Estimation of the solution of a differential equation with endogenous effect

J-M. Loubes and A. Vanhems

November 28, 2004

Abstract

The objective of this paper is to study the statistical properties of solutions of a differential equation which depends on the data set and whose underlying random variables are endogenous. The problems of endogeneity are numerous in economic fields and we will briefly motivate our study by an application in demand theory. We have two problems to solve: to smooth the data set with endogenous variables and to solve the differential equation. We show how solving a differential equation can improve the properties of a nonparametric estimator. The estimated solution depends on two smoothing parameters: the bandwidth parameter of the kernel method and the regularization parameter of the Tikhonov methodology. We present results on the consistency and the optimal choice of the parameters.

Keywords: Inverse Problems, Kernel estimator, Endogenous effect, Instrumental variable, Econometrics
Subject Class. MSC-2000 : 63G07, 34K29 .

Contents

1 Introduction

1.1 Presentation of the problem

Studying the solutions of differential equations depending on the data set is a very common problem in statistics, on both theorical and practical point of view. Indeed, many interest parameters in economics, physics, or finance are defined as the solutions of a differential equation. They can be estimated using nonparametric kernel analysis, which has been developed by Vanhems [Van01] or other methods, see for

instance [] or []. Here, in the case of endogeneity of the underlying random variables, we build an fully tractable estimator of the solution of the differential equation and give its asymptotic behaviour.
More precisely, consider a random vector (Y, Z, W) which follows an unknown cumulative distribution function F. The function m is defined by the following relation:

$$\begin{cases} Y & = m\left(Z\right) + U \\ E\left(U \left| W\right.\right) & = 0 \end{cases} \tag{1}$$

where Z is the endogenous variable and W is the instrument, chosen such that $(Z, W) \in L^2\left(I\!R^2\right) \times L^2\left(I\!R^2\right)$. We have set $Z = \left(Z^1, Z^2\right)$ (resp. $W = \left(W^1, W^2\right)$).
The analysis of endogenous regressors, and more generally of simultaneity, has a great impact in structural econometrics. Since the earliest works of Amemiya in [Am74] and Hansen in [Han82], extensions to nonparametric and semiparametric models have been considered. Estimating nonparametric models with endogenous explanatory variables has been in particular studied by Darolles, Florens, Renault [DFR01], Blundell and Powell in [blund00] and Florens in [Flo00]. Our objective is therefore to introduce some nonparametric instrumental regression into a differential equation and to study the asymptotic properties of the associated estimated solution.
As a matter of fact, the interest parameter is the function λ solution of:

$$\begin{cases} \lambda'(x) & = m\left(x, \lambda(x)\right) \\ \lambda\left(x_0\right) & = \lambda_0,\ \ (x_0, \lambda_0) \in I\!R^2 \end{cases} \tag{2}$$

Note that λ is defined by an implicit nonlinear relation (there is no restrictive assumption on the form of the function m). The second equation represents the initial condition in order to have uniqueness of a solution in a neighborhood of (x_0, λ_0). For sake of simplicity and without loss of generality, throughout all the article, we will assume that $x_0 = \lambda_0 = 0$.
Hence we will first estimate non parametrically the function m by $\widehat{m}_n$. Then we will find the solution of the estimated differential equation:

$$\begin{cases} \lambda'(x) & = \widehat{m}_n\left(x, \lambda(x)\right) \\ \lambda\left(x_0\right) & = \lambda_0, (x_0, \lambda_0) \in I\!R^2 \end{cases} \tag{3}$$

Finally, we aim at studying the convergence of the estimated solution to the true one, under natural conditions of existence and uniqueness.
The difficulty in of this work lies in the fact that we are facing two inverse problems. Inverse problems have been intensively studied by several authors. For general references, we refer to the following papers [JS90] [CT00] [CIK99] [Erm89] or [Osu96]. Contrary to the first one (1), the inverse problem (2) is well-posed in the sense of Tikhonov. But in the case studied here, we approximate this problem by the problem (3), which is an inverse problem where the operator is unknown, depending on the efficiency of the estimator of the first inverse problem set in (1). The estimation of $\hat{m}_n$ is tackled in the work of Darolles, Florens and Renault [DFR01]. They construct a nonparametric kernel estimator of the equation (1). They show in particular that this problem is ill-posed in the sense of Tikhonov []. That is the reason why they construct, using Tikhonov regularization method, an estimator of m: $\widehat{m}_{n,\alpha_n}$ where α_n is a smoothing regularization term necessary to transform the initial problem into a well-posed one. They study the asymptotic properties of this estimate using kernel approximation results. We will use their results as a starting point for the construction of the solution of (3). We will be interested in particular to compare our asymptotic results with the previous ones.

1.2 Application in microeconomics

Let us now present an example of application in microeconomics. It is taken from an article by Hausman and Newey in [HN]. The objective is to measure the impact on the consumer welfare of a price change for one good. Therefore, we consider one consumer; we define y its income, q the demand in good and p^1 the price of a unique good. We assume that there exists a price variation from p^0 to p^1. A way to capture the impact on the consumer is to calculate the variation of exact consumer surplus λ: it represents the cost to pay to the consumer so that his welfare does not change for a price change. It is a monetary measure of the variation of utility (see Varian in [Var92]).
Our first objective is to find a relation that links the functions λ and q. For that purpose, let us consider a price path $p\left(t\right), t \in [t_0, t_1]$ where $p\left(t_0\right) = p^0$ and $p\left(t_1\right) = p^1$ and $\lambda\left(p\left(t\right)\right)$ is the variation of

exact consumer surplus between $p(t)$ and $p(t_1)$. Then, we can derive the following relation between our interest parameter λ and the demand function q:

$$\begin{cases} \lambda'(p(t)) & = -q(p(t), y - \lambda(p(t))) . p'(t) \\ \lambda(p(t_1)) & = 0 \end{cases}$$

assuming that p^1 is a price reference, that is all price variations are calculated with respect to p^1. By a change of variable, we find that:

$$\begin{cases} \lambda'(p) & = -q(p, y - \lambda(p)) \\ \lambda(p^1) & = 0 \end{cases}$$

This is clearly a particular case of differential equation of order one.

We can then present some econometric model to estimate the demand function q:

$$\ln q = \ln m(p, y) + \varepsilon$$

Assuming that $I\!E[\varepsilon | p, y] = 0$ is quite a common assumption to make in order to estimate the function m by a simple regression. However, such a simplified hypothesis is often not realistic and the price is usually an endogenous variable. Therefore, the interesting case to model is to consider that $I\!E[\varepsilon | p, y] \neq 0$. Since the function m is no more identified, we need to introduce some instrumental variable w to solve this model. So, we have the following underlying econometric model to solve:

$$\begin{cases} \ln q & = \ln m(p, y) + \varepsilon \\ I\!E[\varepsilon | w] & = 0 \end{cases}$$

Moreover, a similar problem could be studied for the variation of firm profit and the variation of demand factor.

Therefore, this paper proceeds in the following way. In the following section, we set the mathematical framework of the problem and define the estimators we will use. Their asymptotic behaviour is given in Section 3. Section 4 is devoted to some auxiliary lemmas which enable to prove the asymptotic behavior of the estimator. All the proofs are gathered in Section 5.

2 Statistical framework of the problem

Note first that all the asymptotic results will be given using the L^2 norm which will be written $\|.\|$. The different other norms will be clearly specified.

We recall that the statistical model is the following: we estimate a function λ solution of a differential equation

$$\begin{cases} \lambda'(x) & = m(x, \lambda(x)) \\ \lambda(0) & = 0 \end{cases} \tag{4}$$

where the function m is unknown but is observed in the following framework:

$$Y_i = m(Z_i) + U_i,\ i = 1, \ldots, n \tag{5}$$

We assume that

Assumption over the observations : the random variables Y, Z, W take values in a compact set of $\mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q$. The observations Y_i, Z_i, W_i, $i = 1, \ldots, n$ are an iid sample with density f.

Without constraint over the residual term, the function m is not properly characterized and the inverse problem (4) is not identifiable. In the classical regression scheme, one often assumes that $\mathbf{E}(U|Z) = 0$ or equivalently that $m(Z) = \mathbf{E}(Y|Z)$. However, such assumption is not appropriate in structural econometric models defined in Section 1. As a matter of fact, the relationship between Y and Z is characterized by a third variable W , called an instrumental variable. More precisely, we assume that there exist variables $W = (W_1, W_2)$ such that

$$\mathbf{E}[Y - m(Z)|W] = 0 \tag{6}$$

Instrumental variables analysis has been introduced by Reiersol in [R41] [R45] or Sargan in [S58]. This definition was also the starting point of the analysis of simultaneity in linear or parametric nonlinear models [Am74]. Moreover, the problem of estimating a regression function in this particular setting was studied by Darolles, Florens and Renault in [DFR01].

The estimation method in our work is the following two-step method:

- First, we construct an estimator of m based on a kenerl estimator for inverse problem. Indeed, write F the joint distribution of $S = (Y, Z, W)$ with density function f, and define the following operators:

$$\begin{aligned} T = T_F : L^2(Z) &\longrightarrow L^2(W) \\ g &\longrightarrow T_F(g(Z)) = \mathbf{E}[g(Z)|W] \end{aligned}$$

$$\begin{aligned} T^* = T_F^* : L^2(W) &\longrightarrow L^2(Z) \\ h &\longrightarrow T_F^*(h(W)) = \mathbf{E}[h(W)|Z] \end{aligned}$$

$$r = r_F(W) = \mathbf{E}(Y|W)$$

Such operators are the conditional expectations with respect to the variables W and Z. These linear operators satisfy:

$$< T(\phi(Z)), \psi(W) >=< \phi(Z), T^*(\psi(W)) >= \mathbf{E}(\phi(Z)\psi(W)).$$

Hence, consider the instrumental regression setting (5), which can be written in the following equivalent way

$$Tm - r = 0.$$

This is actually an equation of the type

$$A(m, F) = 0.$$

It is well known that the initial problem is ill-posed. That is the reason why, following ideas from [DFR01], we use a regularization method called Tikhonov regularization [TA77] and transform the original problem into:

$$(\alpha_n I + T^*T)m_{\alpha_n} = r^* \tag{7}$$

for α_n a given positive sequence such that $\alpha_n \to 0$, at a rate that will be made precise later in this paper, and $r^* = T^*r$.

