Working Paper Series n° 615

2009

"The Practice of Non Parametric Estimation by Solving Inverse Problems: the Example of Transformation Models"

Jean-Pierre FLORENS
and Frédérique FÈVE

Institut d'Economie Industrielle (IDEI) – Manufacture des Tabacs
Aile Jean-Jacques Laffont – 21, allée de Brienne – 31000 TOULOUSE – FRANCE
Tél. + 33(0)5 61 12 85 89 – Fax + 33(0)5 61 12 86 37 – **www.idei.fr** – contact@cict.fr

The Practice of Non Parametric Estimation by Solving Inverse Problems: The Example of Transformation Models

F. Fève † and J.P. Florens ‡

† *Toulouse School of Economics, 21 allée de Brienne, 31000 Toulouse, France*

E-mail:**feve@cict.fr**

‡ *Toulouse School of Economics, 21 allée de Brienne, 31000 Toulouse, France*

E-mail:**florens@cict.fr**

Revised version: January 2009

Abstract: This paper considers a semiparametric version of the transformation model $\varphi(Y) = \beta' X + U$ under exogeneity or instrumental variables assumptions ($E(U|X) = 0$ or $E(U|\text{instruments }) = 0$). This model is used as an example to illustrate the practice of the estimation by solving linear functional equations. This paper is specially focused on the data driven selection of the regularization parameter and of the bandwidths. Simulations experiments illustrate the relevance of this approach.

Keywords: Integral Equations. Tikhonov regularization. Instrumental variables. Selection of the regularization parameter and of the bandwidths.

JEL Codes: C14, C20

1 Introduction

The objective of this paper is to provide a simple guideline for the estimation of functional econometric parameters based on Tikhonov regularization of ill posed linear inverse problems.

We concentrate our presentation around a class of examples, namely the transformation models. This model is characterized by the relation:

$$\varphi(Y) = \beta' X + U$$

and has been extensively studied in the econometric literature following in particular the paper by Horowitz (1996). The origin of the transformation models is probably the Box Cox model where $\varphi(Y) = \frac{y^{a-1}}{a}$ if $a \neq 0$ and $ln\ y$ if $a = 0$. Several extensions of this family of transformation have been proposed (see Horowitz (1998) chapter 5 for references). These models are essentially parametric and have been estimated under endogeneity using instruments by GMM. In this paper, we treat this model semiparametrically: φ is a functional element and β is a vector of parameters. We assume that φ is monotonous and a particular example is the case $\varphi = S^{-1}$ where S is the cumulative distribution or the survivor of a random variable. This example covers in particular market shares models (Y is the observed proportion of individuals who take the choice 1 between 0 and 1. The choice 1 is selected if an individual characteristic θ is greater than $\beta' X + U$ and $1 - S$ is the c.d.f. of θ). An extension of this market share model covers the econometric models derived from the theory of two sided markets. For example, let us take the credit card market. The share of users of the credit card depends on the share of stores which accept the credit card and the share of stores depends on the share of users. This creates a system of transformation models which may be analysed in a limited information approach by transformation models with endogenous variables (see Rochet and Tirole (2003), Argentesi and Fillistrucchi, (2007)).

Many others examples and references to previous papers may be found in the Horowitz's book. In particular this class of model includes semiparametric analysis of durations models where φ is the integrated hazard function. Many new references consider this model and these references may be found e.g. in Linton, Sperlich and Van Keilegom (2008).

Two main differences characterize our model. We do not assume independence between U and X and we consider two cases : X exogenous defined by $E(U|X) = 0$ or X endogenous. In that case the model is estimated using instrumental variables.

For identification reasons we normalize β such that one element is equal to one and we consider the model:

$$\varphi(Y) = Z + \beta' W + U \tag{1.1}$$

where Z may be endogenous. In that case we assume that there exists a vector R of instruments such that $E(U|R, W) = 0$.

The simplest case consists in the model $\varphi(Y) = Z + U$ where $E(U|Z) = 0$ (exogeneity condition). Even in this case the parameter of interest φ should be considered as the solution of the equation

$$E(\varphi(Y)|Z) = Z \tag{1.2}$$

or in terms of density function

$$\int \varphi(y) f(y|z) dy = z. \tag{1.3}$$

Then the estimation of φ may be obtained by first the estimation of the conditional expectation operator $E(\varphi(Y)|Z)$ and second by solving the equation (1.2). The more general model (1.1) under an instrumental variables assumption satisfies the condition:

$$E(\varphi(Y)|W, R) = E(Z|W, R) + \beta' W \tag{1.4}$$

where the two conditional expectations may be estimated and the equation (1.3) needs to be solved w.r.t. φ and β in order to estimate the parameters.

This example illustrates the inverse problems approach in econometrics. The economic theory defines a structural model where the (possibly functional) parameters φ are linked with the observation scheme by a (functional) equation $A(\varphi, F) = 0$ where F is the data cumulative distribution function. The statistical analysis is then performed in two steps.

First we estimate the equation using for example an *i.i.d.* sample of data whose distribution is F and secondly we solve the equation in order to recover the parameters of interest. This approach is very common in econometric and a usual example is provided by GMM where the parameters φ and F are linked by a relation $E^F(h(X, \varphi)) = 0$.

The main question coming from the nonparametric approach concern the ill posedness of the inversion. The solution of the equation may not exist or is not in general a continuous function of the estimated part of the equation. The estimation is then not consistent in many cases. There exists several ways to treat this problem: we can reduce the parameter space to be compact (see Ai and Chen (2003) or Newey Powell (2003)) or we can keep general the parameter space by introducing a regularized solution in the equation. Instead of solving $A(\varphi, F) = 0$ the principle is to minimize a penalized criterium

$$\|A(\varphi, F)\|^2 + \alpha\|\varphi\|^2 \tag{1.5}$$

where the norms are suitably chosen and where α goes to zero at a suitable rate. The minimization of (1.5) leads to the Tikhonov regularization approach but other regularizations may be used.

The regularized solutions of ill posed inverse problems are standard in numerical analysis and in image treatment and have been introduced in econometric to solve GMM estimation in infinite dimension (see Carrasco and Florens (2000)) and in non parametric estimation using instrumental variables (see Florens (2003), Darolles, Florens and Renault (2003), Hall and Horowitz (2005), Carrasco, Florens and Renault (2007)).
The main objective of this paper is to present an introduction to inverse problems, both on its practical implementation and on the main mathematical arguments of the derivation of the rate of convergence. This paper also contains different contributions to this literature. Identification of the transformation model without independence is based on standard tools but it contains new results. For example, the estimation of the transformation model under mean independence condition is a contribution of this paper. The rate of convergence of the estimators is not derived in previous articles on inverse problems on instrumental variables. the demonstration is founded on general arguments which have been developed in, e.g. Carrasco et al. (2007). The selection of α we suggest is derived from a known technic and

cross validation selection of bandwidth is standard but the recursive application of these approaches have not been presented previously in the literature. This paper is not a survey of inverse problems in econometrics (see e.g. Florens (2003) and Carrasco et al. (2007) for more examples of application of this theory). However we may locate our basic examples in the general class of ill posed inverse problems.
The main characteristics of our example is to be linear, with an integral unknown operator. This operator is a conditional expectation operator. Linear inverse problems take the form $T\varphi = r$ and usually only r is estimated and T is given. This is not true in our case and $T\varphi$ is equal to the conditional expectation of φ given some random elements. Other relevant models belong to this class, essentially the basic nonparametric instrumental variables model $(Y = \varphi(Z) + U, E(U|W) = 0)$ which leads to $E(\varphi(Z|W)) = E(Y|W)$, very similar to the model treated in section 5. This question has been addressed in Darolles, Florens and Renault (2003) and Hall, Horowitz (2003) in particular. This nonparametric inverse problem has been extended to additive models (see Florens, Johannes and Van Bellegem (2005)) or has been used to test parametric (see Horowitz(2006)) or exogeneity assumptions (see Blundell, Horowitz (2007)). In all that cases, this problem is ill posed because T is a compact integral operator. The problem becomes well posed if equations $\varphi + T\varphi = r$ are considered where T may be still an unknown conditional expectation operator (see Mammen and Yu (2008)). The literature on inverse problems is essentially theoretical in econometrics but an empirical application is presented in Blundell, Chen and Kristensen (2007). The link between instrumental variables and simultaneous equations models is treated in Chernozhukov, Imbens and Newey (2007).
Linear ill posed inverse problems where the operator is not the expectation operator are relevant in econometrics. A class of examples is based on the covariance operators $(T\varphi = E(X < W, \varphi >))$ estimated by $\hat{T}\varphi = \frac{1}{n}\Sigma x_i < w_i, \varphi >$ which defines an ill posed problem if the data are functional data (see Cardot and Johannes (2009), Florens and Van Belleghem (2009)). An illustration is the linear instrumental regression model with many regressors and instruments (see Carrasco (2008)). The different forms of deconvolution problems (when the operator is the convolution with a known or unknown density) has generated a huge literature and is particularly adapted to many econometrics problems. (see e.g. Carrasco and Florens (2000) or Bonhomme and Robin (2008)). More recently the researchers have been interested to non linear inverse problems motivated by non separable models (to treat quantile regression under endogeneity , for example, see Horowitz and Lee (2007), Gagliardini and Scaillet (2007), Chernozhukov, Gagliardini and Scaillet (2008) or by a general approach to GMM with functional parameters (see,

Chen and Pouzo (2008 a and b). These models create difficult numerical questions. Other non linear inverse problems come from game theoretic models (see Florens and Sbai (2009)). Inverse problems in more complex spaces or using other classes of operators may be founded in Hoderlein, Klemelä and Mammen (2009) and in Gautier and Kitamura (2008)).

The paper is organized as follows. In section 2, the model is described and the identification is examined. Section 3 presents a simple example in more details. Some asymptotic properties are considered in section 4 and semiparametric extension and instrumental variable approach are studied in section 5. Numerous simulations and some technical details are reported in Appendix I and II.

