Inverse Problems and Structural Econometrics : The Example of Instrumental Variables[1]

Jean-Pierre Florens[2]

February 2002

[1]Invited Lecture to the World Congress of the Econometric Society, Seattle 2000.
Acknowlegements : This paper is strongly related to researches with M. Carrasco, S. Darolles and E. Renault. I also want to acknowledge C. Gouriéroux, L.P. Hansen, J. Heckman, C. Meghir, C. Protopopescu, J.F. Richard, E. Sbaï, A. Vanhems and E. Vytlacil for useful discussions on this topic. The reduction of this paper had benefited of numerous remarks and comments of L.P. Hansen

[2]University of Toulouse (IDEI-GREMAQ), Manufacture des Tabacs, 21, allée de Brienne, F-31000 Toulouse, France, e-mail : florens@cict.fr

Abstract

A structural functional model is characterized by a functional equation relating the infinite-dimensional parameter of interest φ and the distribution F of the sample. In linear cases this equation can be written $K_F\varphi = \psi_F$ where K_F is a linear operator. This inverse problem is said to be ill posed if the inverse of K_F does not exist or is not continuous. In that case an approximated continuous solution of this equation may be computing using a Tikhonov regularization ($\varphi = (\alpha I + K_F^* K_F)^{-1} K_F^* \psi_F$). We analyze this procedure where F is estimated non parametrically and where α decreases to zero. Applications to instrumental variable estimation are developed.

JEL Classification : C14, C30.

Keywords : Inverse problem, Tikhonov regularization, Instrumental variables, Spectral decomposition, Nonparametric estimation.

1 Introduction

The development of nonparametric estimation in econometrics has been extremely important in the last fifteen years. Inference was first concentrated on the data's distribution, described for example by its density or by its hazard function, or by some characteristics of the conditional distributions, such as the conditional expectations. This approach is typically a reduced form analysis oriented to sophisticated data description, even if the selection of conditioning variables may depend on a theoretical model. On the other side, the structural econometric analysis is focused on the estimation of the (possibly functional) parameters which describe the economic agent's behavior and which are not, in general, "simple" transformations of the sampling distribution. An excellent state of the art discussion of nonparametric econometrics is given by Pagan and Ullah (1999).

A first objective of this paper is to introduce a general framework for structural functional inference in connection with the inverse problems literature. An inverse problem is the resolution of a functional equation with a particular attention to the sensitivity of the solution to possible errors in the specification of the equation due for instance to an estimation procedure (see e.g. for recent surveys of the literature Colton et al (2000)).

We analyse more specifically linear inverse problems where the parameter of interest is a function φ solution of a linear equation $K_F\varphi = \psi_F$ in which both the linear operator K_F and the r.h.s. depend on the (unknown) distribution F of the sample. This linear problem may be ill posed if K_F does not admit a continuous inverse and this problem must be regularized (see Tikhonov and Arsenin (1977) or Wahba (1973)).

One of the fundamental question of structural econometrics is the treatment of endogeneity. This question is addressed in terms different from the definition of exogeneity based on the notion of cut (see Engle et al (1983) and Florens and Mouchart (1985)). The problem is to define a relation (such as $Y = \varphi(Z) + U$) in absence of the "exogeneity assumption" ($E(U|Z) = 0$). Different possible definitions are given in the paper and the instrumental variable definition is featured ($E(U|W) = 0$ where W is a set of instruments). This presentation is more in the tradition of Frisch (1934) and (1938), Reiersol (1941 and 1945), Sargan (1958), Basman (1959) or Theil (1953).

Pratical implementation of the solution of a linear inverse problem is developed and we finally present a synthesis of some of our previous works on the asymptotic properties of the Tikhonov regularization of the solution of an ill posed linear inverse problem.

2 Functional structural econometrics and inverse problems

A structural analysis in nonparametric (i.e. with functional parameters) econometrics can be introduced by considering the following elements.

i) The functional parameter of interest is denoted by φ and this unknown function is an element of a Banach space Φ.

ii) The observation mechanism is characterized by a random element S, which is in general a random vector in $\mathbb{R}^m$ but S could be also infinite dimensional. The probability measure of S is defined by a cumulative distribution function F. This c.d.f. is an element of a topological space $\mathcal{F}$. The econometrician observes a sample $s_1, ..., s_N$ of S. In the last sections of this paper we essentially consider i.i.d. samples but the extension to weakly dependent observations (e.g. strong mixing stationary processes) does not deeply modify our analysis.

iii) The economic model defines the parameter of interest φ and connects this parameter to the probability distribution F of the sample by a functional equation :

$$A(\varphi, F) = 0, \tag{1}$$

where A is an operator defined on $\Phi \times \mathcal{F}$ and valued in a Banach space $\mathcal{E}$. The main feature of this presentation is that φ is implicitely related to F which allows to set fundamental questions of structural econometrics as identification (unicity of the solution of (1) for given F) or overidentification (existence of the solution). Statistical nonparametric inference or reduced form analysis are in general concerned by explicit definitions of the parameter of interest, like the regression function or the cumulative hazard function for example.

