems more powerful (from simulations made with the same value of $n = 100$).
No effect tests statistics based on T_n and $T_{2,n}$ have been compared on three regression models

$$Y^k = 5(k-1)exp(-\int_0^1 X(t)cos(7.5t)dt) + \epsilon.$$

For each model, 10000 samples of 300 pairs (X_i, Y_i^k) have been considered. Each curve is simulated in a nonparametric way from the simulation of a brownian motion with the same approach as in Delsol (2008). The 300 pairs are splitted into three samples of size 100. Then, to make our comparison as fair as possible, T_n is computed from the three sub-samples while $T_{2,n}$ is computed on one sample of size 100 in such a way for both tests $n = 100$.

The first regression model (X, Y^1) corresponds to the null hypothesis, while (X, Y^3) and (X, Y^6) correspond to alternatives. The empirical levels of our tests (see column 2) are comparable and close to the nominal levels. However, $T_{2,n}$ seems to be more powerful than T_n (see columns 3 and 4).

4. Detection of informative features in explanatory curves with no effect tests

The effect of the explanatory curve on the response sometimes reduces to the effect of some of its features (parts, pointwise values, ...). Here are two simulations studies to observe if our testing procedures are relevant to detect such informative features. By simplicity, we consider situations where the informative and non-informative features of the curves are independent to check if the test is relevant to detect their respective nature. However, the features of a curve (derivatives, parts) may be dependent. In this case, the detection of informative features is more complex and require to consider variable selection tests. No effect tests on estimated residuals might be seen as an heuristic alternative.

4.1 A first attempt with test statistic T_n

In this section we use the no effect test statistic based on T_n. Hence, the orginal sample of size 300 is splitted into three subsamples of the same size to compute the test statistic.

Assume first explanatory curves (defined as before) are observed with an additive independent white noise η, that is to say we observe $\tilde{X}_i(t) = X_i(t) + \eta(t)$ instead of $X_i(t)$. A spline approximation (with three nots and splines of order 3) Z of each curve X is used to remove the independant noise (see Fig 1.a). The empirical rejection probabilities presented in Table 2 show our no effect testing procedure is able to detect the effect of the de-noised curve Z and does not detect any significative effect of the residual curve $R = \tilde{X} - Z$.

Table 2: No effect tests for the original curve X (unobserved), the denoised curve Z and the residual part R. Empirical rejection probabilites obtained on 1000 samples with $N_{boot} = 100$, $\alpha = 0.05$ and a data-driven choice of $h = h_0$ (see Delsol, 2008).

	Explanatory variable		
Model	X	Z	R
$M_0 : \; Y = \epsilon$	0.056	0.055	0.054
$M_1 : \; Y = 10 \exp\left(-\int_0^1 X(t)\, a(t) dt\right) + \epsilon$	0.502	0.505	0.056

Table 3: No effect tests for BM, BM_3 and R_3. Empirical rejection probabilites obtained on 1000 samples with $N_{boot} = 100$, $\alpha = 0.05$ and a data driven choice of $h = h_0$ (see Delsol, 2008).

	Explanatory variable		
Model	BM	BM_3	R_3
$M_1' : \; Y = \epsilon$	0.050	0.043	0.062
$M_2' : \; Y = 5 \exp\left(-\int_0^1 BM_3(t)\, cos(7.5t)\, dt\right) + \epsilon$	0.493	0.509	0.050
$M_3' : \; Y = 24 \exp\left(-\int_0^1 R_3(t)\, cos(7.5t)\, dt\right) + \epsilon$	0.065	0.046	0.523