2 An example of a semi parametric transformation model

We assume that all the variables and functions that we consider are square integrable. The model satisfies:

$$\begin{array}{l} \varphi(Y) = Z + \beta' W + U \\ Y \in I\!R \quad Z \in I\!R \quad W \in I\!R^k \end{array} \tag{2.1}$$

where U is an unobservable noise. The model is semiparametric and the parameter space contains a non decreasing function φ and a vector $\beta \in I\!R^k$. Equation (2.1) may be completed by one of these hypothesis.

$$\text{Exogeneity: } E(U|Z,W) = 0 \tag{2.2}$$

$$\text{Instrumental Variables: } E(U|R,W) = 0 \tag{2.3}$$

where R is a random vector.
As we will see below, these mean independence conditions are sufficient, up to some regularity assumptions, to identify the φ and β elements and an estimation procedure will be naturally derived from condition (2.2) or (2.3). The Box Cox models or their extensions are naturally developed in the regression case, i.e. with $E(U \mid Z) = 0$ and not under an independence assumption between U and Z. The main motivation of the analysis of the problem under these weaker assumptions is to consider cases where high order moments of U may depend on (Z,W) or (R,W). Practically, heteroscedasticity is extremely common in empirical research. The theory where U and (Z,W) are independent is well established but this is not the case where Z is endogenous. In that case the treatment of the problem will lead to a non linear integral

equation problem (as in Horowitz and Lee (2007)) that may be difficult to analyse.It follows that in the endogenous case the mean independence conditions leads to a more simple procedure for the estimation of φ and β. If we trust into the full independence and if we want to use a non linear Tikhonov procedure (or other regularization methods for the non linear inverse problem), our estimator will provide a natural (because consistent) initial value for the required recursive procedure.
To analyze the identification of these model, we need to recall two important concepts extensively used in the theory of resolution of inverse problems involving conditional expectation operators.

First a random element A is say to be strongly identified by B given C if $E(g(A,C)|B,C) = 0$ a.s. implies $g = 0$ a.s. for any square integrable function g. (see Florens, Mouchart and Rolin (1990)). This concept has been introduced in statistics under the name "completeness" in a particular case. Secondly a random element A is said to be measurably separated to an other random element B if equality $g(A) = h(B)$ a.s. implies $g = h =$ constant a.s. (see also Florens, Mouchart, Rolin (1990)). This concept has also along history in statistics in the theory of sufficient and ancillary statistics.

Identification theorem may then be written as follows.

Theorem 2.1: Let us consider model (2.1) under the exogeneity condition (2.2).

Let us assume:

- Assumption A1: $E(WW')$ is invertible and W only contains non constant variables,
- Assumption A2: Y is strongly identified by Z given W,
- Assumption A3: Y and W are measurably separated.

Then φ and β are identified.

Proof: Let us consider two solutions φ_0, β_0 and φ_1, β_1 to equation (2.1). Then if $\varphi = \varphi_1 - \varphi_0$ and $\beta = \beta_1 - \beta_0$ we have

$$E(\varphi(Y)|W, Z) = \beta' W \tag{2.4}$$

and we have to proof that this implies $\varphi = 0$ and $\beta = 0$. Equation (2.4) and A2 implies $\varphi(Y) - \beta' W = 0$ and A3 implies that $\varphi(Y) = \beta' W = c$ where c is a real constant. As W is not constant $\beta' W = c$ implies $c = 0$ and then $\beta = 0$ under 1. Finally $\varphi(Y) = 0$. ■

An analogous proof gives the following generalization :

Theorem 2.2: Let us consider model (2.1) under (2.3). If we assume A1, A3 and A2' where:

- Assumption $A2'$: Y is strongly identified by R given W.

Then φ and β are identifiable.

The assumptions of Theorems 2.1 and 2.2 do not seem to be immediately interpretable. However they can be illustrated by the following comments. The assumption $A3$ (Y and W are measurably separated) is essentially a support condition (for a precise statement see Florens, Heckman, Meghir and Vytlacil (2008), theorem 2). It means that there does not exist an exact relation between W and Y or equivalently that the derivative of W is w.r.t. Y is zero. Then $\varphi(Y) - \beta' W = 0$ implies $\frac{d\varphi}{dY} = 0$ and $\varphi = constant$. This hypothesis is false is $Z + U$ is constant which is an extreme dependence between Z and the noise U. More generally it is sufficient that $Z + U$ may vary independently of W to verify the assumption. Assumption A2 is more severe. For simplicity we may eliminate W (or we can consider the question with respect to the conditional distribution of $W = w$, w fixed). The assumption A2 is a dependence condition between Y and Z. It is known that if Y and Z are jointly normal, this assumption is equivalent to rank $Cov(Y, Z)) = 1$. General characterizations of this dependence are more difficult (see for example a recent contribution of d'Haultfoeuille (2008)). Intuitively this assumption means there exists no function of Y non correlated to any function of Z. If this assumption is false the theory is essentially preserved but φ may not be fully estimable but only up to any function of Y orthogonal to any function of Z. The recent developments on "set identification" may applied in that case.

Remark 2.1: The model analysed in theorem 2.1 can be extended to the case where $\varphi(Y)$ becomes a function of $\varphi(Y, \Xi)$ where Ξ are some exogenous variables. However extension of Theorems 2.1 and 2.2 to that case requires that Ξ and W should be measurably separated. This condition excludes the case where Ξ and W have some common elements. Common elements between Ξ and W prevent the identification of the model. Moreover assumptions A2 should be modified by "Y is strongly identified by Z given Ξ and W". Similar extensions of theorem 2.2 may be done also.

Remark 2.2: All the variables we consider in the paper are assumed for simplicity continuous variables. All our results applied but some hypothesis may be false in presence of discrete variables. For example, hypothesis A2 is no longer true if Y is continuous and Z discrete but the case Y discrete and Z continuous usually satisfies A2. In case of instrumental variables approach, R should be continuous if Y is continuous. In all cases, W may contain discrete variables but if Y is discrete the support conditions A3 need to be check carefully. If Y may take only a finite number of values, the functional estimation problem becomes a finite dimensional question and the difficulty of ill posedness disappears.

3 Estimation by Tikhonov regularization

We illustrate our analysis by the particular simple case

$$\varphi(Y) = Z + U \quad E(U|Z) = 0 \tag{3.1}$$

We assume that we are $i.i.d.$ sample of (Y, Z) is available and denoted by $(y_i, z_i) i = 1, ..., n$. Equation (3.1) implies

$$E(\varphi(Y)|Z) = Z \tag{3.2}$$

and the usual kernel smoothing estimation gives the following empirical counterpart of this equation.

$$\frac{\sum_{i=1}^{n} \varphi(y_i) K\left(\frac{z - z_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{z - z_i}{h_n}\right)} = z \tag{3.3}$$

where K is a univariate kernel and h_n the bandwidth. This equation has no solution in general because there does not exist a linear combination of the functions $\frac{K\left(\frac{z-z_i}{h_n}\right)}{\Sigma_{i=1}^{n} K\left(\frac{z-z_i}{h_n}\right)}$ equal to z. The resolution of equation (3.2) is then ill posed. We then adopt a Tikhonov regularization which is based on the minimization of

$$\|T\varphi - Z\|^2 + \alpha \|\varphi\|^2 \tag{3.4}$$

where $T\varphi = E(\varphi(Y)|Z)$ (T is an operator from L^2_Y to L^2_Z defined w.r.t. the true data distributions) and the two norms are L^2 norms. ($\|\varphi\|^2 = \int \varphi^2(z) f(z) dz$ if f is the true density of Z). This minimization leads to the first order condition:

$$\alpha\varphi + T^*T\varphi = T^*Z \tag{3.5}$$

where T^* is the adjoint operator of T. A general theory for adjoint operators is not necessary here and it is sufficient to note that T^* is the conditional expectation operator of functions of Z given Y. Then the first order condition of minimization of (3.4) is:

$$\alpha\varphi(y) + E(E(\varphi(Y)|Z)|Y = y) = E(Z|Y = y) \tag{3.6}$$

The empirical counter part of this equation may be written:

$$\begin{aligned} \alpha\varphi(y) \quad &+ \quad \frac{\displaystyle\sum_{j=1}^{n} \frac{\displaystyle\sum_{i=1}^{n} \varphi(y_i) K\left(\frac{z_j - z_i}{h_n}\right)}{\displaystyle\sum_{i=1}^{n} K\left(\frac{z_j - z_i}{h_n}\right)} K\left(\frac{y - y_j}{h_n}\right)}{\displaystyle\sum_{j=1}^{n} K\left(\frac{y - y_j}{h_n}\right)} \\ &= \quad \frac{\displaystyle\sum_{j=1}^{n} z_j K\left(\frac{y - y_j}{h_n}\right)}{\displaystyle\sum_{j=1}^{n} K\left(\frac{y - y_j}{h_n}\right)} \end{aligned} \tag{3.7}$$

This equation may be solved in two steps. Consider first equation (3.7) for $y = y_1, ..., y_n$. Then (3.7) reduces to a matrix equation:

$$\alpha\bar{\varphi} + C_Y C_Z \bar{\varphi} = C_Y \bar{z} \tag{3.8}$$

where $\bar{\varphi}$ is the vector of the $(\varphi(y_j))_{j=1,...,n}$, $\bar{z}$ the vector of $(z_j)_{j=1,...,n}$ and C_Z and C_Y two $n \times n$ matrices:

$$C_Z = \left(\frac{K\left(\frac{z_j - z_i}{h_n}\right)}{\sum_i K\left(\frac{z_j - z_i}{h_n}\right)} \right)_{j,i=1,...,n} \qquad C_Y = \left(\frac{K\left(\frac{y_l - y_j}{h_n}\right)}{\sum_j K\left(\frac{y_l - y_j}{h_n}\right)} \right)_{l,j=1,...,n}$$

Equation (3.8) has a solution

$$\hat{\bar{\varphi}}^\alpha = (\alpha I + C_Y C_Z)^{-1} C_Y \bar{Z} \tag{3.9}$$

involving the inversion of an $n \times n$ matrix[1]

If we want $\varphi(y)$ for a value y which does not belongs to the sample we may use equation (3.7) for which $\varphi(y)$ may be derived immediately from the knowledge of $\hat{\hat{\varphi}}^{\alpha}$.