In this paper we call *Structural Functional Model* the three elements $\Phi, \mathcal{F}$ and A. This definition will be illustrated by the following examples. In this section, only nonlinear examples are given. Linear examples will be considered in section 2. ■

Example 2.1 *(Conditional moment condition)*
This example covers a large class of particular cases. It gives a natural way to specify a relation between φ and F. Let assume $S = (Y, Z) \in \mathbb{R}^m$ a random vector and h is an operator defined on $\mathbb{R}^m \times \Phi$ and valued in $\mathbb{R}^r$. We assume that h is integrable for any φ and we defined A by :

$$A(\varphi, F) = E^F(h(S, \varphi)|Z = z).$$

The usual (conditional) moment condition is obtained where φ is finite dimensional ($\Phi \subset \mathbb{R}^k$) and this example also covers the marginal moment condition $E^F(h(S, \varphi)) = 0$. Following the Hansen (1982) paper, a huge literature examines this conditions (see e.g. Hall (1993)).

Most of this literature considers finite dimensional parameters and finite number of moment conditions but infinite dimensional extensions are given by Carrasco and Florens (2000a).

Moment or conditional moment conditions are in general derived from economic models by assuming that the first order conditions of optimisation programs which characterized the behavior of economic agents are satisfied in average(see e.g. Ogaki (1993)).

■

Example 2.2 *(Surplus analysis and non linear differential operators)*
Let us assume $S = (Y, Z, W) \in \mathbb{R}^3$ and define

$$m_F(z, w) = E(Y[Z = z, W = w).$$

This conditional expectation function is assumed to be smooth and the parameter of interest φ is a differentiable function from $\mathbb{R}$ to $\mathbb{R}$. This function is assumed to be solution of a Cauchy-Lipschitz differential equation

$$\varphi'(z) = m_F(z, \varphi(z)),$$

under a boundary condition $\varphi(z_0) = a_0$. In that case

$$A(\varphi, F) = \varphi' - m_F(., \varphi),$$

and $\mathcal{E}$ is the set of real variable real valued functions.

Nonparametric estimation of the surplus function of a consumer gives an example of this functional equation. Following Hausman (1981 and 1985), Hausman and Newey (1995) the surplus function φ satisfies the equation

$$\varphi'(z) = m_F(z, w_0 - \varphi(z)),$$

where Y is the consumption of a good, Z the price, W the revenue of the consumer, m_F the demand function and (z_0, w_0) an initial value of the price and the revenu. The boundary condition assumes that $\varphi(z_0) = 0$. A general treatment of functional parameters solutions of Cauchy-Lipschitz differential equations and others applications are given by Vanhems (2000).

■

Example 2.3 *(Game theoretic model)*
We consider here incomplete information symmetric games which can be simplified in the following way. A player of a game receives a private signal $\xi \in \mathbb{R}$ and plays an action $S \in \mathbb{R}$. We consider cases where the ξ are i.i.d. generated for all the players and all the games and the distribution of any ξ, characterized by its c.d.f. φ, is common knowledge for the players. Actions are related to signals by a strategy function

$$S = \sigma_\varphi(\xi),$$

which is obtained, for example, as a Nash equilibrium and depends on the c.d.f. φ. For simplicity σ_φ is supposed to be one to one and increasing. The econometrician observes a sample of the action S but ignores the signals and the parameter of interest is φ. The strategy function (as a function of ξ and φ) is known. Let F be the c.d.f. of the actions. This distribution satisfies $F = \varphi \circ \sigma_\varphi^{-1}$ and the operator A can be defined by :

$$A(\varphi, F) = \varphi - F \circ \sigma_\varphi.$$

The private value first price auction model gives a particular case of this class of examples. In this case, the strategy function verifies :

$$\sigma_\varphi(\xi) = \xi - \frac{\int_{\xi_0}^{\xi} \varphi(u)^K du}{\varphi(\xi)^K},$$

where the number of bidders is $K+1$ and $\xi \in [\xi_0, \xi_1] \subset \mathbb{R}$. This example was treated in numerous papers (see Guerre et al (2000)) for a recent nonparametric analysis). A general treatment of the game theoretic models (including several extensions) is given by Florens et al (1997).

■

For a given F, φ is identified if two solutions of (1) are necessarily equal and φ is locally identified if, for any solution, there exists a neighborhood in which no other solution exists. Local identification is a useful concept on non linear cases. If A is differentiable in the Frechet sense, the implicit function theorem (for a discussion of several differentiability concepts in relation with the implicit function theorem see Van der Vaart and Wellner (1996)) gives a sufficient condition for local identifiability. If (φ, F) satisfies $A(\varphi, F) = 0$ let us compute the Frechet derivative of A with respect to φ at (φ, F). This derivative is a linear operator from Φ to $\mathcal{E}$ and if this linear operator is one to one, local identification in a neighborhood of φ is warranted (For application at the game theoretic models Florens et al (1997) of Florens and Sbai (2000)).