Consider $\hat{F}$ a kernel estimator of F defined through its density with respect to Lebesgue measure

$$\hat{f}_n(y, z, w) = \frac{1}{n}\sum_{i=1}^{n} K_{y,h_n^y}(y - Y_i)K_{z,h_n^z}(z - Z_i)K_{w,h_n^w}(w - W_i) \tag{8}$$

where K_y, K_z, K_w are three kernels and h_n^y, h_n^z, h_n^w are three bandwidths. Now, set the associated operators $\hat{T} = T_{\hat{F}}$, $\hat{T}^* = T_{\hat{F}}^*$ and $\hat{r} = r_{\hat{F}}$.

Definition 2.1. The estimator we will consider it $\hat{m}_{n,\alpha_n}$ the solution of

$$(\alpha_n I + \hat{T}^*\hat{T})\hat{m}_{n,\alpha_n} = \hat{r}^*. \tag{9}$$

It is the Tikhonov regularized of an inverse problem whose operator is estimated using kernel estimators.

Remark 2.2. It is interesting to notice that this transformation can be seen as a penalized minimization problem:

$$m_{\alpha_n} = \arg\min_{f\in L^2(Z)} \left(\|T(f) - r\|^2 + \alpha_n\|f\|^2\right)$$

Using this particular expression, the estimator can be viewed as a penalized M-estimator, see for instance Loubes and van de Geer in [LvdG00]. This point of view could also lead to different estimators obtained by changing the quadratic penalty into an l^1 penalty. It gives rise to non linear estimators such as thresholded estimators.

- The second step consists in plugging this particular estimator in the differential equation in the following way.
The differential equation (4), with the initial condition $\lambda(0) = 0$, can be solved using Cauchy-Lipschitz's theorem. This classical theorem provides existence and uniqueness of a solution λ in a compact neighborhood of the initial condition $(0, 0)$. This theorem holds as soon as the Lipschitz condition is satisfied for the function m, which is clearly the case assuming that m is, for example, continuously differentiable of order 1. So, the idea is to replace our implicit definition of the functional λ by an explicit relation that links λ and m and this can be done since we are under the assumptions of existence and uniqueness of a solution. Actually, the implicit function theorem for infinite dimension spaces shows that, under some assumptions, there exists a unique solution λ which depends on the function m. As a consequence, we introduce an operator Φ such that the solution has the following form
$$\lambda(x) = \Phi[m](x) \tag{10}$$
As an estimator of the parameter of interest λ, we take:
$$\hat{\lambda}(x) = \Phi[\hat{m}](x) \tag{11}$$
In order to get a well-posed inverse problem, we need to check that the function Φ is continuous. Moreover, it is possible to prove that Φ is continuously differentiable. All this problem is one of the main issues tackled in [Van01]. Therefore, once we are able to derive the properties of the function m and its estimator $\hat{m}$, given by the relation 6, and using an adaptation of the results from [Van01], we will be able to study the properties of our interest parameter λ and its estimator $\hat{\lambda}$.

One inverse problem or two nested problems ?

We do not observe directly the function m, since $\mathbf{E}(U|Z) \neq 0$, but only the relationship with the instrumental variable W:
$$\mathbf{E}(Y|W) = \mathbf{E}(m(Z)|W).$$
As a result, the estimation of Tm is a necessary step in the estimation process. Since the operators m and T do not commute, there is no way to look for an estimator of the form: $\widehat{\Phi[m]}$. That is the reason why we consider the estimator $\Phi[\hat{m}]$, constructed with a preliminary estimator of the function m. As a result, it turns impossible to write the estimation problem into a single estimation, but it must be studied following the two steps described above. Therefore, using a methodology in two steps to solve the two inverse problems 4 and 6 is unavoilable.

3 Main Results

In this section, we aim at giving the asymptotic behaviour of the solution of the differential equation obtained after estimating the regression function observed in an endogenous settings. Hence we begin to prove that the following expansion holds

$$\begin{aligned} \left(\widehat{\lambda}_{n,\alpha_n} - \lambda\right)(x) &= \left(\Phi\left[\hat{m}_{n,\alpha_n}\right] - \Phi\left[m\right]\right)(x) \\ &= d\Phi[m](\hat{m}_{n,\alpha_n} - m)(x) + R \\ &= H(\widehat{m}_{n,\alpha_n} - m)(x) + R \end{aligned} \tag{12}$$

where $d\Phi$ represents the Frechet-derivative of Φ, and having set $d\Phi[m] = H$. The residual term R is chosen at an order higher than the rest of the usual Taylor development, that is: $R = O_P\left(\|\widehat{m}_{n,\alpha_n} - m\|_\infty^2\right)$. Just note that if we assume that the function Φ is continuously differentiable of order 2, this assumption is automatically satisfied, and is not a constraint any more. Introducing this expansion enables us to transform the nonlinear problem into a linear one, up to a residual term R. Hence the rate of convergence of $\widehat{\lambda}_{n,\alpha_n}(x)$ towards $\lambda(x)$ can be deduced from the the two terms:

- the linear part $d\Phi[m](\hat{m}_{n,\alpha_n} - m)(x)$.
- the second term R, which is the counterpart in the Taylor expansion.

3.1 Assumptions

The assumptions are of three kinds

- Assumptions ensuring the convergence of the preliminary estimator. Such assumptions are similar to the one given in [DFR01].
- Assumptions necessary to linearize the problem using Taylor's expansion. This enables to compute the rate of convergence of the solution of the estimated estimator from the rate of convergence of the estimate of the regression with endogenous effect.
- Assumptions that enforce the convergence of the unknown inverse operator and enable us to provide the rates of convergence of the estimator $\widehat{\lambda}_{n,\alpha_n}$.

Assumptions A:
$[A1] : m \in C^2(D)$,
$[A2] : \hat{m}_{n,\alpha_n} \in C^2(D)$,
$[A3] : \|D_2\hat{m}_{n,\alpha_n} - D_2 m\|_\infty \to 0$.

Remark 3.1. Assumptions $[A1]$ and $[A2]$ provide existence and uniqueness of $\widehat{\lambda}_{n,\alpha_n}$ and λ, assumption $[A3]$ provides stability of the solution.

Assumptions A':
$[A'1] : f \in C^2(\mathbb{R}^5)$,
$[A'2]$: K is a function of $I\!R$ into $I\!R$; $\int_{-\infty}^{\infty} K(x)\,dx = 1$, K is of order $d \geqslant 2$; K is continuously differentiable up to order 2 and its derivatives of order up to 2 are in $I\!L^2(I\!R)$,
$[A'3]$: As $n \to +\infty : h_n \to 0, \frac{nh_n^d}{\log^2 n} \to \infty, \frac{nh_n^{\frac{d}{2}+2}}{\log^2 n} \to \infty, \quad h_n^d \leqslant \alpha_n, \quad \frac{\log n}{\sqrt{n}h_n^{\frac{d}{2}+2}} = o(\alpha_n)$,
$[A'4] : \|m_{\alpha_n} - m\|_2 = O\left(\alpha_n^\beta\right)$, $\|m_{\alpha_n} - m\|_\infty = O\left(\alpha_n^\beta\right)$ and $\|D_2[m_{\alpha_n}] - D_2[m]\|_\infty \to 0$,
$[A'5]$: Hilbert-Schmitt assumption

$$\|T_F - T_{\hat{F}}\| = O_P\left(\sqrt{\frac{1}{\sqrt{nh_n^d}} + h_n^{2\rho}}\right),$$

where $\rho = \min(s,t)$.
This assumption is fulfilled as soon as we have

- $f(y,z,w)$ the joint density is s times continuously differentiable
- $f(y,z,w)$ is bounded from below
- $$\int \frac{f^3(y,z,w)}{f(y,.,.)f(.,z,.)f(.,.,w)} dydwdz < +\infty.$$
- the three kernels $K_{\bullet,\bullet}$ are taken r times differentiable and with the same bandwidth h_n.

3.2 Linearization of the inverse problem

First, we give the expression of the operator H and the conditions of its existence. Indeed, under some regularity conditions on the function m, Cauchy-Lipschitz theorem ensures existence and uniqueness of a solution for the system 4 in a neighborhood of the initial conditions $(0,0)$. More precisely, set $\mathbb{D}$ a compact neighborhood of $(0,0)$ and I a compact neighborhood of 0 of the following form:

$$\begin{aligned} I &= \{x, |x| \leqslant a\}, a > 0 \\ \mathbb{D} &= \{(x,y), |x| \leqslant a, |y| \leqslant b\}, b > 0 \end{aligned}$$

Moreover, we will denote by $\left(C^2(D), \|.\|_\infty\right)$ the Banach space of continuously differentiable functions of order 2 defined on $\mathbb{D}$, with the supremum topology. Previous assumptions on m and $\hat{m}_{n,\alpha_n}$ enable us to define unique solutions $\widehat{\lambda}_{n,\alpha_n}$ and λ. Therefore, under $[A1] - [A3]$, 4 is a well-posed inverse problem and we can write $\lambda = \Phi[m]$ and $\widehat{\lambda}_{n,\alpha_n} = \Phi[\widehat{m}_{n,\alpha_n}]$. Moreover, we are able to provide the exact expression of the first term of the Taylor expansion. Write $D_2[.]$ the derivative with respect to the second variable.