Remark 3.1: In the particular case of market share models, the random variable Y is constrained to belong to the $[0,1]$ interval. In that case we are faced to boundary problems in the kernel estimation. We solved this difficulty by using boundary kernels in the estimation of conditional expectations given Y. We use boundary gaussian kernel defined e.g. in Li and Racine (2006).

Remark 3.2: The model implies that φ is monotonous non increasing. We don't impose this restriction in our estimation even if the minimization of (3.4) under constraint is feasible (see e.g. Engl, Hanke and Neubauer (2000)). The estimation without monotony constraint illustrate in a better way the impact of the selection of α because the monotony constraint is a regularization and the distinction between the impact of the penalization by $\alpha\|\varphi\|^2$ and the constraint is not easy. Moreover our model is then a little more general and not restricted to usual transformation models.

The implementation of our method depends on the selection of the bandwidths in the different kernel estimations and on the value of the regularization parameter α.
The bandwidths may be chosen using many methods. We will compare two of them:

i) "Naive" bandwidth. As recommended by many authors (see e.g. Silverman (1986))we may choose $1.059 \times$ standard deviation of the variable $\times n^{-\frac{1}{5}}$.

ii) Cross validation. Recall that the bandwidth may be chosen for the estimation of $E(g(A)|B)$ by minimization of the sum of square of the residuals computed using the leave-out-method (the residual of an observation is computed by dropping out this observation in the estimation). We then have three bandwidths to compute: the one corresponding to $E(Z|Y)$, the one corresponding to $E(\varphi(Y)|Z)$ and finally the bandwidth of the estimation of $E(E(\varphi(Y)|Z)|Y)$. The last two bandwidths require a preliminary estimation of φ in order to be computed.

[1] The Tikhonov regularization needs this inversion of a possibly large matrix. If n is very large, other methods like Landweber-Fridman regularization may be used which do not requires inversion.

The selection of α (given the bandwidth) is also a standard issue in regularized solution of linear equations. We adopt a version of the principle described in Engl et al (2000). This method consists in the following procedure.

i) For any (small) value of α compute the estimation of φ by an iterated Tikhonov approach. This estimation is defined by:

$$\begin{aligned} \hat{\hat{\varphi}}_2^{\alpha} &= (\alpha I + C_Y C_Z)^{-1} C_Y \bar{z} + \alpha \hat{\hat{\varphi}}^{\alpha} \\ &= (\alpha I + C_Y C_Z)^{-1} [C_Y + (\alpha I + C_Y C_Z)^{-1} C_Y] \bar{z} \end{aligned}$$

   Even if our final estimate will be based on usual (non iterated) Tikhonov regularization, iterated method is necessary to determine α optimal for the non iterated scheme

ii) minimize the following sum of square

$$SS(\alpha) = \frac{1}{\alpha} \sum_{j=1}^{n} \left[\frac{\sum_{i=1}^{n} \widehat{\overline{\varphi}}_{(2)}^{\alpha}(y_i) K\left(\dfrac{z_j - z_i}{h_n}\right)}{\sum K\left(\dfrac{z_j - z_i}{h_n}\right)} - z_j \right]^2 \tag{3.10}$$

The idea is to minimize the norm of the residuals of the integral equation $E(\varphi(Y)|Z) = z$ where the conditional expectation is replaced by its estimator, the norm by the empirical mean of the squares and the φ by its estimator. This norm should be divided by α in order to admit a minimum. We will show in section 4 that the value of α which leads to the optimal rate of convergence of our estimator should be proportional to $n^{-\frac{4}{5(\beta+1)}}$.Engl et al.(2000) show that the value of α which minimizes (3.1) satisfies this condition and then this value of α may be call optimal in the rate sense. [2].

The bandwidth and α may be chosen sequentially: we start by naive bandwidths an we minimize (3.10) in order to get a first value of φ which may be used to improve the bandwidth by cross validation. A new value of α is then obtained from the minimization of α. The process may be then recursively updated.

[2]The proof given by Engl et al. (2000) is done in the case where the operator T is known. The extension of the proof in the case of unknown T operator is given in the Appendix II

4 Asymptotic properties

Even if this paper is focused on practical implementation, this section gives a low technical flavor of the asymptotic analysis. The objective of this section is to provide the general method for the analysis of the rate of convergence of an estimator derived from a Tikhonov regularization. We concentrate this study to the case of model (3.1) and refers to different papers for more general cases.

Let us recall that the estimator $\hat{\varphi}^\alpha$ is the solution of the equation (3.6) where the conditional expectations operators $T\varphi = E(\varphi(Y)|Z)$ or $T^*\psi = E(\psi(Z)|Y)$ are replaced by kernel estimators $\hat{T}$ or $\hat{T}^*$ defined analogously to $\hat{T}$. In an abstract way we have

$$\hat{\varphi}^\alpha = (\alpha I + \hat{T}^*\hat{T})^{-1}\hat{T}^*Z \tag{4.1}$$

The asymptotic properties of $\hat{\varphi}^*$ are based on two properties of the kernel estimation.

i) First we consider $\|\hat{T}\varphi - Z\|^2 = \int((\hat{T}\varphi)(z) - z)^2 f(z)dz$.

   Using usual results on the kernel smoothing we will assume that our problem is sufficiently regular in order to have

$$\|\hat{T}\varphi - Z\|^2 \sim O\left(\frac{1}{nh_n} + h_n^{2\rho}\right).$$

   In this expression ρ is the minimum value between the smoothness of φ and the order of the kernel. We simplify our presentation by considering probability kernels and twice continuous functions and then [3]$\rho = 2$. All the O in the paper are in probability.

ii) We also assumed that the two norms of $\|\hat{T} - T\|^2$ and $\|\hat{T}^* - T^*\|^2$ are $O\left(\frac{1}{nh_n^2} + h_n^4\right)$. Intuitively these results are based on the rate of convergence of the kernel estimator of the joint density of Y and Z to the true density. [4]. Note that $\hat{T}^*$ is an estimator of T^* and not the adjoint of $\hat{T}$.

An important component of the calculus of the rate of convergence is the regularity assumption on φ. As we will see in Appendix II the asymptotic analysis involves a term:

[3] See Darolles et al.(2003) or Hall and Horowitz (2005)

[4] See Darolles, Florens et al. (2003)

$$C = (\alpha I + T^*T)^{-1}T^*T\ \varphi - \varphi = -\alpha(\alpha I + T^*T)^{-1}\ \varphi$$

This term represents the difference between the true function and the regularized solution of the "true" problem $T\varphi = r \quad ((\alpha I + T^*T)^{-1}\ T^*T\varphi)$.

The value $\|C\|$ is called the regularization bias. $\|C\| \to 0$ if $\alpha \to 0$ but not uniformly w.r.t. φ. In order to control the rate of decline of $\|C\|^2$ when $\alpha \to 0$ φ should be constrained to be an element of a regularity class : φ is said to have the regularity $\beta > 0$ (w.r.t. the joint data generating process) if $\|C\|^2 \sim O(\alpha^\beta)$. For example (see Carrasco et al.(2007)) if there exists a function $\psi(z)$ such that $\varphi(y) = E(\psi(Z)|Y = y)$ φ has the regularity 1. The characterization of the regularity is a very complex question which is not treated here (see e.g. Carrasco et al (2007)). Note finally that a constraint imposed by Tikhonov regularization is that $\beta \leq 2$. If $\beta > 2$, it should be replaced by 2.

Theorem 4.1 Under the previous hypothesis i, ii and the regularization condition on φ we have:

$$\|\hat{\varphi}^\alpha - \varphi\|^2 = O\left(\frac{1}{\alpha n h_n} + \frac{h_n^4}{\alpha} + \frac{1}{\alpha n h_n^2}\alpha^{\min(\beta+1,2)} + \frac{h_n^4}{\alpha}\alpha^{\min(\beta+1,2)} + \alpha^\beta\right)$$

■

The estimator is then consistent if $\alpha \to 0$ and $h_n \to 0$ such that $\alpha n h \to \infty$, $\frac{h^4}{\alpha} \to 0$ and $\alpha^{[1-\min(\beta+1,2)]} n h_n^2 \to \infty$.

The question is now to select the optimal value of α and to derive the speed of convergence. In our approach h_n is selected by cross validation constructed from estimation of conditional expectations given a single variable. Then h_n is proportional to $n^{-\frac{1}{5}}$. In that case the optimal choice of α is obtained by balancing $\frac{1}{\alpha n^{\frac{4}{5}}}$ and α^β and leads to $\alpha = n^{-\frac{4}{5}(\beta+1)}$.

In that case it is clear that the two other terms are negligible and we get:

$$\|\hat{\varphi}^\alpha - \varphi\|^2 \sim O\left(n^{-\frac{4\beta}{5(\beta+1)}}\right)$$

The component $n^{-\frac{4}{5}}$ is due to non parametric estimation and the factor $\frac{\beta}{\beta+1}$ follows from the resolution of the integral equation. Note that this rate is the actual rate of our procedure characterized by a specific choice of the regularization parameters.