Identifiability or local identifiability is typically a property of F. Its analysis in specific models should exhibit conditions on F that imply identification. It is natural to construct models such that identification is satisfied for the true c.d.f. (i.e. the Data Generating Process). However in numerous particular cases, identification is not verified for the estimated $\hat{F}_N$ (which is in general the empirical c.d.f. or a smooth regularization of the empirical c.d.f.). Linear models will provide examples of this lack of identification and solutions will be given Section 4.

Existence of a solution to equation (1) is also a property of F. If a solution exists for F in a strict subset of $\mathcal{F}$ only, the model will be said overidentified. In that case, it is natural to assume that the true D.G.P. F_0 satisfies the existence condition but in general the equation $A(\varphi, \hat{F}_N) = 0$ has no solution where $\hat{F}_N$ is an usual unconstrained estimator.

If there exists a neighborhoud of the true F_0 such that a solution of $A(\varphi, F) = 0$ exists for any F_* in this neighborhoud and if $\hat{F}_N$ converges to F_0 (relatively to the same topology) then overidentification will necessarily disappear for finite (possibly large) sample size and is not a major issue (this is for example the case in the private value first price auction model). However, in general, a solution does not exist for any sample size. Two types of treatments to this problem are adopted (see Manski (1988)). The first one consist in a modification of the original definition of the parameter of interest (e.g. φ becomes the *argmin* of $||A(\varphi, F)||$ instead of the solution of $A(\varphi, F) = 0$ or is the solution of a new functional equation $A_*(\varphi, F) = 0$ which extend the original one). This solution is essentially adopted in the GMM estimation and our analysis belongs to this methodology. A second way to beat overidentification is to constrain the estimation of F in order to satisfy existence conditions. This is done in finite dimensional parameter estimation by using unequal weights to the observations (see Owen (1990) Quin and Lawless (1994) and Kitamura and Stutzer (1997)).

3 Linear inverse problems

We analyse in this section particular models where the equation $A(\varphi, F) = 0$ is linear (up to an additive term) in φ.

The presentation will be simplified by assuming that Φ is an Hilbert space. Let us consider an other Hilbert space Ψ. A linear structural model is defined by an equation :

$$A(\varphi, F) = K_F \varphi - \psi_F = 0, \tag{2}$$

where $\psi_F \in \Psi$ and K_F is a linear operator from Φ to Ψ. Both the linear operator and the constant term depend in general on $F \in \mathcal{F}$.

Linear operators constitue a very large class of transformations of φ. Important families of operators are integral operators and differential operators and the properties of equation (2) will depend on topological properties of K_F (continuity, compactness...). This diversity will be illustrated by the following examples.

Example 3.1 *(Density)*
As noticed by Hardle and Linton (1994) density estimation may seen as a linear inverse problem defined, in the real case ($S \in \mathbb{R}$), by

$$\int_{-\infty}^{s} \varphi(u)du = F(s).$$

In that case φ is the density of F w.r.t. the Lebesgue measure and K_F is an integral operator. This presentation is interesting because it will be used to point out that density estimation is an ill posed problem (in a sense which will be defined later on).

■

Example 3.2 *(Differential operators)*
Let us assume that φ is a continuously differentiable function from $\mathbb{R}$ to $\mathbb{R}$ and that the model is characterized by :

$$\varphi^{(p)} + \alpha_{1F}\varphi^{(p-1)} + ... + \alpha_{P_F}\varphi = \psi_F,$$

where $\varphi^{(k)}$ is the k-th derivative of φ and α_{JF} are functions dependent on F. The solution is constrained to a set of limit conditions. Extensions to partial differential operators in case of functions of several variables can also be consider. This case covers estimation of integral of the regression ($S = (Y, Z, W) \in \mathbb{R}^z$ $\psi_F = (z, w) = E(Y|Z = z, W = w)$ $\varphi : \mathbb{R}^2 \to \mathbb{R}$ solution of $\frac{\partial}{\partial z}\varphi = \psi_F$ with $\varphi(z_0, w_0) = y_0$ see Florens and Vanhems (2000) for an application). Extension to some partial differential equations is given in dynamic models by Banon (1978) and Aït-Sahalia (1996).

■

Example 3.3 *(Backfitting estimation in additive nonparametric regression)*
Let $S = (Y, X_1, X_2) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q$. The parameter of interest φ is $(\varphi_1, \varphi_2) \in L^2(X_1) \times L^2(X_2)$($L^2(X_1)$ and $L^2(X_2)$ are the Hilbert spaces of square integrable functions of X_1 and X_2 respectively). The underlying probability measure used for the definition of these space is the D.G.P. The functions φ_1 and φ_2 are defined as the functions which minimize $E(Y - \varphi_1(X_1) - \varphi_2(X_2))^2$ or

equivalently which are solution of the linear inverse problem (see e.g. Hastie and Tibshirani (1990)).