Theorem 3.2. *Under the following assumptions (A), $\forall x \in I$, we have:*

$$d\Phi[\widehat{m}_{n,\alpha_n} - m](x) = H[\widehat{m}_{n,\alpha_n} - m](x) = \int_0^x (\widehat{m}_{n,\alpha_n} - m)(t, \lambda(t)) \exp\left(\int_t^x D_2 m(u, \lambda(u))\, du\right) dt$$

Write

$$v(x,t) = \exp\left(\int_t^x D_2 m(u, \lambda(u))\, du\right)$$

a given continuous function. Then we have:

$$H[\widehat{m}_{n,\alpha_n} - m](x) = \int_0^x (\widehat{m}_{n,\alpha_n} - m)(t, \lambda(t)).v(x,t)\, dt$$

Proof. The proof of this theorem can be found in [Van01]. In the appendix, for sake of completness, we recall the guidelines of the proof. □

Moreover, the operator H is a one to one integral operator and can be written as

$$\forall g \in \mathcal{C}(\mathbb{D}), \forall x \in I, H[g](x) = \int_0^x v(x,t) g(t, \lambda(t))\, dt. \tag{13}$$

This fundamental property is stated in Lemma 4.1.

Remark 3.3. Write $r-1$ the maximum regularity of the function m. Since H is an integral operator, the maximum regularity of $H[m]$ is r. As a result, the rate of convergence of the estimated solution of the differential equation (4) is expected to be greater than the rate of convergence of the estimator of the function m since there is a gain in regularity. That is the reason why we study directly the term $H(\hat{m}_{n,\alpha_n} - m)$ and do not use rough upper bounds implying the term $\hat{m}_{n,\alpha_n} - m$. Moreover, we also expect a gain in dimension since we transform a function of two arguments into a function in one argument.

3.3 Asymptotic behaviour of solution of estimated differential equation

Our aim is to prove consistency and give rates of convergence for the estimator of the solution of the differential equation (4). The following theorem holds.

Theorem 3.4. *Under the assumptions A', we can take $d = 3$. Then, the estimator $\hat{\lambda}_{n,\alpha_n}$ is consistent and, for an optimal choice $h_n \approx n^{-\frac{1}{6}}$ and $\alpha_n \approx n^{-\frac{1}{(2+\beta)}}$, its rate of convergence is given by:*

$$\mathbf{E}\|\hat{\lambda}_{n,\alpha_n} - \lambda\|^2 = O\left(\sup\left[n^{-\frac{\beta}{2+\beta}}, n^{-\frac{2\beta}{2+\beta}+\frac{1}{3}}\right]\right) \tag{14}$$

The rate of convergence depends on two terms: an approximation term which corresponds to the smoothing problem and leads to an optimal choice of a sequence α_n, and an error term which corresponds to the estimation issue and leads to an optimal choice of the kernel's bandwidth h_n. Of course, there is a trade-off between these two contributions that must be solved. The gain in the rate of convergence is obtained in the choice of the optimal bandwidth since, more precisely, the regularization enables us to choose a smoother kernel which leads to a gain in the bias term. Indeed we can choose h_n^6 instead of h_n^4 without the regularization operator.

Proof of Theorem 3.4:

Proof. The proof divides into 3 steps. First we show that the rate of convergence of $\hat{\lambda}_n$ is deeply related to the rate of convergence of $\hat{m}_n$. Then we prove the consistency of the instrumental regression estimator, whose proof falls into two parts. Finally, we compute the rate of convergence of the Tikhonov's regularized term to find the global rate of convergence of the estimator.
We recall that the estimator we consider is given by:

$$\hat{\lambda}_{n,\alpha_n} = \Phi[\hat{m}_{n,\alpha_n}](x)$$

First, prove that $H[\hat{m}_{n,\alpha_n}] \to H[m]$. So we have the following decomposition:

$$\begin{aligned}\|H[\hat{m}_{n,\alpha_n}] - H[m]\| &\leqslant \|H[\hat{m}_{n,\alpha_n}] - H[m_{\alpha_n}]\| + \|H[m_{\alpha_n}] - H[m]\| \\ &\leqslant (I) + (II)\end{aligned} \tag{15}$$

We first control the first term of the decomposition.
Using the definition of the estimators (7) and (9), we have the following equalities:

$$\begin{aligned}
&H[\hat{m}_{n,\alpha_n}] - H[m_{\alpha_n}] \\
&= H(\alpha_n I + \hat{T}_F^* \hat{T}_F)^{-1}(\hat{T}_F^* \hat{r}_F - \hat{T}^* \hat{T} m) + H(\alpha_n I + \hat{T}^* \hat{T})^{-1} \hat{T}^* \hat{T} m - H m_{\alpha_n} \\
&= (A) + (B)
\end{aligned}$$

Since the operator H is invertible due to Lemma 4.1, we can write:

$$(A) = H(\alpha_n I + \hat{T}^* \hat{T})^{-1} H^{-1} [H \hat{T}^* \hat{r}^* - H \hat{T}^* \hat{T} m] \tag{16}$$

Since H is invertible, the norm of the first term is unchanged

$$\|H(\alpha_n I + \hat{T}^* \hat{T})^{-1} H^{-1}\| = \|(\alpha_n I + \hat{T}^* \hat{T})^{-1}\|,$$

so we get

$$\begin{aligned}
\|(A)\| &\leqslant \|(\alpha_n I + \hat{T}^* \hat{T})^{-1}\| \|H \hat{T}^* \hat{r} - H \hat{T}^* \hat{T} m\| \\
&\leqslant O\left(\frac{1}{\alpha_n}\right) \|H \hat{T}^* \hat{r} - H \hat{T}^* \hat{T} m\|
\end{aligned}$$

By the result of Lemma 4.3 we can conclude that we get the following bound:

$$\|(A)\|^2 = O\left(\frac{1}{{\alpha_n}^2}\left[\frac{1}{n} + h_n^{2\rho}\right]\right) \tag{17}$$

The second term is such that:

$$\begin{aligned}
&(B) \\
&= H(\alpha_n I + \hat{T}^* \hat{T})^{-1} \hat{T}^* \hat{T} m - H m_{\alpha_n} \\
&= H(\alpha_n I + \hat{T}^* \hat{T})^{-1} \hat{T}^* \hat{T} m - H m_{\alpha_n} + H(\alpha_n I + \hat{T}^* \hat{T})^{-1}(T^* T m - T^* T m) \\
&= H(\alpha_n I + \hat{T}^* \hat{T})^{-1}(\hat{T}^* \hat{T} - T^* T)(m_{\alpha_n} - m) \\
&= H(\alpha_n I + \hat{T}^* \hat{T})^{-1} H^{-1}(H \hat{T}^* \hat{T} - H T^* T)(m_{\alpha_n} - m)
\end{aligned}$$

We can conclude that the second term is such that

$$\|(B)\|^2 \leqslant \|H(\alpha_n I + \hat{T}^* \hat{T})^{-1} H^{-1}\|^2 \|H \hat{T}^* \hat{T} - H T^* T\|^2 \|m_{\alpha_n} - m\|^2 \tag{18}$$

Using Lemma 4.2, we can deduce that

$$\|H \hat{T}^* \hat{T} - H T^* T\|^2 = O\left(\frac{1}{n h_n^p} + h_n^{2\rho}\right)$$

As a result we get the following upper bound

$$\|(B)\|^2 = O\left(\frac{1}{{\alpha_n}^2}[\frac{1}{n h_p^2} + h_n^{2\rho}] \|m_{\alpha_n} - m\|^2\right).$$

The assumption over the model gives the rate of convergence of the approximation error $\|m_{\alpha_n} - m\|^2$. Finally we get:

$$\|(B)\|^2 = O\left(\frac{1}{{\alpha_n}^2}[\frac{1}{n h_p^2} + h_n^{2\rho}] \alpha_n^\beta\right). \tag{19}$$

The second term of the sum $\|H m_{\alpha_n} - H m\|$ is the error made when approximating the function by the solution of the regularized problem. It is the bias of the regularization issue. By Lemma 4.4, we have an upper bound for this term. To conclude the proof, it remains to be seen that the remainder term in the expansion (15) is of the right order. This statement is proved by the Lemma 4.5. As a consequence we get:

$$\mathbf{E}\|\hat{\lambda}_{n,\alpha_n} - \lambda\|^2 = O\left({\alpha_n}^\beta + \frac{1}{{\alpha_n}^2}[\frac{1}{n h_n^p} + h_n^{2\rho}] \alpha_n^\beta + \frac{1}{{\alpha_n}^2}\left[\frac{1}{n} + h_n^{2\rho}\right]\right) \tag{20}$$

There is a trade-off in this expression which can be minimized by adjusting the values of the smoothing parameter α_n and the bandwidth h_n. Note that an optimal choice is given by $\alpha_n \approx n^{-\frac{1}{2+\beta}}$ and $h_n \approx n^{-\frac{2\rho}{2\rho+p}}$, so we get for a given constant C:

$$\mathbf{E}\|\hat{\lambda}_{n_{\alpha_n}} - \lambda\|^2 \leqslant Cn^{-\frac{\beta}{2+\beta}} + n^{-\frac{2\beta}{2+\beta}+\frac{1}{3}} \tag{21}$$

The previous upper bound proves the statement of the theorem. □

Remark 3.5. The rate of convergence depends on the range of the value β. Indeed we have

$$n^{-\frac{\beta}{2+\beta}} \leqslant n^{-\frac{2\beta}{2+\beta}+\frac{1}{3}}$$

as soon as

$$\beta \leqslant 1.$$

As a result if on the one hand β, defined in Assumptions A', is such that $0 \leqslant \beta \leqslant 1$ then the rate of convergence is given by

$$\mathbf{E}\|\hat{\lambda}_{n,\alpha_n} - \lambda\|^2 = O\left(n^{-\frac{2\beta}{2+\beta}+\frac{1}{3}}\right)$$

On the other hand, for $2 \geqslant \beta \geqslant 1$, we get

$$\mathbf{E}\|\hat{\lambda}_{n,\alpha_n} - \lambda\|^2 = O\left(n^{-\frac{\beta}{2+\beta}}\right).$$

The rate of convergence is quicker than the rate we could have obtained by a direct rough upper bound and the results proved by Darolles, Florens and Renault in [DFR01]. Indeed, their method gives a rate of convergence in $n^{-\frac{2\beta}{2+\beta}+\frac{1}{2}}$ for all β. The gain we obtain is due to a better upper bound for the bias term even if no gain is obtained for the variance term not even for the approximation term. That is the reason why the improvement in the endogenous case is less than the one we could have expected in the classical setting, as it is stated by Vanhems in [Van01]. We point out that the case $\beta = 2$ is the border of the set Φ_β.