The optimality of this rate of convergence is a complex question and we just give in this paper an intuitive answer. Our rate is optimal under our hypothesis which do not link the regularity conditions of the kernel estimation and of the inverse problem. Intuitively the speed of convergence of the kernel estimation is based on differentiability conditions of φ and on the joint density of Y and Z. The source condition ($\| C \|^2 \sim O(\alpha^\beta)$) is based on the Fourier decomposition of φ on the singular vectors basis of T. In general ρ and β are not related. However if the source condition is derived from a degree of ill posedness of T and from a regularity condition on φ both measured relatively to the differential operator (defining an Hilbert scale), the rate may be improved under this set of stronger hypothesis. This analysis has been done by Chen and Reiss (2007) and Johannes, Van Bellegem and Vanhems (2007). See also Darolles et al.(2003) for a discussion on the minimax property of inverse problems solutions. We just consider here the consistency and the rate of convergence but the asymptotic normality may also be examined (see Darolles et al. (2003) and Horowitz (2007)).

5 Extensions to endogenous variables and semi-parametric models

Let us first consider the model $\varphi(Y) = Z + U$ where Z is endogenous and $E(U|R) = 0$ where R is a real instrumental variable. We now solve the empirical counterpart of:

$$\alpha\varphi(y) + E(E(\varphi(Y)|R)|Y) = E(E(Z|R)|Y) \tag{5.1}$$

Using the same arguments as in section 3 it may be shown that the vector $\hat{\bar{\varphi}}^\alpha$ verifies:

$$\hat{\bar{\varphi}}^\alpha = (\alpha I + C_Y C_R)^{-1} C_Y C_R \bar{Z}$$

where C_R is defined analogously to C_Y or C_Z.

The asymptotic properties of these estimator are very similar to these studied in Darolles et al (2003). The choice of α and of the bandwidth is done analogously to the case where Z is exogenous.

Let us now analyze semiparametric estimation : We first consider the simple case of model (1.1) where $W \in I\!R$ under an exogeneity assumption:

$$\varphi(Y) = Z + \beta W + U \quad E(U|Z, W) = 0 \tag{5.2}$$

We adopt a sequential approach extending the backfitting principle frequently used in semiparametric statistics.

- If φ is given, β may be obtained OLS method where the dependent variable is $\varphi(Z)-Z$ and the explaining variable is W because $E(U|W)=0$

- If β is given, our approach is identical to the one presented in section 3 replacing Z by $Z+\beta W$ because $E(U|Z+\beta W)=0$.

The algorithm iterates these two steps up to convergence. An initial value for β should be selected and should be not too far to the true value. In many cases 0 may be a suitable initial value.

This algorithm converges to the solution of the set of the the two equations:

$$E(W(\varphi(Y)-\beta W))=0 \tag{5.3}$$

$$E(\varphi(Y)|Z+\beta W)=Z+\beta W \tag{5.4}$$

The second equation is actually regularized and transformed into

$$\begin{aligned}\alpha\varphi(y)+E[E(\varphi(Y)|Z+\beta W)|Y=y]\\ =E[Z+\beta W|Y=y]\end{aligned} \tag{5.5}$$

We extend this analysis by considering Z as an endogenous variable and we use two instruments R and W. The computations are also realized using a recursive algorithm:

- The step where φ is given is analogous to the first step if Z is exogenous.

- The step where β is given is performed by solving:

$$\begin{aligned}\alpha\varphi(y)+E(E(\varphi(Y)|R,W)|Y=y)\\ =E(E(Z+\beta W|R,W)|Y=y)\end{aligned} \tag{5.6}$$

where the conditional expectations are replaced by their empirical counterparts.

The different bandwidths and the α parameter are computed by purely data driven methods as in section 3. In the sequential algorithms these parameters are updated at each step of the algorithms.

The key question concern the asymptotic properties of the estimator of the parametric β. It has been proved (see Ai and Chen (2003) and Florens et al (2005)) that β is asymptotically normal and converges at $\sqrt{n}$ speed.

6 Conclusion

This paper proposes an approach to the transformation model based on a conditional mean hypothesis and not on an independence condition between the exogenous variables (or the instrumental variables) and the residual. This weaker assumption leads to estimate the functional parameter by solving an integral equation of type I and then to construct estimators with different rates of convergence from the usual $\sqrt{n}$ rate. The treatment of the endogeneity of some variables is however easier under this weaker assumption.

This family of semiparametric transformation models is taken in this paper as a class of examples of econometric inverse problems. We want to show that despite the technicality of the mathematical framework the technology of Tikhonov regularization is easy to implement. We illustrate this simplicity using numerous simulations presented in the paper.

The usual difficulty of the practical use of nonparametric technics is the selection of the bandwidths and of the regularization parameters. We present in this paper a purely data driven strategy for these bandwidths and parameters. We illustrate by simulations the relevance of our methods.

This paper is not a pure theoretical contribution but we present in section 4 the main intuitions of the analysis of asymptotic properties, essentially the consistency and the rate of convergence.

Acknowledgements: The authors thank F. Collard and P. Fève for helpful suggestions and comments. They also thank S. Van Bellegem for numerous discussions and the participants of the EC^2 conference (Roma December 2008) and of the workshop at the Economic department of the Brown University. We also thank the editor and two anonimous referees for their remarks and suggestions

References

Ai, C. and X. Chen (2003). Efficient Estimation of Models with conditional Moment Restrictions Containing Unknown Functions. *Econometrica, 71(6)*,1795-1843.

Argentesi E. and L. Fillistrucchi (2007). Estimating market power in a two-sided market: The case of newspapers. *Journal of Applied Econometrics.*

Blundell, R., X. Chen and D. Kristensen (2007). Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves. *Econometrica 75(6)*,1613-1669.

Blundell, R. and J. Horowitz (2007). A Non-Parametric Test of Exogeneity. *Review of Economic Studies, Volume 74-4*,1035-1058.

Bonhomme S. and J.M. Robin (2008). Generalized Non parametric Deconvolution with an application to earning Dynamics. Forthcoming in *Review of Economic Study.*

Cardot H. and J. Johannes (2009). Thresholding Projection Estimators in Functional Linear Models. Forthcoming in *Journal of Multivariate Analysis).*

Carrasco M.(2008). A regularization approach to the many instruments problem. Mimeo Montreal.

Carrasco, M. and J.P. Florens (2000). Generalization of the GMM in Presence of a Continuum of Moment Conditions. *Econometric Theory, 16*, 797-834.

Carrasco, M. and J.P. Florens (2009). Spectral method for deconvolving a density, forthcoming in *Econometric Theory.*

Carrasco, M., J.P. Florens and E. Renault (2007). Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization in Heckman J.J. and E.E. Leamer (ed.), *Handbook of Econometrics, volume 6B*, North Holland Amsterdam.

Chen, X. and M. Reiss (2007). On Rate Optimality for Ill-posed Inverse Problems in Econometrics. Forthcoming in *Econometric Theory.*

Chen X. and Pouzo D.(2008). Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals. Forthcoming in *Journal of Econometrics.*

Chen X.and Pouzo D.(2008). Efficient of Nonparametric Conditional Moment Models with Possibly Nonsmooth Moments. Mimeo, Yales.

Chernozhukov V., P. Gagliardini and Scaillet O., (2007). Nonparametric instrumental variable estimators of quantile structural effects. HEC, swiss finance institute, technical report.
Chernozhukov, V., G.W. Imbens, and W.K. Newey (2007). Instrumental variable identification and estimation of nonseparable models via quantile conditions. *Journal of Econometrics, 139*, 4-14.

Darolles, S., J.P. Florens and E. Renault (2003). Non Parametric Instrumental Regression. IDEI Working Paper.

d'Haultfoeuille, X., (2008). On the Completeness Condition in Nonparametric Instrumental Problems. ENSAE, CREST-INSEE, Université Paris I Panthéon Sorbonne. Forthcoming in *Econometric Theory.*

Engl, H.W., M. Hanke and A. Neubauer (2000). *Regularization of Inverse Problems*, Kluwer Dordrecht.

Florens, J.P. (2003). Inverse Problems in Structural Econometrics. The Example of Instrumental Variables. In Dewatripont M., Hansen L.P., Turnovsky, S.J. (Eds) *Advances in Economics and Econometrics, Volume 2*, Cambridge University Press, Cambridge.

Florens J.P. and E. Sbai, (2009). Local Identification in Empirical games of incomplete information. Forthcoming in *Econometric Theory.*

Florens J.P., J.J. Heckman, C. Meghir and E. Vytlacil (2008). Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica,76*, 1191-1206.

Florens, J.P. , J. Johannes and S. Van Bellegem (2005). Instrumental Regression in Partially Linear Models. Forthcoming in *Econometric Journal.*

Florens, J.P., J. Johannes and S. Van Bellegem (2007). Identification and estimation by penalization in Nonparametric Instrumental Regression. Forthcoming in *Econometric Theory.*

Florens, J.P., M. Mouchart and J.M. Rolin (1990), *Elements of Bayesian Statistics*, Marcel Dekker, New York.

Florens J.P. and S. Van Belleghem (2009). Functional Instrumental Linear Regression. Mimeo Toulouse.

Gagliardini P. and O. Scaillet (2007). A specification test for non parametric instrumental variable regression, working paper HEC, Swiss Finance Institute.

Gautier E. and Y. Kitamura (2008)," Non parametric estimation in Random Coefficients Binary Models". Mimeo Yales.

Hall, P. and J. Horowitz (2003). Non Parametric Methods for Inference in the Presence of Instrumental Variables. *Annals of Statistics, 33 (6)*, 2904-2929.

Horowitz, J.L. (2006). Testing a parametric model against a nonparametric alternative with identification through instrumental variables. *Econometrica 74*, 521-538.

Horowitz, J.L. (2007). Asymptotic normality of a nonparametric instrumental variables estimator. *International Economic Review, 48*, 1329-1349.

Hoderlein S., Klemelä J. and E. Mammen (2009), "Reconsidering the Random Coefficient Model", forthcoming, in *Econometric Theory.*

Horowitz, J.L. (1998), *Semiparametric methods in econometrics*, lecture notes in statistics vol 131, Springer Verlag, New-York.

Horowitz, J.L. (1996). Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable. *Econometrica, 64* 103-137.

Horowitz, J.L. and S. Lee (2007). Nonparametric Instrumental Variables Estimation of a Quantile Regression Model," *Econometrica, 75* 1191-1208.