$$\begin{cases} \varphi_1(x_1) + E(\varphi_2(X_2)|X_1 = x_1) = E(Y|X_1 = x_1) \\ \\ E(\varphi_1(X_1)|X_2 = x_2) + \varphi_2(x_2) = E(Y|X_2 = x_2) \end{cases} .$$

■

Example 3.4 *(Repeated measurement model)*
This example is very closed of the previous one. Suppose we have two ways to measure an unobserved value η*. The measurement equation are given by* $Y_j = \eta + \varphi(X_j) + u_j (j = 1, 2)$ *where* u_j *is a zero mean (given the* X_j*'s) error term and* φ *is a bias function depending on observable variables* X_j*. The order of the measurements is not relevant* $((Y_1, X_1, Y_2, X_2)$ *is distributed as* $(Y_2, X_2, Y_1, X_1))$*. We observe an i.i.d. sample of* (Y_1, X_1, Y_2, X_1) *corresponding to an i.i.d. sample of* η*. The unknown value* η *is eliminated by difference and if* $Y = Y_2 - Y_1$*, it follows that* $E(Y|X_1 = x_1, X_2 = x_2) = \varphi(x_2) - \varphi(x_1)$ *(where* φ *is a square integrable function). This relation implies that* φ *is solution of :*

$$\varphi(x) - E(\varphi(X_1)|X_2 = x) = E(Y|X_2 = x),$$

which is a particular case of (2). The function φ *is then used to forecast* η *by* $Y_j - \varphi(X_j)$*. See for details and applications Gaspar and Florens (1998)).*

Note that if the joint distribution of (X_1, X_2) *has a density* $f(x_1, x_2)$ *w.r.t. the Lebesgue measures the previous equation may rewrite :*

$$\varphi(x) - \int \varphi(u) \frac{f(u, x)}{f(x)} du = r_F(x),$$

where $r_F(x) = E(Y|X_1 = x)$*. This equation is a Fredholm type II equation (see e.g. Tricomi (1985), Debrath and Mikusinski (1999)). The system of equations which characterize* φ_1 *and* φ_2 *in example III-3 is also a Fredholm type II integral equation.*

■

Example 3.5 *The following example is motivated by the extension of GMM to a continuous number of moment conditions. It also apply to regressions with a continuous number of regressors.*

Let us consider $u(S, t)$ *a function of* $t \in [0, 1]$ *dependent on the random element* S *and* $h(S, \tau, t)$ *a function of* $(\tau, t) \in [0, 1] \times [0, 1]$ *also* S *dependent. The parameter of interest is a function* $\varphi(t)$ *defined on* $[0, 1]$*, real valued and solution of :*

$$\int E^F(h(S,\tau,t))\varphi(\tau)d\tau = E^F(u(S,t)).$$

This equation is a Fredholm type I integral equation. It is the natural extension of a linear equation system from finite dimension to infinite dimension. Despite this simplicity this type of equation rises complex question as we will see in section 4. This equation is motivated by models with continuous numbers of regressors. Consider for example the model

$$Y = \int_0^1 X(t)\beta(t)dt + U,$$

where the regressors are indexed by $t \in [0,1]$. *In this model the random elements is equal to* $(Y, X(.))$ *where* Y *is real and* X *is a random element of the set* $L^2_{[0,1]}$ *of square integrable function defined on* $[0,1]$ *provided with the uniform measure. The model assume* $E(Y|X) = \langle X, \beta \rangle$ *where* β *is a element of* $L^2_{[0,1]}$. *This moment condition implies*

$$\int E(X(\tau)X(t))\beta(t)dt = E(YX(\tau)),$$

and may be treated on a particular case of the previous relation.

The GMM with a continuous number of moments conditions also gives a motivation for this kind of equation. Let us consider a moment condition

$$E(h(S,\theta,t)) = 0,$$

where S *is a random element,* θ *a vector of parameter and* $t \in [0,1]$ *indexes the moment conditions.*

The overidentification issue is solved by replacing this equation h_j *by the minimisation of*

$$\int E(h(S,\theta,t))E(h(S,\theta,\tau))k(t,\tau)dsd\tau,$$

where k *is a weighting linear operator. Optimal GMM are considered by Carrasco and Florens (2000a) and are shown to be the solution of the minimisation of*

$$\int E(h(S,\theta,t))\varphi(\theta,t)dt,$$

where $\varphi(\theta,t)$ *is solution of*

$$\int E(h(S,\theta,t)u(S,\theta,\tau))\varphi(\theta,t)dt = E(h(S,\theta,\tau)).$$

The formalisation of the inversion of the variance of the moment conditions then lead to a linear integral equation which is a part of the implementation of Optimal GMM.

■

The class of linear inverse problem is very large, it covers cases with very different statistical properties.

Identification and overidentification can be reformulated in the linear case. The function φ is identified if K_F is one to one and this property is equivalent to $\mathcal{N}(K_F) = \{0\}$ where $\mathcal{N}(K_F)$ is the nullset of K_F. A solution of equation (2) exists if ψ_F belongs to the range of K_F (denoted $\mathcal{R}(K_F)$).