Remark 3.6. The choice of the bandwidth is not optimal in the sense of nonparametric estimation. As a matter of fact, a minimization in h_n leads to a natural choice of $h_n \approx n^{-\frac{1}{8}}$. It also implies a choice of $\alpha_n \approx n^{-\frac{6}{8(2+\beta)}}$. But the corresponding rate of convergence, $n^{-\frac{3}{4}\frac{\beta}{2+\beta}}$ is slower than the speed given in the theorem. An explanation of this phenomena is that the choice of such h_n, which decreases faster than the optimal choice $h_n \approx n^{-\frac{1}{6}}$, provides an oversmoothing effect.

Remark 3.7. The assumption A' describes a category of set, named Φ_β in Darolles, Florens and Renault in [DFR01]. Such sets are complex since they involve both the smoothness of the function and the dependence between the endogenous and the instrumental variable. Indeed, for $m \in \Phi_\beta$, they rely on the behavior of the coefficients $< m, m_i >, \forall i \geqslant 0$ and on the decreasing rate of the eigenvalues λ_i. The greater β, the more regular is the function. If β is greater than one, the rate of convergence is in $n^{\frac{\beta}{2+\beta}}$. It is important to notice that it does not depend on the dimension of the random variables, nor the regularity $r-1$ of the function m. This result suggests that the regularity condition imposed on the space Φ_β implicitly implies enough regularity, in the classical nonparametric sense. So the rate of convergence of the estimation issue is given by the leading term of the approximation error in Tikhonov's regularization. More attention should be paid to the definition of such spaces.

4 Auxiliary lemmas

Let d be the regularity of the kernel we use and r be the regularity of the regularized function Hm. The following lemma expresses the one to one property of the operator H.

Lemma 4.1. *The operator H defined by*

$$\forall g \in \mathcal{C}(\mathbb{D}), \forall x \in I, H[g](x) = \int_0^x v(x,t) g(t, \lambda(t))\, dt$$

is an invertible operator.

The convergence of the estimators is linked with the two next lemmas, which describe the convergence of $H\hat{T}^*\hat{T}$ and of $\hat{T}^*\hat{r}$. The rate of convergence depends on the choice of an optimal bandwidth h_n.

FIXME: rajouter assumptions B1 et B3 sur support compact et bounds

Lemma 4.2. *Under the assumptions A' and for $\rho = r \wedge d$, we get*

$$\|H\hat{T}^*\hat{T} - HT^*T\|^2 = O\left(\frac{1}{nh_n^p} + h_n^{2\rho}\right) \tag{22}$$

Lemma 4.3. *The following upper bound holds under assumptions A':*

$$\|H\hat{T}^*\hat{r} - H\hat{T}^*\hat{T}m\|^2 = O\left(\frac{1}{n} + h_n^{2\rho}\right) \tag{23}$$

The following lemma describes the approximation error of Tikhonov's regularization problem. This non random term depends on the choice of a smoothing sequence α_n. This choice depends on the smoothness of the function m as well as the behavior of the eigenvalues of the operator T.

Lemma 4.4. *If the function m satisfies the conditions A', and if H is such that*

$$\sum_{i=0}^{\infty} \|Hm_i\|^2 < \infty,$$

we get the following bound:

$$\|Hm_{\alpha_n} - Hm\|^2 = O\left(\alpha_n^\beta\right) \tag{24}$$

The following gives the rate of convergence of the remainder term in the Taylor's expansion and shows that it is negligible.

Lemma 4.5. *As soon as $\frac{nh_n^d}{\log^2 n} \to \infty$ and*

$$\|m_{\alpha_n} - m\|_\infty = O\left(\alpha_n^\beta\right),$$

we get the following uniform convergence:

$$\|\hat{m}_{n,\alpha_n} - m\|_\infty \to 0.$$

Moreover, we get

$$\|\hat{m}_{n,\alpha_n} - m\|_\infty^2 = O\left(\|H\hat{m}_{n,\alpha_n} - Hm\|_2^2\right).$$

The next lemma is an auxiliary lemma, giving the rate of uniform convergence of the derivative of a conditional expectancy. It is used to prove Lemma 4.7.

Lemma 4.6. *Set X, Y, Z random variables with joint density function l, two times continuously differentiable. Define $L(x,y) = \mathbf{E}(Z|X=x, Y=y)$ and $\hat{L}_n$ the associated kernel estimator. We get the following asymptotic behaviour:*

$$\|D_2[\hat{L}_n] - D_2[L]\| \to 0$$

as soon as $\frac{nh_n^{\frac{d}{2}+2}}{\log^2 n} \to \infty$.

The following lemma proves that assumption (A) can be deduced from more drastic conditions above the choice of the bandwidth h_n.

Lemma 4.7. *Under the assumptions that*

$$\|D_2[m_{\alpha_n}] - D_2[m]\|_\infty \to 0$$
$$\frac{nh_n^{\frac{d}{2}+2}}{\log^2 n} \to \infty, \quad h_n^d \leqslant \alpha_n, \quad \frac{\log n}{\sqrt{n}h_n^{\frac{d}{2}+2}} = o(\alpha_n)$$

we get the following asymptotics

$$\|D_2[\hat{m}_{n,\alpha_n}] - D_2[m]\|_\infty \to 0.$$

5 Proofs

5.1 Proof of Auxliary Lemmas

Proof of Lemma 4.1:

Proof. We want to prove that the operator H is invertible. Assume that there exists a function g different from the function 0 such that $H[g]=0$. Without loss of generality, we can assume that g is non negative in a neighborhood of the origin of the form $[0,x_1]$, otherwise consider the function $-g$. Moreover, since g is non equal to zero everywhere, there exists $x_0 < x_1$ such that g is positive over the compact $[x_0,x_1]$. But $H[g](x_1)=0$ implies that $\int_{x_0}^{x_1} v(x_1,t)g(t)\,dt = 0$. Since $v(x,t)>0$, by continuity of the function $t \to v(x,t)g(t)$, we get that $g(t)=0, \forall t \in [x_0,x_1]$ which contradicts the definition of x_0. So H is one to one. ☐

The two following proofs are adaptations of the proofs in [DFR01]. We give here the difference and recall te guidelines of the proofs.

Proof of Lemma 4.2:

Proof. The norm of the operator $H\hat{T}^*\hat{T} - HT^*T$ defined by

$$\|H\hat{T}^*\hat{T} - HT^*T\| = \sup_{\|g\|\leqslant 1} \|(H\hat{T}^*\hat{T} - HT^*T)g\|$$

is bounded by the Hilbert-Schmidt norm of the operator, ass stated in [DS88]. Hence, the norm satisfies:

$$\begin{aligned} A_n^2 &= \|H\hat{T}^*\hat{T} - HT^*T\|_{HS}^2 \\ &= \int\int f(.,u,.)\left[\int v(x,z)1_{0\leqslant z\leqslant x}\int (\hat{e}_n(z,u,w) - e(z,u,w))\,dw\,dw\right]^2 dudx \end{aligned}$$

where $e(u,z,w) = \dfrac{f(.,u,w)f(.,z,w)}{f(.,z,.)f(.,.,w)}$ and $\hat{e}_n(u,z,w)$ is a kernel estimator of $e(u,z,w)$. As in [DFR01], we obtain by linearizing:

$$A_n^2 = R_n + \sum_{j=1}^{4} B_j$$

where, if we set

$$\begin{aligned} \tilde{v}(x,z) &= x(x,z)1_{0\leqslant z\leqslant x} \\ b_1(z,w) &= \frac{f(.,z,w)}{f(.,z,.)f(.,.,w)} \quad b_2(u,z,w) = \frac{f(.,u,w)}{f(.,z,.)f(.,.,w)} \\ b_3(u,z,w) &= \frac{f(.,u,w)f(.,z,w)}{f^2(.,z,.)f(.,.,w)} \quad b_4(u,z,w) = \frac{f(.,u,w)f(.,z,w)}{f(.,z,.)f^2(.,.,w)} \end{aligned}$$

we have written:

$$\begin{aligned} B_1 &= \int\int f(.,u,.)[\int\int \tilde{v}(x,z)b_1(z,w)(\hat{f}(.,u,w) - f(.,u,w))dwdz]^2 dxdu \\ B_2 &= \int\int f(.,u,.)[\int\int \tilde{v}(x,z)b_2(u,z,w)(\hat{f}(.,z,w) - f(.,z,w))dwdz]^2 dxdu \\ B_3 &= \int\int f(.,u,.)[\int\int \tilde{v}(x,z)b_3(u,z,w)(\hat{f}(.,z,.) - f(.,z,.))dwdz]^2 dxdu \\ B_4 &= \int\int f(.,u,.)[\int\int \tilde{v}(x,z)b_4(u,z,w)(\hat{f}(.,.,w) - f(.,.,w))dwdz]^2 dxdu \end{aligned}$$