Johannes, J., Van Bellegem, S. and A. Vanhems (2007). A unified approach to solve ill-posed inverse problems in econometrics. Core Discussion paper,

2007/83.
Li, Q.I. and J.S. Racine (2006) *Nonparametric Econometric*, Princeton University Press, Princeton.

Linton, O., S. Sperlick and I. Van Keilegem (2008). Estimation of a Semi Parametric Transformation Model. *Annals of Statistics*, 36,pp. 686-718.

Mammen E. and Yu K., (2009). Nonparametric estimation of noisy integral equations of the second kind. *Journal of the Korean Statistical Society, 38(2)*, 99-110.

Newey, W. and J. Powell (2003). Instrumental Variable Estimation of Nonparametric Models. *Econometrica ,71* 1565-1578.

Rochet, Tirole, (2003). Platform competition in Two Sided Markets. *Journal of the European Economic Association, 1(4)* 990-1029.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*, London. Chapman & Hall.

APPENDIX I: Simulation results

AI 1 -Simulation in the non parametric model under exogeneity
This Appendix illustrates our procedures by a Monte Carlo simulation [5].

The exogenous variable Z is drawn from a $N(0, 1.2^2)$ and U is independently generated by a $N(0, 0.3^2)$. The survivor S is a logistic function ($S(t) = \frac{1}{1+e^{-t}}$) or equivalently $\varphi(y) = \ln \frac{1-y}{y}$. Figure 1 shows the sample (the y_i is on the horizontal axes and z_i on the vertical one), the true function and three estimations with naive bandwidth and arbitrarily values of α. The sample size is $n = 500$. Intuitively great values of α lead to a flat line and very small values to a curve oscillating around the true one.

The minimization of α is represented by the curve in figure 2 where the function $SS(\alpha)$ defined in (3.10) is represented (the corresponding estimation is represented in figure 3). In figure 4 same estimation using optimal α is represented for a smaller sample of 200. In these figures, bandwidth are naive bandwidth. We show in figure 5 the change of estimation by two recursive evaluations of the bandwidth by cross validation and by selection of optimal α. In all these first 5 figures, one draw of the sample only is treated.

Finally, 50 Monte Carlo replications of the model (where $n = 100$) are drawn and 50 curves are estimated (and represented in figure 6) using naive bandwiths and optimal α for each simulation. This figure illustrates graphically the distribution of the estimator. We have check the sensitivity of our results by modifying some assumptions. We first increased the variance of the error term in the equation (0.3^2 is replaced by 0.6^2). The results are very similar to the previous one and the Monte Carlo simulation for a sample size of 100 are represented in figure 6a. Secondly we have replaced the function φ by $\varphi(Y) = tg(2\pi\, Y)$. The design of Z is modified ($Z_i \sim N(0, 0.4^2)$) and the results are given in figure 6b by a Monte Carlo simulation. Here also results are very similar.
A theory for the joint determination of α and the bandwiths has not yet be developed. Our simulations experiments show that many couples α and h_n will give an identical result. If h_n is fixed arbitrarily in a suitable interval, α will "adapt" itself to give "good" results. This conjecture needs to be check theoretically.

[5] All the codes are available from the authors upon request

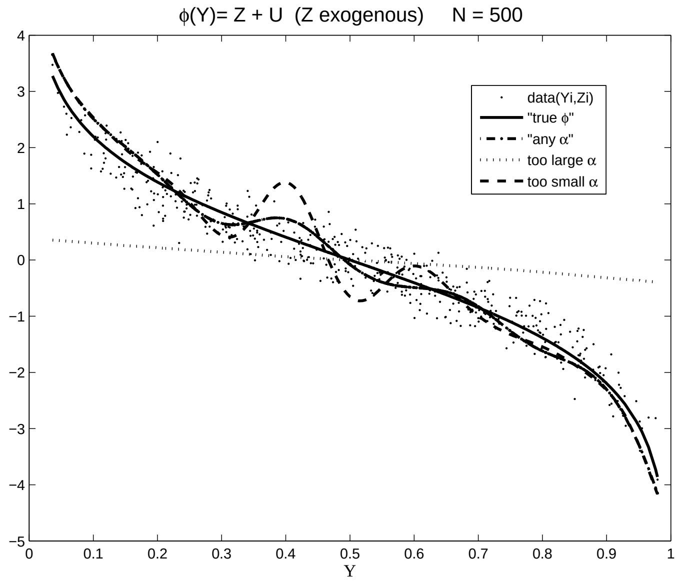

Figure 1: Estimation under different values of α (one draw)

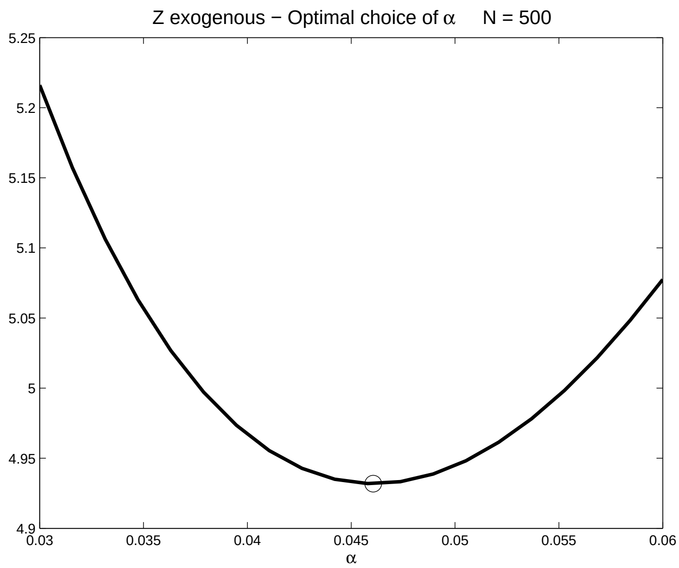

Figure 2: Representation of $SS(\alpha)$ (one draw) and selection of the minimum

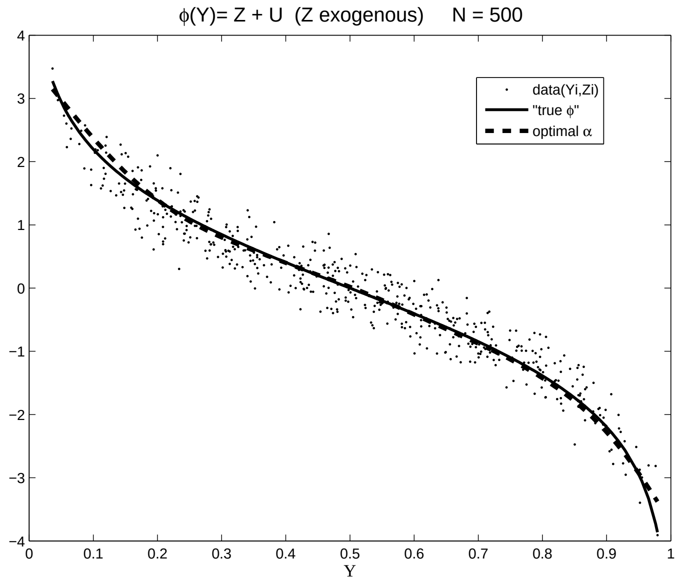

Figure 3: Estimation under data driven value of α

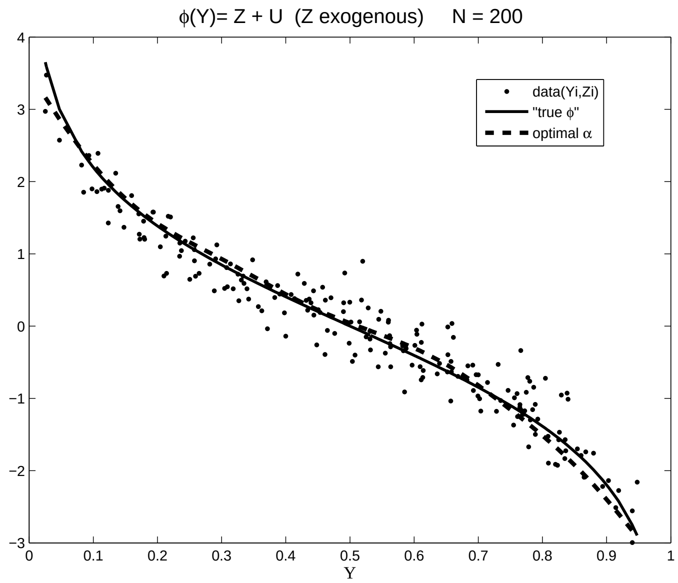

Figure 4: Estimation under data driven value of α (one draw, small sample)

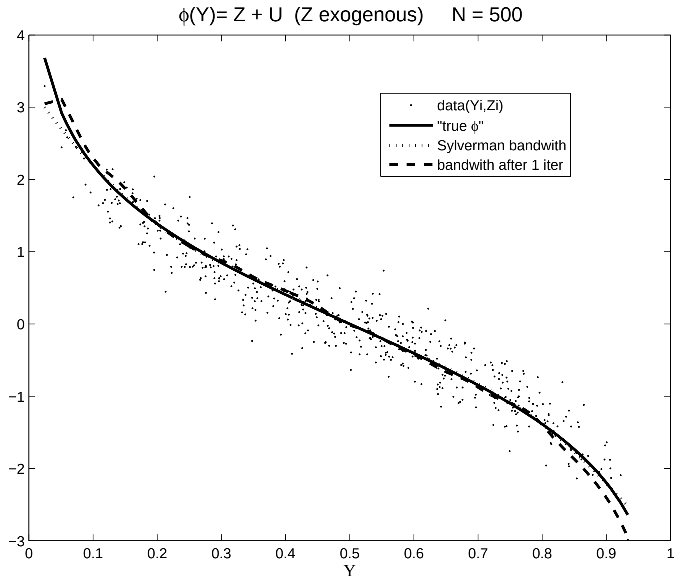

Figure 5: Estimation under two bandwiths (one draw)