The main question raised by a linear equation is the existence of the inverse of K_F and if its continuity. Intuitively we want to estimate φ by $\hat{\varphi}_N = K_{\hat{F}_N}^{-1}\psi_{\hat{F}_N}$. This computation requires an inversion of $K_{\hat{F}_N}$ and the continuity of $K_{\hat{F}_N}^{-1}$ because even if $\psi_{\hat{F}_N}$ and $K_{\hat{F}_N}$ converge to ψ_{F_0} and K_{F_0}, this continuity is necessary for the consistency of $\hat{\varphi}_{\hat{F}_N}$. This continuity is not always satisfied because a linear operator is not necessarily continuous in the infinite dimension case.

A linear inverse problem is said well posed if K_F^{-1} exists and is continuous (This notion is due to Hadamard (see e.g. Nashed and Wahba (1974) and Tikhonov and Arsenin (1977)).). This problem is ill posed otherwise. As we will see later on, some of important econometric issues, like instrumental variables estimation, define ill posed inverse problems.

4 Ill Posed Linear Inverse Problems

Let two Hilbert spaces Φ and Ψ and $\mathcal{F}$ a family of c.d.f. of a random element S. We simplify our presentation by considering Hilbert spaces and not Banach spaces. Hilbert spaces are self adjoint and we can used orthonormal basis and spectral decomposition of operators. On a statistical viewpoint convergences will be obtained in norm and normal distributions in Hilbert spaces are more easy to deal with than in Banach spaces. In many examples Φ and Ψ are L^2 type functions spaces and their topological structure is dependent on a probability measure. Suppose that the definition of the sets Φ and Ψ is construct in such a way that these sets do not depend on the probability F in $\mathcal{F}$ (for example all the F have a support included in a compact set of $\mathbb{R}^m$ and Φ and Ψ are spaces of continuous functions) but the scalar product is relative to the true DGP F_0 in $\mathcal{F}$.

We consider a linear inverse problem $K_F\varphi = \psi_F$ where K_F is a linear operator from Φ to Ψ and ψ_F is an element of Ψ.

We restrict our analysis to an important but specific case of operators.

Hypothesis 4.1 $\forall F \in \mathcal{F} \quad K_F$ *is a compact operator.*

■

Recall that K_F is compact if the closure of the image of the closed unit sphere is compact. We give in the application to instrumental variables an interpretable sufficient condition which implies compactness of an operator.

A compact operator is bounded ($\sup_{||\varphi|| \leq 1} ||K_F \varphi||$ finite) or equivalently continuous. Its dual operator K_F^* (from Ψ to Φ) (characterized by $\langle K_F \varphi, \psi \rangle = \langle \varphi, K_F^* \psi \rangle$) is also compact and the two self adjoint operators $K_F^* K_F$ (from Φ to Φ) and $K_F K_F^*$: (from Ψ to Ψ) are also compact.

Compact operators have only a discrete spectrum. More precisely their exits two orthonormal families $(\varphi_{jF})_{j=0,1,2,...}$ and $(\psi_{jF})_{j=0,1,2,...}$ of Φ and Ψ and a sequence of decreasing positive numbers $\lambda_{0F} \geq \lambda_{1F} \geq ... > 0$ such that

$$
\begin{aligned}
& K_F^* K_F \varphi_{jF} = \lambda_{jF}^2 \varphi_{jF} \quad K_F K_F^* \psi_{jF} = \lambda_{jF}^2 \psi_{jF} \\
& K_F \varphi_{jF} = \lambda_{jF} \psi_{jF} \quad K_F^* \psi_{jF} = \lambda_{jF} \varphi_{jF} \\
& \forall \varphi \in \Phi \ \varphi = \sum_{j=0}^{\infty} \langle \varphi, \varphi_{jF} \rangle \varphi_{jF} + \bar{\varphi}_F \text{ where } K_F \bar{\varphi_F} = 0 \\
& \forall \psi \in \Psi \ \Psi = \sum_{j=0}^{\infty} \langle \psi, \psi_{jF} \rangle \psi_{jF} + \bar{\psi}_F \text{ where } K_F^* \bar{\psi}_F = 0.
\end{aligned}
\tag{3}
$$

The spectrums of $K_F^* K_F$ and if $K_F K_F^*$ are discrete and included in $\{\lambda_{0F}^2, \lambda_{1F}^2, ...\} \cup \{0\}$. If K_F is one to one the spectrum of $K_F^* K_F$ reduces to the family of λ_{jF}^2 but 0 may be an eigenvalue of $K_F K_F^*$.

Let us come back to the equation $K_F \varphi = \psi_F$. A unique solution exists if K_F is one to one. Compact operators have a range in general strictly smaller than the space Ψ (in particular if $\Phi = \Psi$ a compact operator can be onto Φ only if Φ has a finite dimension (See Wahba (1973)) and then a solution to $K_F \varphi = \psi_F$ does not exist in general. We denote as before $\mathcal{F}_0$ the set of F such that a unique solution exists and the true c.d.f. F_0 is assumed to be an element of $\mathcal{F}_0$. If $F \in \mathcal{F}_0$ 0 is not an eigenvalue of $K_F^* K_F$. In that case we can compute the solution using the decompositions given in (3).