Since the B_i, $i=1,\ldots,4$ are positive, by Chebychev's inequality, its rate of convergence is the rate of $\mathbf{E}(B_i)$, $i=1,\ldots,4$. Recall that the nonparametric estimator of a density we use is defined using a kernel estimator. Recall that we consider kernels of order r and that the estimator of the joint density has the following expression, given in (8):

$$\hat{f}_n(y,z,w) = \frac{1}{n}\sum_{i=1}^{n} K_{y,h_n^y}(y-Y_i)K_{z,h_n^z}(z-Z_i)K_{w,h_n^w}(w-W_i).$$

As a consequence, the behavior of each $\mathbf{E}(B_i)$, $i=1,\ldots,4$ can deduced from the rate of convergence of $\mathbf{E}(\hat{f}_n - f)^2$, extended to each particular case.
The first term gives

$$\mathbf{E}\left(\int\int \tilde{v}(x,z)b_1(z,w)(\hat{f}(.,u,w) - f(.,u,w))dwdz\right)^2 = O\left(\frac{1}{nh_n^p} + h_n^{2\rho}\right)$$

Indeed, the integration implies that only the dimension of Z appears in the exponent of h_n. So we get

$$B_1 = O\left(\frac{1}{nh_n^p} + h_n^{2\rho}\right)$$

The second term is such that:

$$\mathbf{E}\left(\int\int f(.,u,.)\left[\tilde{v}(x,z)b_2(u,z,w)(\hat{f}(.,z,w) - f(.,z,w))dwdz\right]^2 dudx\right) = O\left(\frac{1}{n} + h_n^{2\rho}\right)$$

which implies that

$$B_2 = O\left(\frac{1}{n} + h_n^{\rho}\right).$$

The same arguments apply for the other terms. Then, we can use elementary extensions of usual arguments on the asymptotic behavior of the quadratic loss in kernel estimation (see Bosq in [Bos96]). Therefore we can show that, provided $nh_n^{p+2(r\wedge s)} \to 0$, the remainder term R_n is such that $R_n = O_P\left(\frac{1}{nh_n^2}\right)$. So the statement of Lemma 4.2 is proved. □

Proof of Lemma 4.3:

Proof. It relies on the convergence of Gaussian process and use results from empirical process theory to be found in [vdVW96].
In a first step, we prove that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f(.,z,w_i)}{f(.,z,.)f(.,.,w_i)}(y_i - m(z_i)) \to \mathcal{N}\left(0, \sigma^2 T^* T\right), \tag{25}$$

where $\sigma^2 = \mathrm{Var}(U|W)$. For this, note $V_n(z) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f(.,z,w_i)}{f(.,z,.)f(.,.,w_i)}(y_i - m(z_i)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta_i(z)$. The β_i are independent variables, with zero mean, finite variance and such that $\mathbf{E}(\|\beta_i(Z)\|_2^2) < \infty$, as a result, V_n converges to a Gaussian process, with variance given by an operator K such that for any functions $m, \psi \in L_Z^2$:

$$\begin{aligned}
< K\psi, m > &= \mathbf{E}(< V_n, m >< V_n, \psi >) \\
&= \int\int \mathbf{E}\left[V_n(z)V_n(u)\right] m(z)\psi(u)f(.,z,.)f(.,u,.)dzdu \\
&= \sigma^2 \int\int \mathbf{E}\left[\frac{f(.,z,w_i)f(.,u,w_i)}{f(.,z,.)f(.,.,w_i)}\right] m(z)\psi(u)f(.,z,.)f(.,u,.)dzdu
\end{aligned}$$

So we get

$$K\psi = \sigma^2 \int\int \frac{f(.,z,w)f(.,u,w)}{f(.,z,.)f(.,.,w)}\psi(u)dzdz.$$

Finally we obtain the asymptotic distribution which proves (25).
In a second step, we linearize the quantity $A_n = \hat{T}^*\hat{r} - \hat{T}^*\hat{T}m$ in the following way:

$$A_n = R_0 + R_1 + \frac{1}{n}\sum_{i=1}^{n}\frac{f(.,z,w_i)}{f(.,z,.)f(.,.,w_i)}(y_i - m(z_i))$$

For the first term, we use the first asymptotic result, while the remaining term can be written

$$R_1 \asymp \frac{1}{n}\sum_{i=1}^{n}\left[\int a(s)K_{h_n}(s - s_i)ds - a(s_i)\right]$$

where the function $s = (y, u, w)$ and a is such that

$$a(s) = \frac{f(.,z,w)}{f(.,z,.)f(.,.,w)}(y - m(u)).$$

The usual approximating properties of the kernel estimator proves that this term is in $O_P\left(h_n^{2(r\wedge s)}\right)$ which gives the upper bound given by the lemma.
Now see, that

$$R_0 = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{f(.,z,w_i)}{f(.,z,.)f(.,.,w_i)}(y_i - m(z_i)) - \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\int a(s)K_{h_n}(s-s_i)ds - a(s_i)\right]$$

With the regularity assumption we assume over the function m, the last term goes to zero, which concludes the proof. ☐

Proof of Lemma 4.4:

Proof. m_{α_n} is the solution of the equation 7. It can be computed, using its decomposition onto the bases m_i, $i \geqslant 0$ and can be written:

$$m_{\alpha_n}(z) = \sum_{i=0}^\infty \frac{\lambda_i}{\alpha_n + \lambda_i^2} < r, \psi_i > \phi_i(z).$$

As a result we have

$$\|Hm_{\alpha_n} - Hm\|^2 = \|\sum_{i=0}^\infty \frac{\alpha_n}{\alpha_n + \lambda_i^2} < m, \phi_i > H\phi_i\|^2 \tag{26}$$

Using the decomposition onto the bases ϕ_i we get:

$$H\phi_i = \sum_{j=0}^\infty < H\phi_i, \phi_j > \phi_j.$$

Then we obtain using the linearity of the scalar product:

$$\begin{aligned}\|Hm_{\alpha_n} - Hm\|^2 &= \|\sum_{j=0}^\infty(\sum_{i=0}^\infty \frac{\alpha_n}{\alpha_n + \lambda_i^2} < m, \phi_i >< H\phi_i, \phi_j >)\phi_j\|^2 \\ &= {\alpha_n}^2 \sum_{j=0}^\infty \left(\sum_{i=0}^\infty \frac{1}{\alpha_n + \lambda_i^2} < H\phi_i, \phi_j >< m, \phi_i >\right)^2\end{aligned}$$

As a result, such spaces are of interest in the range $0 \leqslant \beta \leqslant 2$. Indeed, regularity assumptions over the regression function m give the convergence of series $\sum_{i=0}^\infty c_i < m, \phi_i >^2$, for well chosen c_i. For instance if $c_i \approx i^{2s+1}$, spaces $\{\phi, \sum_{i=0}^\infty i^{2s+1} < \phi, \phi_i >^2 < \infty\}$ are balls of Sobolev spaces H^s.
If we write

$$< m, \phi_i >= \lambda_i^2 c_i \tag{27}$$

with $\sum_{j=0}^\infty c_j < \infty$ and $\sum_{i=0}^\infty \|H\phi_i\|^2 < \infty$, it is obvious that we get the case $\alpha = 1$. Indeed, by Cauchy-Schwarz's inequality we can write:

$$\begin{aligned}\|Hm_{\alpha_n} - Hm\|^2 &= \sum_{j=0}^\infty \left(\sum_{i=0}^\infty \frac{\alpha_n}{\alpha_n + \lambda_i^2} < H\phi_i, \phi_j >< m, \phi_i >\right)^2 \\ &\leqslant \sum_{j=0}^\infty \left[\left(\sum_i \frac{\alpha_n^2}{(\alpha_n + \lambda_i^2)^2} < m, \phi_i >^2\right)\sum_{i=0}^\infty < H\phi_i, \phi_j >^2\right] \\ &\leqslant \sum_i \frac{\alpha_n^2}{(\alpha_n + \lambda_i^2)^2} < m, \phi_i >^2 \sum_{i,j} < H\phi_i, \phi_j >^2 \\ &\leqslant O(\alpha_n)\sum_{i=0}^\infty \|H\phi_i\|^2 \\ &= O(\alpha_n)\end{aligned}$$

Using the same ideas, the condition

$$< m, \phi_i >= \lambda_i^4 c_i, \quad \sum_{i=0}^{\infty} c_i < \infty$$

leads to the conclusion that $\alpha = 2$.
The previous condition (27) links the decay of the eigenvalues λ_i with the decay of the Fourier coefficients of the function ϕ. The same assumption is done in [DFR01]. This assumption is about both the smoothness properties of the operator T and of the function m. It is not possible to break this relationship as it is stated in [LV04]. Even if this kind of assumption is not usual in statistics, we point out that, in the litterature of inverse problems in numerical analysis, the regularity of the function and the regularity of the operator are linked and both give the rate of convergence. For more references, we refer to [1], [2] or [3].
We could have expected that the regularization via the integral operator to improve the upper bound. But, with the same arguments as in Remark 3.5 we can see that the regularization does not improve the rate of convergence of the approximation term, due to the two-step estimation. □

<u>Proof of Lemma 4.5</u>:

Proof. We want to prove the uniform convergence of $\widehat{m}_{n,\alpha_n}$ to m
We have the following decomposition:

$$\|\widehat{m}_{n,\alpha_n} - m\|_\infty \leqslant \|\widehat{m}_{n,\alpha_n} - m_{\alpha_n}\|_\infty + \|m_{\alpha_n} - m\|_\infty$$

As in the general theorem, the idea is to control each term of the decomposition.
We know that:

$$\|\widehat{m}_{n,\alpha_n} - m_{\alpha_n}\|_\infty \leqslant \|A\|_\infty + \|B\|_\infty$$

where, we have used the same notations as in 3.4:

$$\begin{aligned} \|A\|_\infty &\leqslant \left\| \left(\alpha_n I + \widehat{T}^*\widehat{T}\right)^{-1} \right\|_\infty . \left\| \widehat{T}^*\widehat{r} - \widehat{T}^*\widehat{T}m \right\|_\infty \\ &\leqslant O\left(\frac{1}{\alpha_n}\right) . \left\| \widehat{T}^*\widehat{r} - \widehat{T}^*\widehat{T}m \right\|_\infty \end{aligned}$$

and

$$\begin{aligned} \|B\|_\infty &\leqslant \left\| \left(\alpha_n I + \widehat{T}^*\widehat{T}\right)^{-1} \right\|_\infty . \left\| \widehat{T}^*\widehat{T} - T^*T \right\|_\infty . \|m_{\alpha_n} - m\|_\infty \\ &\leqslant O\left(\frac{1}{\alpha_n}\right) . \left\| \widehat{T}^*\widehat{T} - T^*T \right\|_\infty . \|m_{\alpha_n} - m\|_\infty \end{aligned}$$

The next step is to study the two terms $\left\| \widehat{T}^*\widehat{r} - \widehat{T}^*\widehat{T}m \right\|_\infty$ and $\left\| \widehat{T}^*\widehat{T} - T^*T \right\|_\infty$. Let's begin with the second one. By definition, we have:

$$\begin{aligned} \left\| \widehat{T}^*\widehat{T} - T^*T \right\|_\infty &= \sup_{\|g\| \leqslant 1} \left\| \left(\widehat{T}^*\widehat{T} - T^*T\right) g \right\|_\infty \\ &\leqslant \left\| \widehat{T}^*\widehat{T} - T^*T \right\|_{HS} \end{aligned}$$

where $\|.\|_{HS}$ is the Hilbert-Schmidt norm of this operator (see Dunford and Schwartz [DS88]). Therefore, using Darolles, Florens and Renault, we find that:

$$\|B\|_\infty^2 = O\left(\frac{1}{\alpha_n^2}\left[\frac{1}{nh_n^2} + h_n^{2(r\wedge d)}\right] \|m_{\alpha_n} - m\|_\infty^2\right)$$

The first term is such that:

$$\begin{aligned}
&\left\|\widehat{T}^*\widehat{r}-\widehat{T}^*\widehat{T}m\right\|_\infty \\
&=\left\|\int\int\int (y-m(u))\frac{\widehat{f}_n(.,z,w)}{\widehat{f}_n(.,.,w)\widehat{f}_n(.,z,.)}\widehat{f}_n(y,u,w)\,dudydw\right\|_\infty \\
&\leqslant\left\|\int\int\int (y-m(u))\left(\frac{\widehat{f}_n(.,z,w)}{\widehat{f}_n(.,.,w)\widehat{f}_n(.,z,.)}-\frac{f(.,z,w)}{f(.,.,w)f(.,z,.)}\right)\widehat{f}_n(y,u,w)\,dudydw\right\|_\infty \\
&+\left\|\int\int\int (y-m(u))\frac{f(.,z,w)}{f(.,.,w)f(.,z,.)}\widehat{f}_n(y,u,w)\,dudydw\right\|_\infty \\
&\leqslant\left\|\int\int\int (y-m(u))\frac{f(.,z,w)}{f(.,.,w)f(.,z,.)}\widehat{f}_n(y,u,w)\,dudydw\right\|_\infty
\end{aligned}$$

Then,we can decompose this term as follows:

$$\left\|\widehat{T}^*\widehat{r}-\widehat{T}^*\widehat{T}m\right\|_\infty\leqslant\|R\|_\infty+\left\|\frac{1}{n}\sum_{i=1}^n\frac{f(.,z,w_i)}{f(.,z,.)f(.,.,w_i)}(y_i-m(z_i))\right\|_\infty$$

where

$$R=\frac{1}{n}\sum_{i=1}^n\left\{\int a(s)K_{h_n}(s-s_i)\,ds-a(s_i)\right\}$$

and

$$a(s)=\frac{f(.,z,w)}{f(.,z,.)f(.,.,w)}(y-m(u)),\,s=(y,u,w)$$

We know, using the asymptotic distribution given in the proof of Lemma 4.3 that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{f(.,z;,w_i)}{f(.,z,.)f(.,.,w_i)}(y_i-m(z_i))\xrightarrow{L}\mathcal{N}(O,\sigma^2T^*T).$$

Since we have for $\epsilon>0$

$$\begin{aligned}
&\mathbf{P}\left(|\frac{1}{n}\sum_{i=1}^n\frac{f(.,z,w_i)}{f(.,z,.)f(.,.,w_i)}(y_i-m(z_i))|>\epsilon\right)\\
&\leqslant\mathbf{P}\left(|\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{f(.,z,w_i)}{f(.,z,.)f(.,.,w_i)}(y_i-m(z_i))|>\sqrt{n}\epsilon\right),
\end{aligned}$$

we get, using Billingsley's inequality [Bi95] that

$$\mathbf{P}\left(\sum_n\left|\frac{1}{n}\sum_{i=1}^n\frac{f(.,z,w_i)}{f(.,z,.)f(.,.,w_i)}(y_i-m(z_i))\right|>\epsilon\right)<\infty$$

and Borel Cantelli's lemma leads to the conclusion that

$$\|\frac{1}{n}\sum_{i=1}^n\frac{f(.,z,w_i)}{f(.,z,.)f(.,.,w_i)}(y_i-m(z_i))\|_\infty^2\to 0$$

at a range of convergence of $\frac{\log n}{n}$ as soon as the following condition is fulfilled:

$$\frac{n}{(\log n)^2}->+\infty\;.$$

The residual term can be viewed as a bias term, so we can prove the uniform convergence of that term provided that $\frac{nh_n^d}{(\log n)^2}->+\infty$

As a result, the rate of convergence is, at a logarithmic term, the same as for the L^2 rate of convergence. As a consequence, as it is stated in the proof of Theorem 1 in [Van01], the remainder term in Taylor's expansion is negligible since we only lose a logarithmic term in the rate of convergence. □

Proof of Lemma 4.6:

Proof. Define the following function:

$$\begin{aligned}\mathbf{E}(Z|X=x,Y=y) &= L(x,y)\\ &= \int z\frac{l(x,y,z)}{l(x,y)}\,dz\\ &= \frac{q(y,z)}{l(x,y)}\end{aligned}$$

where $l(x,y)=\int f(x,y,z)dz$.
An estimator of this function is given by

$$\begin{aligned}\hat{L}_n(x,y) &= \int z\frac{\hat{l}_n(x,y,z)}{\hat{l}_n(x,y)}dz\\ &= \int z\hat{e}_n(x,y,z)dz,\end{aligned}$$

with

$$\hat{e}_n(x,y,z)=\frac{\sum_{i=1}^n K_h(x-X_i)K_h(y-Y_i)K_h(z-Z_i)}{\sum_{i=1}^n K_h(x-X_i)K_h(y-Y_i)}.$$

Our goal is to provide conditions over the bandwidth h_n such that the derivative of the estimator with respect to the second variable converges uniformly to the similar derivative of the true function.

$$\|D_2[\hat{L}_n]-D_2[L]\|_\infty \to 0.$$

For this, use the following decomposition:

$$\begin{aligned}D_2[\hat{L}_n]-D_2[L] &= \frac{lD_2[\hat{q}_n]-l\hat{L}_nD_2[\hat{l}_n]-\hat{l}_nD_2[q]+L\hat{l}_nD_2[l]}{l\hat{l}_n}\\ &= (l-\hat{l}_n)\frac{D_2[q]-LD_2[l]}{l\hat{l}_n}+(L-\hat{L}_n)\frac{D_2[\hat{l}_n]}{\hat{l}_n}\\ &+\frac{L}{\hat{l}_n}(D_2[l]-D_2[\hat{l}_n])+\frac{1}{\hat{l}_n}(D_2[\hat{q}_n]-D_2[q])\end{aligned}$$

As a result, the uniform convergence of $D_2[\hat{L}_n]$ to $D_2[L]$ can be deduced from the uniform convergence of the previous four terms.
We have already proven that, under suitable conditions, $\|\hat{L}_n-L\|_\infty\to 0$, when bounding the remainder term in the Taylor expansion.
Moreover, we already know that $\hat{l}_n$ converges uniformly to l.
So it remains to be seen the two following uniform convergence:

$$\|D_2[\hat{l}_n]-D_2[l]\|_\infty \to 0 \tag{28}$$
$$\|D_2[\hat{q}_n]-D_2[q]\|_\infty \to 0 \tag{29}$$

If we have proved the first uniform convergence, since

$$q(x,y)=\int zl(x,y,z)\,dz$$

Lebesgue theorem as well as the fact that the functions are defined over compact sets, give the second uniform convergence. As a consequence it suffices to prove the first asymptotics (28).
The estimator $\hat{l}_n$ is defined as follows:

$$\hat{l}_n(x,y,z)=\frac{1}{nh_n^d}\sum_{i=1}^n K(\frac{x-X_i}{h_n})K(\frac{y-Y_i}{h_n})K(\frac{z-Z_i}{h_n}).$$