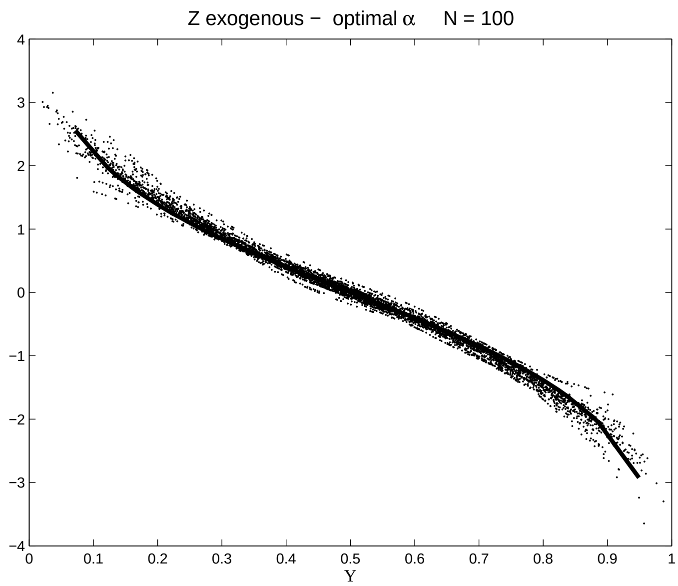

Figure 6: Monte Carlo simulations of the estimation

AI 2 - Simulation in the non parametric model under endogeneity
We illustrate section 5 extension by some simulations. The function $\varphi(y))$ is still equal to $\ln \frac{1-y}{y}$, U and R are independent and both $N(0;0,3^2)$. The variable Z is equal to $aR + bU + \varepsilon$ where $a = 2.5, b = 2, \varepsilon$ is $N(0, 0.015^2)$. Figure 7 to 9 have the same definitions as figure 1 to 3 but with Z endogenous. Figure 10 shows the impact of the bandwidth improvement. This graph concerns a single drawn of the data set but figure 11 shows the results of 50 monte carlo simulations with optimal α and selection of bandwidths by cross validation for each simulation (case $n = 100$).

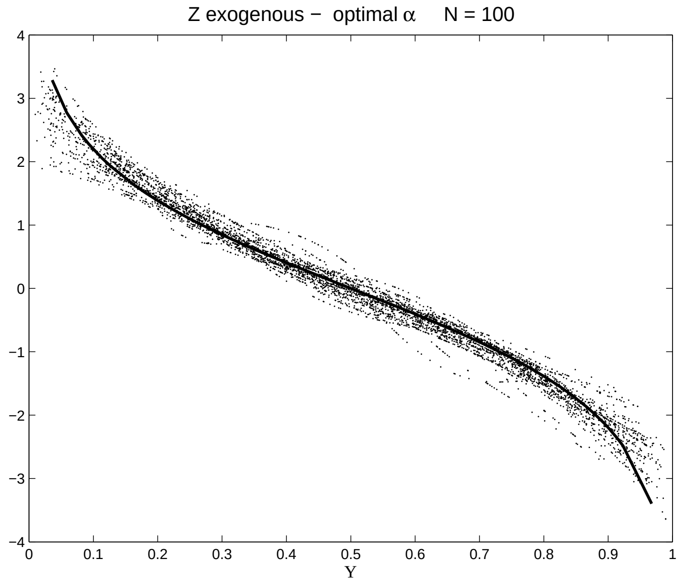

Figure 6a: Monte Carlo simulations of the estimation "largest" variance

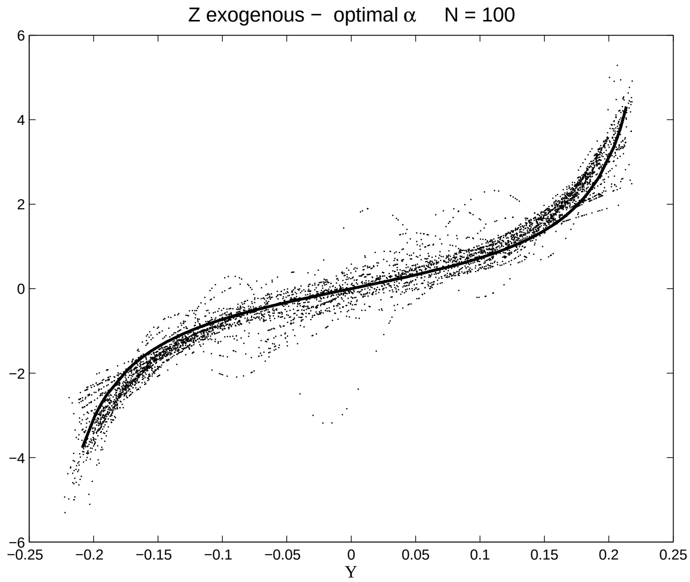

Figure 6b: Monte Carlo simulations of the estimation
Other specification

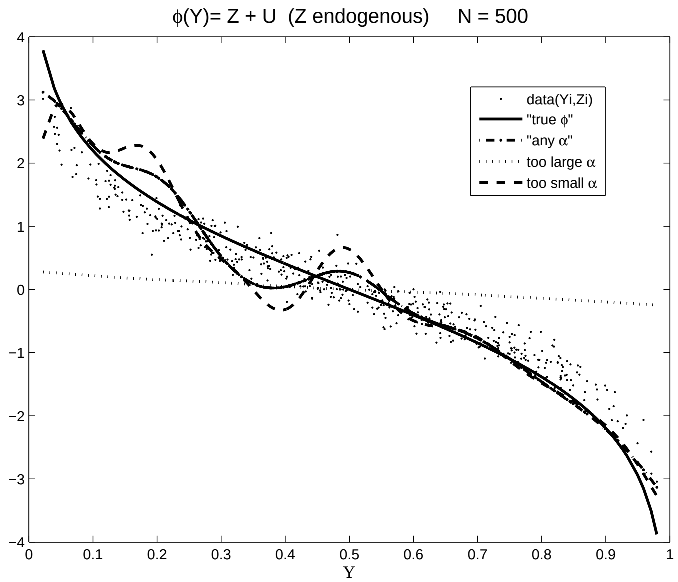

Figure 7: Estimation under different values of α (one draw)

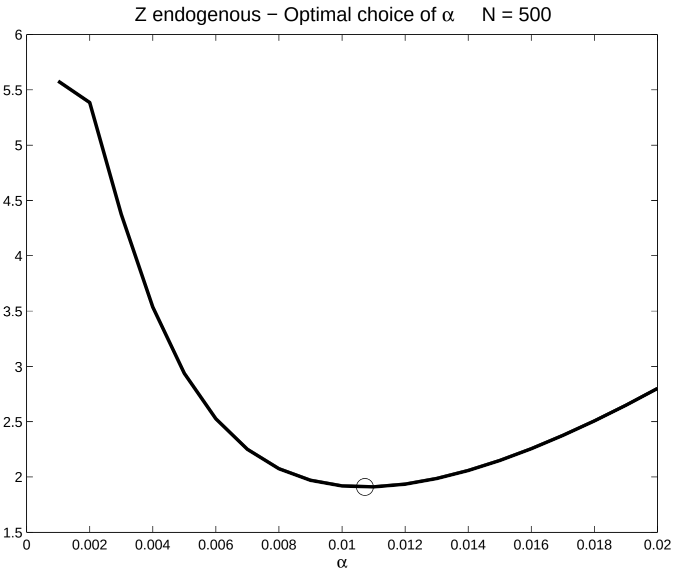

Figure 8: Representation of $SS(\alpha)$ (one draw) and selection of the minimum

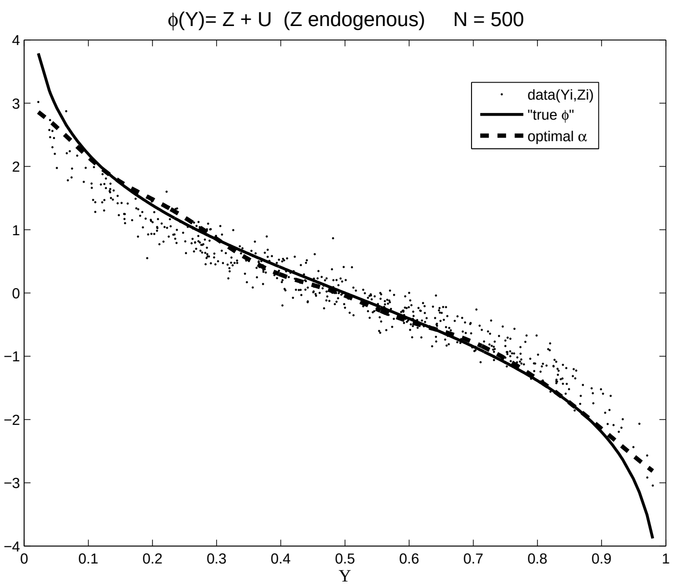

Figure 9: Estimation under the data driven selection of α (one draw)

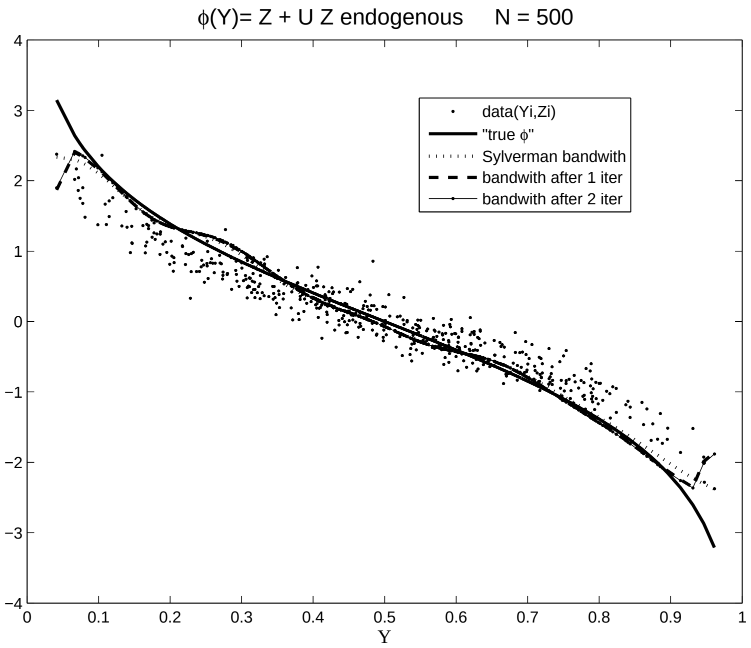

Figure 10: Estimation under different bandwiths (one draw)