First, let us write :

$$
K_F \varphi_F = \sum_{j=0}^{\infty} \lambda_{jF} \langle \varphi, \varphi_{jF} \rangle \psi_{jF},
$$

because as ψ_F is an element of the range of K_F, $\bar{\psi}_F$ must be 0. Then using the unicity of decomposition on the φ_{jF}, we have :

$$\lambda_{jF}\langle\varphi, \varphi_{jF}\rangle = \langle\psi, \psi_{jF}\rangle,$$

and

$$\varphi_F = \sum_{j=0}^{\infty} \frac{1}{\lambda_j} \langle\psi_F, \psi_{jF}\rangle \varphi_{iF}. \qquad (4)$$

A solution exists if and only if this series converges.

If K_F is not one to one and/or ψ_F does not belong to the range of K_F inversion of K_F may be replaced by generalized inversion. Equivalently, it can be proved (see e.g. Luenberger (1969)) that if ψ_F belongs to $\mathcal{R}(K_F) + \mathcal{N}(K_F^*)$ their exists a unique function φ of minimal norm which minimizes $||K_F\varphi - \psi_F||$. This solution may be decomposed into

$$\varphi_F = \sum_{j/\lambda_j \neq 0} \frac{1}{\lambda_j} \langle\psi_F, \psi_{jF}\rangle \varphi_{jF}. \qquad (5)$$

This series converges under the assumption $\psi_F \in \mathcal{R}(K_F) + \mathcal{N}(K_F^*)$. Let $\mathcal{F}_*$ the set of F such that $\psi_F \in \mathcal{R}(K_F) + \mathcal{N}(K_F^*)$. $\mathcal{F}_*$ contains $\mathcal{F}_0$ because if $F \in \mathcal{F}_0$ $\psi_F \in \mathcal{R}(K_F)$. However the condition $\psi_F \in \mathcal{R}(K_F) + \mathcal{N}(K_F^*)$ is not always satisfied. It is always true that $\Psi = \overline{\mathcal{R}(K_F)} + \mathcal{N}(K_F^*)$ but $\mathcal{R}(K_F)$ is not closed in general.

As we will see in the examples, usual estimators of F define operators $K_{\hat{F}_N}$ with a finite dimensional range. This range is then closed and $\hat{F}_N$ is an element of $\mathcal{F}_*$.

The inverse of a compact operator or the generalized inverse are not continuous operators. A small perturbation of ψ_F in the direction of a ψ_{jF} corresponding to a small λ_{jF} will generate a large perturbation of φ. Then even if K_F is know and if ψ_F only is estimated, the estimation of φ obtained by replacing ψ_F by $\psi_{\hat{F}_n}$ is in general not consistent. Examples later on will illustrate this problem.

A regularization is then necessary to obtain consistent estimation. In this paper we privilege the so called Tikhonov regularization methods. Others approaches play similar roles, like the spectral cut off regularization or the Landweber-Fridman iterations which will be define but not studied on a statistical viewpoint.

Tikhonov regularization (see Tikhonov and Arsenin (1977), Groetsch (1984), Kress (1999)) generalizes to infinite dimension the well known ridge regression method used to deal with colinearity problems.[1] The initial linear equation $K_F\varphi = \psi_F$ is replaced by a modified equation

$$(\alpha I + K_F^* K_F)\varphi = K_F^* \psi_F,$$

where α is a strictly positif number and I the identity operator on Φ. If α is not an eigenvalue of $K_F^* K_F$ the linear operator $\alpha I + K_F^* K_F$ has a continuous inverse on the range of K_F^* and the solution of this equation has the following Fourier decomposition :

$$\varphi_F^\alpha = \sum_{j=0}^{\infty} \frac{\lambda_{jF}}{\alpha + \lambda_{jF}^2} \langle \psi_F, \psi_{jF} \rangle \varphi_{jF}.$$

If F is estimated by $\hat{F}_N$, previous formulae defined $\varphi_{\hat{F}_N}^\alpha$ and we will see that the norm of $\varphi_{F_0} - \varphi_{\hat{F}_N}^\alpha$ decreases to zero if α goes to zero at a suitable speed.

An equivalent interpretation of Tikhonov regularization is the following : the minimisation of $\|K_F\varphi - \psi_F\|^2$ which defines the generalized inverse is replaced by the minimisation of $\|K_F\varphi - \psi_F\|^2 + \alpha\|\varphi\|^2$ and α can be interpreted as a penalization parameter. This approach is extensively used in spline estimation for example (see Wahba (1990)). More efficient estimation may be found out of the L^2-norm analysis. The Tikhonov method uses all the eigenvalues of $K_F^* K_F$ but prevent their convergence to zero by adding the positive value α. A spectral cutoff method controls the decrease of the λ_{jF}'s by retaining only the eigenvalues greater to a given ρ :

$$\varphi_F^\rho = \sum_{\lambda_{jF} > \rho} \frac{1}{\lambda_{jF}} \langle \psi_F, \psi_{jF} \rangle \varphi_{jF}. \tag{6}$$

The Tikhonov regularization requires the inversion of $\alpha I + K^* K$ and the spectral cut off regularization requires the computation of the spectrum. These two computations may be difficult. An other regularization scheme only involves successive applications of an operator and may be implemented recursively.