A similar use of no effect test is finally considered to check if the effect of a brownian motion BM starting from 0 may be reduced to the effect of its first three principal components scores (explaining more than 90 percent of the variability). For each simulated sample, three model M_1', M_2', and M_3' have been introduced to cover no-effect of BM, effect of BM reduced to BM_3 (projection of BM on its first three components), and effect of BM reduced to $R_3 = BM - BM_3$. As expected, the effect of BM_3 (respectively R_3) , which may be regarded as the global shape (see Figure 1.b) of BM (respectively the deviation of BM from its global shape), is well detected as significative for model M_2' (respectively model M_3') and not significative elsewhere. However, even if signal to noise ratios in models M_2' and M_3' are similar, a significative effect of BM itself is only detected for M_2'. The use of the $\mathbb{L}^2$ metric gives a lot of importance to the first components of BM (explaining a great part of variability of BM) and is hence not relevant when the effect only comes from the residual part of the trajectory. The use of the semi metric induced by the three first PC scores is equivalent to consider BM_3 as explanatory variable, while the use of a metric based on remaining scores would lead to consider R_3. These no effect tests may be seen as tools to check if $r_d(x) = \mathbb{E}[Y|d(X, x) = 0]$ is constant or not. The use of various semimetrics may be relevant to detect the effect of some features of a curve.

4.2 Using the test statistic $T_{2,n}$

Because no effect tests based on $T_{2,n}$ seem more powerful and do not require to split the original sample, let us now consider the no effect test based on $T_{2,n}$. The test statistic is computed directly from an original sample of size 100. This makes fair the comparison with the results presented in the previous paragraph. The empirical rejection probabilities presented in Tables 4 and 5 show our new test statistic based on $T_{2,n}$ is able to detect informative parts of the curves. As discussed in the previous section, the use of the L^2 metric on the whole trajectory of the brownian motion has a negative impact on the capacity of the test to detect the effect of R_3. However, because no effect tests constructed from $T_{2,n}$ seem more powerful this effect is fairely often detected even when a L^2 metric is used directly on BM.

Table 4: No effect tests for the original curve X (unobserved), the denoised curve Z and the residual part R. Empirical rejection probabilites obtained on 1000 samples with $T_{2,n}$, $N_{boot} = 100$, $\alpha = 0.05$ and $h = 5,\ 5,\ 5,$ and 0.5 respectively.

	Explanatory variable		
Model	X	Z	R
$M_0 :\ Y = \epsilon$	0.058	0.058	0.054
$M_1 :\ Y = 10 \exp\left(-\int_0^1 X(t)\, a(t) dt\right) + \epsilon$	0.997	0.999	0.053

Table 5: No effect tests for BM, BM_3 and R_3. Empirical rejection probabilites obtained on 1000 samples with $T_{2,n}$, $N_{boot} = 100$, $\alpha = 0.05$ and $h = 0.4,\ 0.4,$ and 0.2 respectively.

	Explanatory variable		
Model	BM	BM_3	R_3
$M_1' :\ Y = \epsilon$	0.052	0.046	0.055
$M_2' :\ Y = 5 \exp\left(-\int_0^1 BM_3(t)\, cos\,(7.5t)\, dt\right) + \epsilon$	0.901	0.960	0.021
$M_3' :\ Y = 24 \exp\left(-\int_0^1 R_3(t)\, cos\,(7.5t)\, dt\right) + \epsilon$	0.455	0.043	0.982

5. Application in spectrometry

Spectrometric curves are an interesting example of functional data. They correspond to the measure of the absorption of a light emitted in direction of a product in function of its wavelength. Spectrometric curves have been used to give an estimation of the chemical content of a product without spending time and money in a chemical analysis (see for instance Borggaard and Thodberg, 1992). It is usual in chemometrics to make a pretreatment of the original curves (corresponding in some sense to considering derivatives). The approach described in this work may be used in this context to provide part of an answer to questions dealing with

- the validity of a model proposed by specialists.
- the existence of a link between one of the derivatives and the chemical content to predict.
- the nature of the link between the derivatives of the spectrometric curve and the chemical content of the product
- the validity of models in which the effect of the spectrometric curve is reduced to the the effect of some of its features (parts, points).