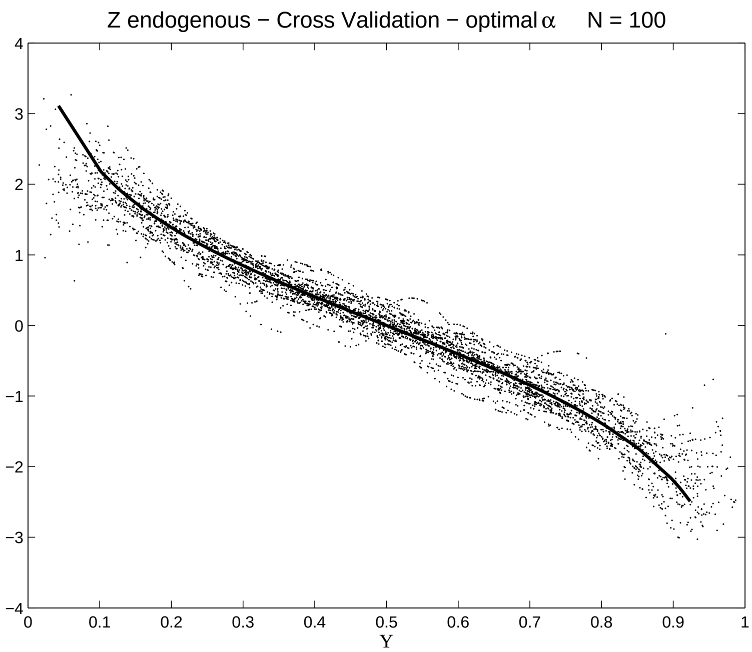

Figure 11: Monte Carlo simulations of the estimation of φ

AI 3 - Asymptotic properties of the estimator of the parametric β Two models have been simulated. In the exogenous case $n = 50$, $\varphi(y)$ and U remain identical, $Z \sim N(0, 0.8^2)$ $W \sim N(0,1)$ and $\beta = 0.5$. In the endogenous case, $n = 200$ $Z = aR + bU + \varepsilon$ where $a = 2,6$ $b = 2,1$. The others elements remain the same. In each case 50 monte carlo replications are generated and we represent the monte carlo distribution of the estimator of β with the values of the mean (figure 12 and 14). The figures 13 and 15 represent the different estimators of φ with naive bandwidth and optimal α for each simulation. Our conclusion deduced from the simulation concerning the bandwith and the regularization parameter is the following. The simultaneous choice of h_n and α is not "identified" in the sense that there probably exists a curve of h_n and α space such that each element gives the same result. In other terms, the selection procedure of α adapts to the choice of h_n (in a reasonable "set") in order to give a "good" result. This conjecture will be examined in the future works.

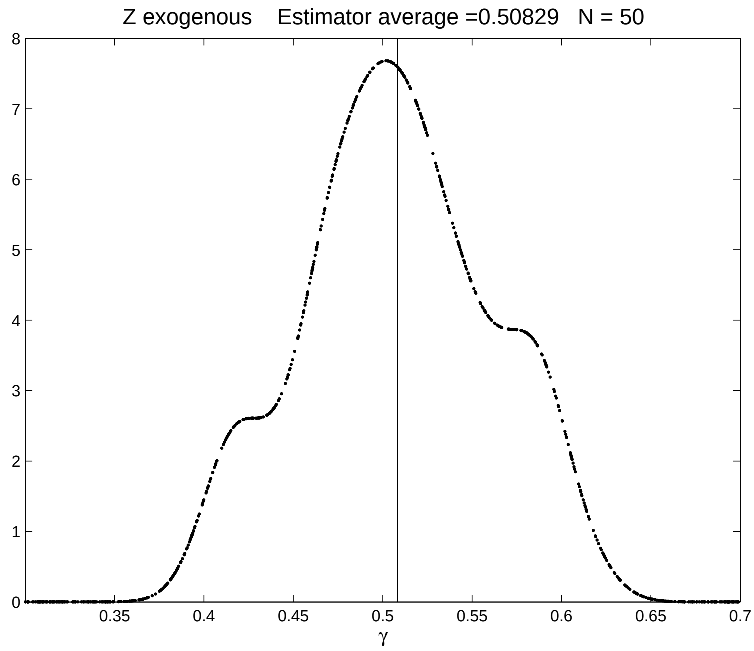

Figure 12: Monte Carlo distribution of β

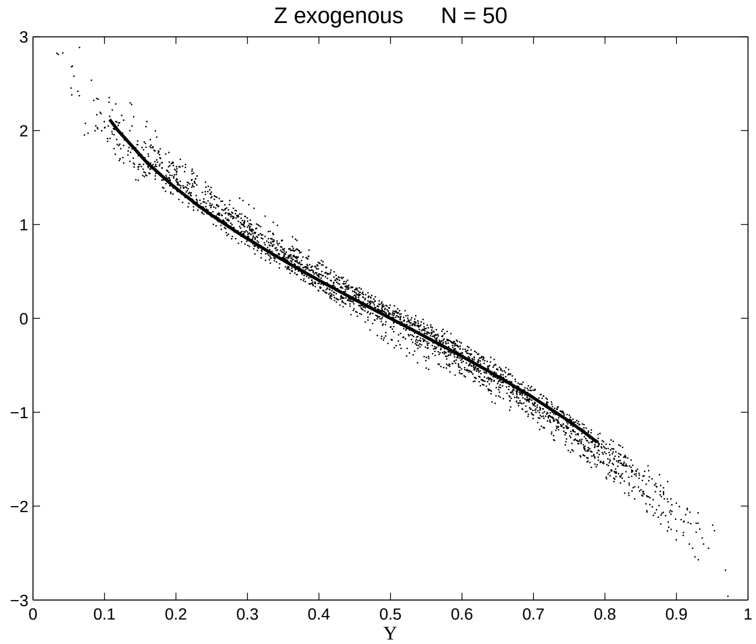

Figure 13: Monte Carlo simulations of the estimation of φ

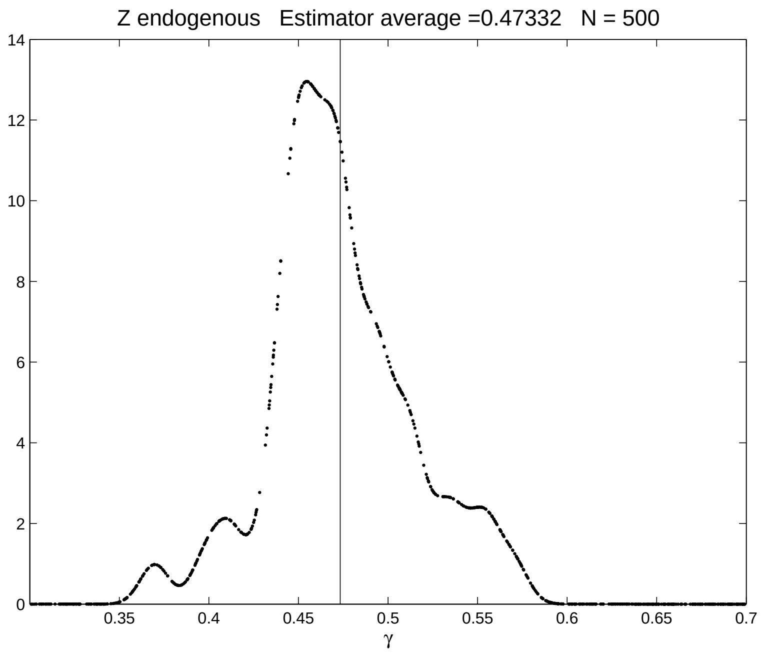

Figure 14: Monte Carlo distribution of β

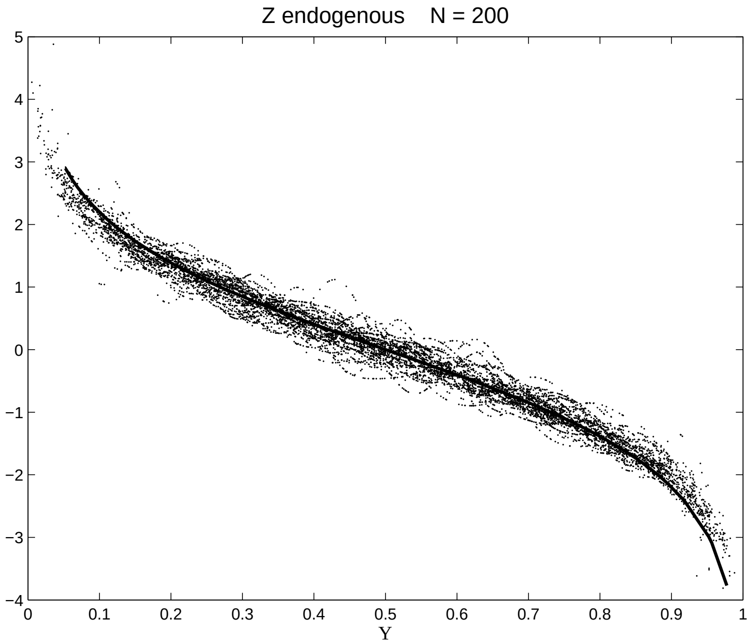

Figure 15: Monte carlo simulations of the estimation of φ.