[1] Using standard notations, the ridge regression estimator of a linear model $y = X\beta + u$ is defined by $\hat{\beta}_\alpha = (\alpha N I + X'X)^{-1} X'y$ where α is a positive number and I the identity matrix. This estimator is used when $X'X$ is singular or quasi singular. Bayesian analysis of linear models provides a natural interpretation of this estimator as a posterior mean of β.

Let us a positive number that $a < 1/\|K\|^2$. We call the Landweber-Fridman regularization the value

$$\varphi_F^m = \sum_{j=0}^{m} (I - aK_F^* K_F)^j K^* \psi_F.$$

This function may be computed through the following recursive relation :

$$\varphi_F^\ell = (I - aK_F^* K_F)\varphi_F^{\ell-1} + aK^* \psi_F,$$

starting by $\varphi_F^0 = 0$ and used until $\ell = m$.

Most compact operators are integral operators operating on functions of real variables. In those cases K_F is characterized by its kernel $k_F(s,t)$ ((s, t are vectors of real numbers) and

$$K_F \varphi = \int k_F(\tau, t)\varphi(\tau) d\tau. \tag{7}$$

The compactness of K_F is equivalent in that case to a more interpretable condition on k_F (k_F^2 must integrable w.r.t. z and t). Operator like $I - K_F$ i.e. :

$$(I - K_F)\varphi = \varphi(t) - \int k_F(\tau, t)\varphi(\tau) d\tau,$$

are not compact operators and their inverses are continuous. Then, inverse problems presented in examples 3.3 (backfitting) and 3.4 (measurement) are not ill posed and may be solved without regularization. We illustrate by developing previous example 3.5 a case of ill posed problem.

Example 4.1 *Let us assume that* $(s_1, ..., s_N)$ *is an i.i.d. sample of* $S \in \mathbb{R}^m$ *and the parameter of interest is a real valued continuous function* $\varphi(t)(t \in [0,1])$ *solution of :*

$$\int_0^1 E^F(v(S,\tau)v(S,t))\varphi(\tau) d\tau = E^F(u(S,t)).$$

The function h *of example 3.5. has now the product form* $h(S,\tau,t) = v(S,\tau)v(S,t)$. *If* v *is a zero mean process, the* K_F *operator is the covariance operator of* v. *As we have seen, this example covers the case of a continuous number of regressors.*

If $k_F(\tau,t) = E^F(v(S,\tau)v(S,t))$ *is a continuous function of* $(\tau,t) \in [0,1] \times [0,1]$ *it is square integrable and the operator* K_F *is an Hilbert Schmith operator and then is compact (see Dunford and Schwartz (1963)). The kernel* k_F *is symmetric. Then* K_F *is self adjoint* ($K_F = K_F^*$).

Take for example $v(S,t) = S - t$ where S is a zero mean square integrable random variable. Then $k_F(\tau,t) = E^F((S-\tau)(S-t)) = \tau t + V(V = Var(S))$. This operator is not one to one (two functions φ_1 and φ_2 such that $\int \tau\varphi_1(\tau)d\tau = \int \tau\varphi_2(\tau)d\tau$ and $\int \varphi_1(\tau)d\tau = \int \varphi_2(\tau)d\tau$ have the same image). The range of K_F is the set of affine functions. A one to one example is given by the covariance operator of a Brownian motion : let $S = (W_t)_{t\in[0,1]}$ be a Brownian motion. Assume that $v(S,t) = W_t$. Then $k_p(s,t) = s \wedge t$ whose null set is $\{0\}$ and

$$\mathcal{R}_F(K_F) = \{\psi/\psi \in \mathcal{C}^1[0,1]\psi(0) = 0 \text{ and } \psi'(1) = 0\}.$$