So we get

$$D_2[\hat{l}_n(x,y,z)]=\frac{1}{nh_n^{d+1}}\sum_{i=1}^n K(\frac{x-X_i}{h_n})K^{'}(\frac{y-Y_i}{h_n})K(\frac{z-Z_i}{h_n}).$$

Using partial integration, it is well known that

$$
\begin{aligned}
l\mathbf{E}(D_2[\hat{l}_n]) &= \frac{1}{h_n^{d+1}} \int\int\int l(u,v,w)K(\frac{x-u}{h_n})K^{'}(\frac{y-v}{h_n})K(\frac{z-w}{h_n})dudvdw \\
&= \frac{1}{h_n^d} \int\int\int D_2[l(u,v,w)]K(\frac{x-u}{h_n})K(\frac{y-v}{h_n})K(\frac{z-w}{h_n})dudvdw, \qquad (30)
\end{aligned}
$$

which proves that $D_2[\hat{l}_n] \to D_2[l]$.
Hence we can write the following decomposition:

$$
\begin{aligned}
D_2[\hat{l}_n] - D_2[l] &= D_2[\hat{l}_n] - \mathbf{E}(D_2[\hat{l}_n]) + \mathbf{E}(D_2[\hat{l}_n]) - l \\
&= (I) + (II).
\end{aligned}
$$

The second term is the usual bias term. As expected, using the assumption that $l \in \mathcal{C}^2$, which implies that $D_2[l] \in \mathcal{C}^1$ and using (30), we get:

$$
\begin{aligned}
(II) &= \int\int\int \frac{1}{h_n^3}(D_2[l(u,v,w)] - D_2[l(x,y,z)])K(\frac{x-u}{h_n})K(\frac{y-v}{h_n})K(\frac{z-w}{h_n})dudvdw \\
&= \int\int\int K(u)K(v)K(w)[D_2[l(x-h_nu, y-h_nv, z-h_nw)] - D_2[l(x,y,z)]]dudvdw \\
&= o(h_n^d)
\end{aligned}
$$

and that bound is uniform over x, y, z.
The first term can be written as follows:

$$
(I) = \frac{1}{nh_n^{d+1}} \sum_{i=1}^{n} \left[K(\frac{x-X_i}{h_n})K^{'}(\frac{y-Y_i}{h_n})K(\frac{z-Z_i}{h_n}) - \mathbf{E}(K(\frac{x-X_i}{h_n})K^{'}(\frac{y-Y_i}{h_n})K(\frac{z-Z_i}{h_n})) \right]
$$

Using empirical process theory, we get the following upper bound:

$$
\begin{aligned}
&\mathbf{P}(|(I)| > \epsilon) \\
&= \mathbf{P}(|\frac{1}{nh_n^{d+1}} \sum_{i=1}^{n} \left[K(\frac{x-X_i}{h_n})K^{'}(\frac{y-Y_i}{h_n})K(\frac{z-Z_i}{h_n}) - \mathbf{E}(K(\frac{x-X_i}{h_n})K^{'}(\frac{y-Y_i}{h_n})K(\frac{z-Z_i}{h_n})) \right] | > \epsilon) \\
&\leqslant 4\exp\left(- \frac{\epsilon^2 h_n^2 \sqrt{n} h_n^{-\frac{d}{2}} h_n^d}{8\|K\|_\infty^3} \right) \\
&\leqslant 4\exp\left(- \frac{\epsilon^2 \sqrt{n} h_n^{\frac{d}{2}+2}}{8\|K\|_\infty^3} \right).
\end{aligned}
$$

As a result, as soon as the following condition is fulfilled

$$
\frac{nh_n^{\frac{d}{2}+2}}{\log^2 n} \to \infty \qquad (31)
$$

an extensive use of Borel Cantelli's lemma gives the uniform convergence of term (I).
As a conclusion, the two terms converge uniformly, which proves the result. □

Proof of Lemma 4.7:

Proof. We want to prove the uniform convergence of $D_2[\hat{m}_{n,\alpha_n}]$ to $D_2[m]$.
For this recall that m_{α_n} and $\hat{m}_{n,\alpha_n}$ are defined by the following relations:

$$
\begin{aligned}
\alpha_n m_{\alpha_n} + \int m_{\alpha_n}(u)a(u,z)du &= \int yb(y,z)dy \\
\alpha_n \hat{m}_{n,\alpha_n} + \int \hat{m}_{n,\alpha_n}(u)\hat{a}_n(u,z)du &= \int y\hat{b}_n(y,z)dy
\end{aligned}
$$

As a consequence, considering the derivative with respect to the second variable (written D_2), we get

$$\alpha_n D_2[m_{\alpha_n}] + \int m_{\alpha_n}(u) D_2[a(u,z)]du = \int y D_2[b(y,z)]dy$$
$$\alpha_n D_2[\hat{m}_{n,\alpha_n}] + \int \hat{m}_{n,\alpha_n}(u) D_2[\hat{a}_n(u,z)]du = \int y D_2[\hat{b}_n(y,z)]dy \quad (32)$$

So we can write

$$\begin{aligned} D_2[\hat{m}_{n,\alpha_n}] - D_2[m] &= (D_2[\hat{m}_{n,\alpha_n}] - D_2[m_{\alpha_n}]) + (D_2[m_{\alpha_n}] - D_2[m]) \\ &= \frac{1}{\alpha_n}(\alpha_n D_2[\hat{m}_{n,\alpha_n}] - \alpha_n D_2[m_{\alpha_n}]) + (D_2[m_{\alpha_n}] - D_2[m]) \\ &= \frac{1}{\alpha_n}(I) \quad + \quad (II). \end{aligned}$$

For the first term (I) consider the following decomposition using (32):

$$\begin{aligned} (I) &= \underbrace{\int y \left(D_2[\hat{b}_n(y,z)] - D_2[b(y,z)] \right) dy}_{A} \\ &+ \int m(u) D_2[a(u,z)]du - \int \hat{m}_{n,\alpha_n} D_2[\hat{a}_n(u,z)]du \\ &= A + \int (m(u) - \hat{m}_{n,\alpha_n}(u)) D_2[a(u,z)]du + \int \hat{m}_{n,\alpha_n}(u) \left(D_2[b(u,z)] - D_2[\hat{b}_n(u,z)] \right) du. \end{aligned}$$

Recall that we have used the following notations

$$a(u,z) = \int \frac{f(.,u,w)f(.,z,w)}{f(.,.,w)f(.,z,.)}dw$$
$$b(u,z) = \int \frac{f(y,.,w)f(.,z,w)}{f(.,.,w)f(.,z,.)}dw.$$

As a result, we can give an interpretation of the quantities $\int yb(y,z)dy$ and $\int m(u)a(u,z)du$ using conditional expectation. There exist two functions g and h such that

$$\begin{aligned} \int m(u) D_2[a(u,z)]du &= \int g(w) D_2[\frac{f(.,z,w)}{f(.,z)}]dw \\ &= D_2[\mathbf{E}(g(W)|Z)] \end{aligned}$$

$$\int y D_2[b(y,z)]dy = D_2[\mathbf{E}(h(W)|Z)].$$

Uniform convergence is then a consequence of Lemma 4.6. As a result, since it is stated that

$$\begin{aligned} \|(I)\|_\infty &\leqslant o(h_n^d) + O\left(\frac{\log n}{\sqrt{n} h_n^{\frac{d}{2}+2}} \right) \\ &\leqslant o(\alpha_n) \end{aligned}$$

due to the restrictions imposed on the bandwidth in Lemma 4.7, we have proved that

$$\frac{1}{\alpha_n}\|(I)\|_\infty = o(1).$$

Moreover the same assumptions give the convergence of the bias term (II), which concludes the proof. □

5.2 Outline of proof of linearization of the differential equation

Proof of Theorem 3.2:

Proof. Under the assumptions over the statistical model, we know that there exits a unique solution to (4). We can call it $\lambda(x) = \Phi[m](x)$, where Φ is a differentiable functional of m. The idea is to use a first order Taylor development and to study both terms. Thus, we have:

$$\begin{aligned}\left(\widehat{\lambda}_n - \lambda\right)(x) &= \left(\Phi\left[\widehat{m}_{n,\alpha_n}\right] - \Phi[m]\right)(x) \\ &= d\Phi[m](\widehat{m}_{n,\alpha_n} - m)(x) + R\end{aligned}$$

where R is stronger than the rest of Taylor development, that is: $R = O_P\left(\|\widehat{m}_{n,\alpha_n} - m\|_\infty^2\right)$.
Our objective is to study the Fréchet-derivative of Φ. Let us now define the function

$$A : \left\{\begin{array}{c} C^1(D) \times C^1_{b,0}(I) \to C(I) \\ (u,v) \mapsto A(u,v)\end{array}\right.$$

where $C^1(D) = \{u \in C(D) \text{ and continuously differentiable}\}$ and

$$C^1_{b,0}(I) = \{v \in C_{b,0}(I), \text{ continuously differentiable and } \|v'\|_\infty < b/a\}$$

where $D = \{(x,y)\,;|x| \leqslant a, |y| \leqslant b\}$.
$\left(C^1(D), \|.\|_\infty\right)$ and $(C(I), \|.\|_\infty)$ are Banach spaces. Moreover we define the following norm:

$$\|.\|_\infty^{'} = \max\left(\|v\|_\infty, \|v'\|_\infty\right)$$

on $C^1_{b,0}(I)$. We can easily see that $\left(C^1_{b,0}(I), \|.\|_\infty^{'}\right)$ is a Banach space. As a matter of fact, to prove it, we have to use the uniform convergence of functions and its application to differentiability. The use of such a norm allows us to have the continuity and linearity of the following application:

$$D : \left\{\begin{array}{c} \left(C^1_{b,0}(I), \|.\|_\infty^{'}\right) \to (C(I), \|.\|_\infty) \\ y \longmapsto y'\end{array}\right.$$