APPENDIX II: Complements on asymptotic properties

This appendix will first give some details on the derivation of the rate result given in theorem 4.1.

AII 1 -Proof of Theorem 4.1

i) Let us first start with the following remark.

In the previous practical computations the kernel estimation of $E(\varphi(Y)|Z)$ was based on formulae (3.3), i.e.:

$$(\hat{T}\varphi)(z) = \frac{\displaystyle\sum_{i=1}^{n} \varphi(y_i) K\left(\frac{z - z_i}{h_n}\right)}{\displaystyle\sum_{i=1}^{n} K\left(\frac{z - z_i}{h_n}\right)}$$

The asymptotic theory we present in this section is actually based on a slightly different expression of $\hat{T}\varphi$, namely:

$$(\hat{T}\varphi)(z) = \frac{\displaystyle\sum_{i=1}^{n} \left[\int \varphi(y) \frac{1}{h_n} K\left(\frac{y - y_i}{h_n}\right) dy\right] K\left(\frac{z - z_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{z - z_i}{h_n}\right)}$$

and the same modification is done for $\hat{T}^*$. Actually this last expression is obtained by estimating $\int \varphi(y) f(y|z) dy$ (where f is the density of Y and Z) by replacing f by its kernel estimator.

This modification is motivated by the following argument. The first estimator defined above is a non bounded estimator. To see this point we can imagine two functions φ, and φ_2 very closed in the square norm sense $\left(E\left((\varphi_1(Y) - \varphi_2(Y))\right)^2\right.$ small) but such that $\hat{T}\varphi_1$ and $\hat{T}\varphi_2$ are very different. This unboundness property complicates the proofs. However the second estimator defined a bounded operator and $\hat{T}$ is continuous. To see the difference between the two computations we first remark that with this new expression the empirical counter part of (3.6) now becomes:

$$\alpha\varphi(y)+\frac{1}{\sum_{i=1}^{n}K\left(\frac{y-y_j}{h_n}\right)}$$

$$\sum_{j=1}^{n}K\left(\frac{y-y_j}{h_n}\right)\int\left[\sum_{i=1}^{n}\int\varphi(y)\frac{1}{h_n}K\left(\frac{y-y_i}{h_n}\right)dy\times\frac{K\left(\frac{z-z_i}{h_n}\right)}{\sum_{i=1}^{n}K\left(\frac{z-z_i}{h_n}\right)}\right]\times$$

$$\frac{1}{h_n}K(\frac{z-z_j}{h_n})dz$$

$$=\frac{\sum_{j=1}^{n}z_jK\left(\frac{y-y_j}{h_n}\right)}{\sum_{j=1}^{n}K\left(\frac{y-y_j}{h_n}\right)}$$

Let us multiply this equation by $\frac{1}{h_n}K\left(\frac{y-y_\ell}{h_n}\right)$ $(\ell=1,...,n)$. After integration of the two sides of the equation with respect to ℓ we get the same system as in (3.8) except that $K\left(\frac{y_\ell-y_j}{h_n}\right)$ and $K\left(\frac{z_j-z_i}{h_n}\right)$ are now replaced by

$$\int\frac{\frac{1}{h_n}K\left(\frac{y-y_\ell}{h_n}\right)K\left(\frac{y-y_j}{h_n}\right)}{\sum_j K\left(\frac{y-y_j}{h_n}\right)}dy \text{ and } \int\frac{\frac{1}{h_n}K\left(\frac{z-z_j}{h_n}\right)K\left(\frac{z-z_i}{h_n}\right)}{\sum_i K\left(\frac{z-z_i}{h_n}\right)}dz$$

These approximations introduce errors with the same magnitude as the bias of the kernel and then they may be neglected.

ii) Let us now come back to formulae (4.1). First we may remark that $(\alpha I+\hat{T}^*\hat{T})$ is invertible for n sufficiently large. Indeed assumption section 4 implies that $\|\hat{T}\hat{T}-TT\|^2$ goes to 0 which imply that the eigen values of $\alpha I+\hat{T}^*\hat{T}$ converges uniformly to the eigen values of $\alpha I+T^*T$. These eigen values have the form $\alpha+\lambda_j^2$ where λ_j^2 is a (positive) eigen value of T^*T and are then strictly positive. This property is then also true for $\alpha I+\hat{T}^*\hat{T}$.

iii) Now consider $\|\hat{\varphi}^\alpha-\varphi\|^2$. We want to analyze the rate of the convergence to zero of this norm.

We have

$$\begin{array}{rcl} \hat{\varphi}^\alpha - \varphi & = & (\alpha I + \hat{T}^*\hat{T})^{-1}\hat{T}^* Z - \varphi \\ & = & (\alpha I + \hat{T}^*\hat{T})^{-1}\hat{T}^* Z - (\alpha I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\hat{T}\varphi \\ & + & (\alpha I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\hat{T}\varphi - (\alpha I + T^*T)^{-1}T^*T\varphi \\ & + & (\alpha I + T^*T)^{-1}T^*T\varphi - \varphi \\ & = & A + B + C \end{array}$$

From the properties of a norm

$$\|\hat{\varphi}^\alpha - \varphi\| \leq \|A\| + \|B\| + \|C\|$$

Let us consider the first term:

$$\|A\| = \|(\alpha I + \hat{T}^*\hat{T})^{-1}(\hat{T}^* Z - \hat{T}^*\hat{T}\varphi)\|$$

We used here these properties:

$$\|(\alpha I + \hat{T}^*\hat{T})^{-1}\hat{T}^*(Z - \hat{T}\varphi)\| \leq \|(\alpha + \hat{T}^*\hat{T})^{-1}\hat{T}^*\|\|Z - \hat{T}\varphi\|$$

The first norm $\|(\alpha I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\|$ is equal to the larger eigen value of the operator. These eigen values converges to $\frac{\lambda_j}{\alpha+\lambda_j^2}(\lambda_j = \sqrt{\lambda_j^2})$ and are then smaller than $\frac{1}{\sqrt{\alpha}}$. Using the assumption i) of section 4 we get that

$$\|(\alpha I + \hat{T}^*\hat{T})^{-1}\hat{T}^*(Z - \hat{T}\varphi)\|^2 \sim O\left(\frac{1}{\alpha}\left(\frac{1}{nh_n} + h_n^4\right)\right)$$

Using elementary algebra, the second term B verifies:

$$B = (\alpha I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\left[(\hat{T} - T) + (\hat{T}^* - T^*)\right] T\alpha(\alpha I + T^*T)^{-1}\varphi$$

We have first remarked that $\|(\alpha I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\| = O(\frac{1}{\sqrt{\alpha}})$ and that $\|\hat{T} - T\|$ or $\|\hat{T} - T\|$ are $O\left(\frac{1}{\sqrt{n}\,h_n} + h_n^2\right)$. The last term, identical to $\alpha(\alpha I + T^*T)^{-1}T^*\varphi$ is the regularity bias of $T^*\varphi$ equal to $O(\sqrt{\alpha^{\min(\beta+1,2)}})$. Then:

$$\|B\|^2 = O\left(\frac{1}{\alpha}\left(\frac{1}{nh^2} + h^4\right)\alpha^{\min(\beta+1,2)}\right)$$

Finally we have seen in section 4.1 that φ is assumed sufficiently regular

$$\||C\||^2 = \alpha^\beta. \quad \blacksquare$$

AII 2 - Speed of convergence of the data driven selection of α

Let us consider the main elements of the proof. if n is large $SS(\alpha)$ defined in (3.10) is almost equal to:

$$\frac{1}{\alpha}\| \hat{T}\ \widehat{\varphi}^{\alpha}_{(2)} - Z \|^2$$

$$\begin{aligned}
\hat{T}\ \hat{\varphi}^{\alpha}_{(2)} - Z &= \hat{T}(\alpha I + \hat{T}^*\hat{T})^{-1}[\hat{T}^* + \alpha(\alpha I + \hat{T}^*\hat{T})^{-1}\hat{T}^*](Z - \hat{T}\varphi) \\
&+ \hat{T}(\alpha I + \hat{T}^*\hat{T})^{-1}[\hat{T}^* + \alpha(\alpha I + \hat{T}^*\hat{T})^{-1}\hat{T}^*]\hat{T}\varphi \\
&= A + B
\end{aligned}$$

as $\| \hat{T}(\alpha I + \hat{T}^*\hat{T})^{-1}\hat{T}^* + \alpha(\alpha I + \hat{T}^*\hat{T})^{-1}\hat{T}^* \|$ is bounded.

$$\| A \|^2 = O\ (\tfrac{1}{nh_n} + h_n^4)$$

The second term is equal to $B = B_1 + B_2$ where

$$B_1 = T(\alpha I + T^*T)^{-1}(T^* + \alpha(\alpha I + T^*T)^{-1}T^*)T\varphi$$

and

$$\begin{aligned}
B_2 &= \hat{T}(\alpha I + \hat{T}^*\hat{T})^{-1}(\hat{T}^* + \alpha(\alpha I + \hat{T}^*\hat{T})^{-1}\hat{T}^*)\hat{T} \\
&- T(\alpha I + T^*T)^{-1}(T^* + \alpha(\alpha I + T^*T)^{-1}T^*)T
\end{aligned}$$

$\| B_1 \|^2$ is the regularization bias of $T\varphi$ equal to $\alpha^{\beta+1}$ if $\beta \leq 2$ (see Engl et al. (2000). The last term is negligible using arguments identical to the end of the proof of the theorem 4.1 but based on the algebra of iterated Tikhonov regularization.

Then

$$\frac{1}{\alpha}\| \hat{T}\ \widehat{\varphi}^{\alpha}_{(2)} - Z \|^2 \sim\ O\ (\frac{1}{\alpha}(\frac{1}{nh_n} + h_n^4) + \alpha^\beta)$$

and the minimization of this expression gives an α which converges to zero at the optimal rate. $\blacksquare$