6. Prospects

To conclude, the structural procedures presented in this paper open a large potential scope of applications. They may be relevant to test a structural hypothesis formulated by scientists, to check if some structural assumptions used to compute an estimator are not significantly rejected, and to test the validity of some hypothesis coming from a first estimation of the regression operator. They could be used in an interesting way as part of an algorithm allowing to extract informative features (parts, points, ...) of the explanatory curve. An other prospect concerns their use in the choice a the semi-metric d since they may be used to test the regularity of r with respect to a semi-metric d_1 under its regularity with respect to d_2 if $d_1 \leq d_2$. Then, $T_{2,n}$ is an alternative to T_n which does not require to split the original sample and seems to lead to more powerful testing procedures. A lot of work has to be done on the theoretical study of this new test statistic and the development of other innovative testing procedures.

REFERENCES

[1] Ait-Saïdi, A., Ferraty, F., Kassa, R. and Vieu, P. (2008) Cross-validated estimations in the single functional index model. soumis
[2] Aneiros-Perez, G. and Vieu, P. (2008) Time series prediction: a semi-functional partial linear modeling. *Journal of Multivariate Analysis*, **99**, (5), 834-857.
[3] Borggaard, C. and Thodberg, H.H. Optimal minimal neural interpretation of spectra, *Analytical chemistry*, 64, (5), 545-551.
[4] Bosq, D. (2000) *Linear Processes in Function Spaces : Theory and Applications*, Lecture Notes in Statistics, 149, Springer-Verlag, New York.
[5] Cardot, H., Ferraty, F., Mas, A. and Sarda, P. (2003) Testing Hypothesys in the Functional Linear Model, *Scandinavian Journal of Statistics*, 30, 241-255.
[6] Cardot, H., Ferraty, F. and Sarda, P. (1999) Functional Linear Model *Statist. and Prob. Letters*, 45, 11-22.
[7] Cardot, H., Ferraty, F. and Sarda, P. (2000) Etude asymptotique d'un estimateur spline hybride pour le modèle linéaire fonctionnel. (French) [Asymptotic study of a hybrid spline estimator for the functional linear model] *C. R. Acad. Sci. Paris*, 330, (6), 501-504.
[8] Cardot, H., Goia, A. and Sarda, P. (2004) Testing for no effect in functional linear regression models, some computational approaches. *Comm. Statist. SimulationComput.* 33, (1), 179-199.
[9] Cardot, H., Crambes, C., Kneip, A. and Sarda, P (2007) Smoothing splines estimators in functional linear regression with errors-in-variables. *Computational Statistics and Data Analysis, special issue on functional data analysis*, 51, (10), 4832-4848.
[10] Chiou, J.M. and Müller H.-G. (2007) Diagnostics for functional regression via residual processes. *Computational Statistics & Data Analysis*, 51, (10), 4849-4863.
[11] Crambes, C., Kneip, A. and Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *Annals of Statistics*, 37, 35-72.
[12] Delsol, L. (2007) Régression non-paramétrique fonctionnelle : Expressions asymptotiques des moments. *Annales de l'I.S.U.P.*, LI, (3), 43-67.
[13] Delsol, L. (2008) Régression sur variable fonctionnelle: Estimation, Tests de structure et Applications. *Thèse de doctorat de l'Université de Toulouse*.
[14] Delsol, L. (2009) Advances on asymptotic normality in nonparametric functional Time Series Analysis. *Statistics*, 43, (1), 13-33.
[15] Delsol, L., Ferraty, F., and Vieu, P. (2011) Structural test in regression on functional variables. *Journal of Multivariate Analysis.* **102**, (3), 422-447.
[16] Ferraty F., Goia A. and Vieu P. (2002b) Functional nonparametric model for time series : a fractal approach for dimension reduction. *Test*, 11, (2), 317-344.
[17] Ferraty, F. et Romain, Y. (2011) *The Oxford Handbook of Functional Data Analysis.* (2011) Oxford Press.
[18] Ferraty, F. and Vieu, P. (2000) Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés, *Compte Rendus de l'Académie des Sciences*, Paris, 330, 403-406.
[19] Ferraty, F. and Vieu, P. (2006) *Nonparametric Functional Data Analysis: Theory and Practice* , Springer-Verlag, New York.
[20] Gadiaga, D. and Ignaccolo, R.(2005) Test of no-effect hypothesis by nonparametric regression. *Afr. Stat.*, 1, (1), 67-76.
[21] Hall, P. and Cai, T.T. (2006) Prediction in functional linear regression. (English summary) *Ann. Statist.*, 34, (5), 2159-2179.
[22] Härdle, W. and Mammen, E. (1993) Comparing Nonparametric Versus Parametric Regression Fits *Annals of Statistics*, 21, (4), 1926-1947.
[23] Masry, E. (2005) Nonparametric regression estimation for dependent functional data : asymptotic normality, *Stochastic Process. Appl.*, 115, (1), 155-177.
[24] Müller, H.-G. and Stadtmüller, U. (2005) Generalized functional linear models. *Ann. Statist.*, **33**, (2), 774-805.
[25] Mammen, E. (1993) Bootstrap and wild bootstrap for high-dimensional linear models. *Ann. Statist.*, 21, (1), 255-285.
[26] Preda, C.and Saporta, G. (2005) PLS regression on a stochastic process. *Comput. Statist. Data Anal.*, 48, (1), 149-158.