So, we have: $\forall x \in I, A(u,v)(x) = v'(x) - u(x, v(x))$. Let us now define an open subset W of $C^1(D) \times C^1_{b,0}(I)$ and $(m,\lambda) \in W$. We know that A is continuous on W (it is a sum of continuous applications) and that $A(m,\lambda) = 0$. Let us check the hypothesis of the implicit function theorem. A is in fact continuously differentiable (thanks to the same argument) so we can take its derivative with the second variable $\partial_2 A(m,\lambda)$. Moreover, we have:

$$\forall h \in C^1_{b,0}(I), \forall x \in I, \partial_2 A(m,\lambda)(h)(x) = h'(x) - \partial_2 m(x, \lambda(x)).h(x)$$

We have to prove that $\partial_2 A(m,\lambda)$ is a bijection. Let us show first the surjectivity:

$$\forall v \in C(I), \exists ? h \in C^1_{b,0}(I); \forall x \in I, h'(x) - \partial_2 m(x, \lambda(x)).h(x) = v(x)$$

This is a linear differential equation, so we can solve it and find that:

$$\forall x \in I, h(x) = \int_0^x \left(v(s).e^{\left[\int_0^x \partial_2 m(t,\lambda(t))dt - \int_0^s \partial_2 m(t,\lambda(t))dt\right]}\right) ds$$

Therefore, $D_2 A(m,\lambda)$ is surjective. Let us now demonstrate the injectivity, that is

$$Ker\left(D_2 A(m,\lambda)\right) = \{0\}$$

We are going to solve $D_2 A(m,\lambda)h = 0, h \in C^1_{b,0}(I)$. We find again a linear differential equation we can solve and find:

$$\forall x \in I, h(x) = ce^{\int_0^x D_2 m(t,\lambda(t))dt} \text{ and } h(0) = 0$$

Therefore, we get $c = 0$. Thus, we have demonstrated that $D_2 A(m,\lambda)$ is bijective. Let us now demonstrate the bi-continuity of $\partial_2 A(m,\lambda)$. In the usual implicit function theorem, this assumption is not required, but here we consider infinite dimension spaces that is why we need a more general theorem with further assumptions to satisfy. The continuity of $D_2 A(m,\lambda)$ has already been proved since A is continuously differentiable.

The continuity of the reversible function is given by an application of Baire Theorem: if an application is linear continuous and bijective on two Banach spaces, the reversible application is continuous.

Therefore, we can apply the implicit function theorem: $\exists U$ an open subset around m and V an open subset around λ such as:

$$\forall u \in U, A(u, y) = 0 \text{ has a unique solution in } V$$

Let us note: $y = \Phi[u]$ this unique solution for $u \in U$.

Now we are going to differentiate the relation: $A(u, \Phi[u]) = 0, \forall u \in U$ and apply it in $(m, \lambda = \Phi[m])$. Let us first differentiate A: $\forall h \in C^1(D) \times C^1_{b,0}(I)$,

$$\begin{aligned} dA(m,\lambda)(h)(x) &= d_1A(m,\lambda)dm(h)(x) + d_2A(m,\lambda)d\lambda(h)(x) \\ &= -dm(h)(x,\lambda(x)) + (d\lambda(h))'(x) - \partial_2 m(x,\lambda(x))d\lambda(h)(x) \end{aligned}$$

The differential of A leads to a linear differential equation in $d\lambda(h)$ that we can solve. Now we apply it with $dm(h) = \widehat{m}_{n,\alpha_n} - m$ and $d\lambda(h) = d\Phi[m](\widehat{m}_{n,\alpha_n} - m)$ in order to find:

$$d\Phi[m](\widehat{m}_{n,\alpha_n} - m)'(x) = \partial_2 m\left(x, \Phi[m](x)\right).d(\widehat{m}_{n,\alpha_n} - m)(x) + \left(\widehat{m}_{n,\alpha_n} - m\right)\left(x, \Phi[m](x)\right)$$

Solving it leads us to:

$$\begin{aligned} d\Phi[m](\widehat{m}_{n,\alpha_n} - m)(x) &= \int_0^x \left(\left(\widehat{m}_{n,\alpha_n} - m\right)(t, \Phi[m](t)).e^{\left[\int_0^x \frac{\partial m}{\partial e_2}(u,\Phi[m](u))du - \int_0^t \frac{\partial m}{\partial e_2}(u,\Phi[m](u))du\right]} \right) dt \\ &= \int_0^x \left(\left(\widehat{m}_{n,\alpha_n} - m\right)(t, \lambda(t)).e^{\left[\int_0^x \frac{\partial m}{\partial e_2}(u,\lambda(u))du - \int_0^t \frac{\partial m}{\partial e_2}(u,\lambda(u))du\right]} \right) dt \\ &= \int_0^x \left(\left(\widehat{m}_{n,\alpha_n} - m\right)(t, \lambda(t)).v(x,t) \right) dt \end{aligned}$$

So the statement is proved. □

References

[Am74] T. Amemiya. Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica*, 42:999–1012, 1974.

[Bi95] P. Billingsley. Probability and Measure. *John Wiley and Sons Inc.*, 1995.

[blund00] R. Blundell and J. Powell. Endogeneity in nonparametric and semiparametric regression models. *Invited lecture at the 8th World Congress of the Econometric Society*, 2000.

[Bos96] D. Bosq. *Nonparametric statistics for stochastic processes.* Springer-Verlag, New York, 1998. Estimation and prediction.

[CT00] L. Cavalier and A. Tsybakov. Sharp adaptation for inverse problems with random noise. *preprint*, 2000.

[CIK99] P. Chow, Ibragimov I. and Khasminskii R. Statistical approach to some ill-posed problems for partial differential equations. *Prob. Theory and Related Fields*, 113, 421–441, 1999.

[Erm89] M. Ermakov. Minimax estimation of the solution of an ill-posed convolution type problem. *Problems of Information Transmission*, 25, 191–200, 1989.

[DFR01] S. Darolles, J-P. Florens, and E. Renault. Nonparametric instrumental regression. *preprint*, 2001.

[1] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications.* Kluwer Academic Publishers Group, Dordrecht, 1996.

[Flo00] J-P. Florens. Inverse problems and structural econometrics: the example of instrumental variables. *Invited lecture at the 8th World Congress of the Econometric Society*, 2000.

[Han82] L.P. Hansen. Large sample properties of generelaized method of moment estimators. *Econometrica*, 50:1029–1054, 1982.

[HN] J. Hausman and W. K. Newey. Nonparametric estimation of exact consumers surplus and deadweight loss. *Econometrica*, 63:1445–1476.

[JS90] I. Johnstone and B. Silverman. Speed of estimation in positron emission tomography and related inverse problems. *Ann. Stat.*, 18:251–280, 1990.

[Lan68] H. Lancaster. The structure of bivariate distribution. *Ann. Math. Statist.*, 29:719–736, 1968.

[LvdG00] J-M. Loubes and S. van de Geer. Adaptive estimation with soft thresholding penalties. *Statistica Neerlandica*, 56, 1-26, 2002.

[LV04] J-M. Loubes and A. Vanhems. Saturation space for inverse problems in econometry. *Proceedings of the american econometric society*, 2004.

[Osu96] F. O'Sullivan. A statistical perspective on ill-posed problems. *Statist. Science*, 1, 502–527, 1996.

[2] Qinian Jin and Umberto Amato. A discrete scheme of Landweber iteration for solving nonlinear ill-posed problems. *J. Math. Anal. Appl.*, 253(1):187–203, 2001.

[R41] O. Reiersol. Confluence Analysis of Lag Moments and others Methods of Confluence. *Econometrica*, 9, 1-41, 1941.

[R45] O. Reiersol. Confluence Analysis by Means of Instrumental Sets of Variables. *Arkiv for Mathematik*, 32, 1945.

[S58] J.D. Sargan. The Estimation of Economic Relationship using Instrumental Variables. *Econometrica*. 1958.

[DS88] N. Dunford and J. Schwartz. *Linear operators. Part III.* John Wiley & Sons Inc., New York, 1988. Spectral operators, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1971 original, A Wiley-Interscience Publication.

[3] Ulrich Tautenhahn. On a general regularization scheme for nonlinear ill-posed problems. II. Regularization in Hilbert scales. *Inverse Problems*, 14(6):1607–1616, 1998.

[TA77] A. Tikhonov and V. Arsenin. *Solutions of ill-posed problems.* V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York, 1977. Translated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics.

[Van01] A. Vanhems. Nonparametric estimation of differential equation. *preprint*, 2001.

[Var92] H.R. Varian. *Microeconomic Analysis* . W.W. Norton

[vdVW96] A. van der Vaart and J. Wellner. *Weak convergence and empirical processes.* Springer-Verlag, New York, 1996. With applications to statistics.

Jean-Michel Loubes.
Address: UMR 8628 CNRS/Paris-Sud, Bâtiment 425, Département de Mathématiques d'Orsay, Université d'Orsay Paris XI, F-91425, Orsay, Cedex, France.
E-mail: `mailto:Jean-Michel.Loubes@math.u-psud.fr`
Web-site: `http://www.math.u-psud.fr/~loubes/`

Anne Vanhems
Address: ESC Toulouse, 20 bvd Lascrosses, BP 7070 31068 Toulouse Cedex 7
E-mail: `mailto:a.vanhems@esc-toulouse.fr`