[27] Ramsay, J. and Dalzell, C. (1991) Some tools for functional data analysis, *J.R. Statist. Soc. B.*, 53, 539-572.

[28] Ramsay, J. and Silverman, B. (1997) *Functional Data Analysis*, Springer-Verlag, New York.

[29] Ramsay, J. and Silverman, B. (2002) *Applied functional data analysis : Methods and case studies*, Spinger-Verlag, New York.

[30] Ramsay, J. and Silverman, B. (2005) *Functional Data Analysis (Second Edition)*, Spinger-Verlag, New York.

[31] Sood, A., James, G. and Tellis, G. (2009) Functional Regression: A New Model for Predicting Market Penetration of New Products *Marketing Science*, 28, 36-51.

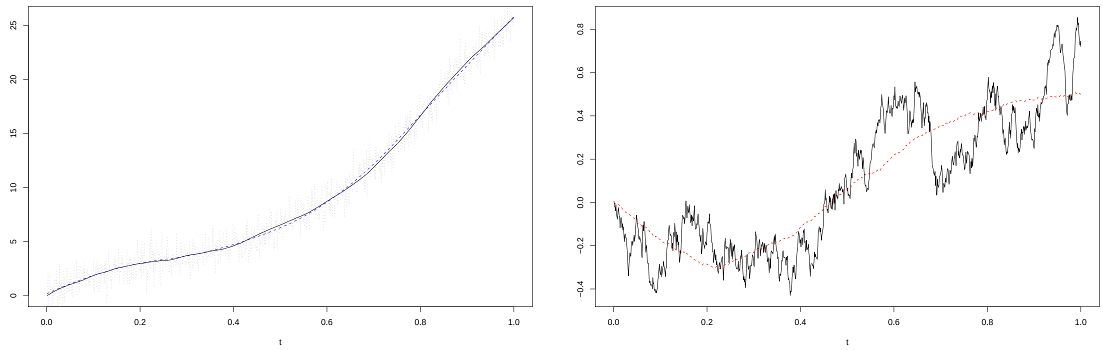

Figure 1: a) An example of noisy curve $\tilde{X}$ (dotted line), its denoised version Z (dashed line) and the true curve X (solid line). b) An example of simulated brownian motion trajectory BM (solid line) and its projection BM_3 (dotted line)