A natural estimator of k_F is obtained by estimating F by the empirical probability measure, i.e.,

$$k_{\hat{F}_N}(\tau,s) = \frac{1}{N}\sum_{n=1}^{N} v(s_n,\tau)v(s_n,t).$$

This kernel defines a socalled Pincherle-Goursat integral operator (or degenerated kernel (see Tricomi (1985)). This operator maps a function φ into a linear combination of the $v(s_n,t)$:

$$K_{\hat{F}_N}\varphi = \frac{1}{N}\sum_{n=1}^{N} v(s_n,t)\int_0^1 v(s_n,\tau)\varphi(\tau)d\tau,$$

and his range is the N-dimensional space spanned by the $v(s_n,t)$ (assumed to be linearly independent). Then, even if K_F is one to one for the true value F_0, the estimated operator $K_{\hat{F}_N}$ is not one to one and only N eigen values of $K_{\hat{F}_N}K_{\hat{F}_N}$ are not equal to zero. Moreover the estimator of the right hand side of the equation is equal to :

$$\psi_{\hat{F}_N} = \frac{1}{N}\sum_{n=1}^{N} u(s_n,t),$$

and is not in general in the range of $K_{\hat{F}_N}$. The generalized inverse solution reduces in that case to solve the linear system $A\underline{\varphi} = b$ where A is the $N \times N$ matrix of general element $\frac{1}{N}\int_0^1 v(s_j,\xi)v(s_n,\xi)d\xi$, b is the vector of general element $\frac{1}{N}\sum_n \int v(s_j,\xi)u(s_n,\xi)d\xi$ and $\underline{\varphi}$ is the vector of $\int \varphi(\tau)v(s_n,\tau)d\tau$.

This procedure is analogous to estimation of a model with incidental parameters (i.e. a model where a new parameter appears with each new observation) and the solution of the equation $A\underline{\varphi} = b$ cannot provided a consistent estimator.

A Tikhonov regularization of this inverse problem leads to solve the following functional equation :

$$\alpha\varphi(t) + \frac{1}{N^2}\sum_{j=1}^{n} v(s_j,t)\sum_{n=1}^{N}\int v(s_j,\xi)v(s_n,\xi)d\xi \times \int \varphi(\tau)v(s_n,\tau)d\tau$$

$$= \frac{1}{N^2}\sum_{j=1}^{n} v(s_j,t)\sum_{n=1}^{N}\int v(s_j,\xi)u(s_n,\xi)d\xi.$$

This functional equation can be solved in two steps. First multiplying by $v(s_\ell,t)$ and integrating w.r.t. t gives a linear $N \times N$ system where unknown variables are the $\int \varphi(\tau)v(s_n,\tau)d\tau$. This system can be solved and $\varphi(t)$ is then obtained from the above expression. This example shows that even if expression in term of Fourier decomposition are useful for analyzing the properties of the estimator, practical computations may be realized by inversion of finite dimensional linears systems.

■

5 Relation between endogenous variables

Let us assume that the observed random vector S can be decomposed into $(Y,Z,X,W) \in \mathbb{R}\times\mathbb{R}^p\times\mathbb{R}^k\times\mathbb{R}^q$. The assumptions derived from the economic models are the following. First X and W are exogenous. This means that no information on the parameter of interest is carried by the marginal distribution generating X and W or equivalently that the parameter of interest may be deduced without loss of information from the conditional distribution of Y and Z given X and W. The second economic assumption says that the parameter of interest is a function of φ (or a transformation of this function) and X which satisfies a relation of Z

$$Y = \varphi(Z,X) + U. \tag{8}$$

Equation (2) involves a normalization (a general function would be $\nu(Y,Z,X,U) = 0$), an additive structure for the residuals and an exclusion of W variables.

These assumptions are not sufficient to characterize φ in an unambiguous way and they need to be completed by an assumption on the residual. This assumption must preserve the endogeneity of both Y and Z. Three different hypothesis have been used in the literature. All these hypothesis define φ as the solution of a linear inverse problem and we call respectively the three possible characterizations of φ Instrumental Variables (IV) definition, Local Instrumental Variable (LIV) definition or Control Function (CF) definition.

i) IV assumption :

This assumption is written :[2]

$$E^F(Y-\varphi(Z,X)|X,W) = E^F(Y-\varphi(Z,X)|X). \tag{9}$$

Usual assumption on IV regression assumed that the conditional expectation of U given all the exogenous variables (X,W) is zero. Following Heckman and Vytlacil (1999) we relax this condition and $E(U|X,W)$ may be function of X.

The main interest of this assumption is to consider a case where (W,X) is not exogeneous if φ is the parameter interest (because $E(U|X,W)\neq 0$) but (X,W) becomes exogeneous if the derivatives of φ with respect to Z are the parameters of interest (See for an application Dearden et al (2002)).

The function φ is solution of a linear inverse problem

$$K_F^{IV}\varphi = \psi_F^{IV},$$

where

$$K_F^{IV}\varphi = E(\varphi(Z,X)|X,W) - E(\varphi(Z,X)|X),$$

and

$$\psi_F^{IV} = E(Y|X,W) - E(Y|X).$$

Using conventional notations for the densities of probability measures $K_F^{IV}\varphi$ may be write :

$$(K_F^{IV}\varphi)(x,w) = \int \varphi(z,x)\{f(z|x,w) - f(z|x)\}dz,$$

and is an integral operator whose kernel is equal to $f(z|x,w)-f(z|x)$.

This linear operator is not one to one because functions of X only are elements of the null space of K_F^{IV}. If the econometrician is interested by

[2]In order to simplify our presentation we can assume that all c.d.f. we consider have the same compact support in $\mathbb{R}^{1+p+k+q}$ and are all equivalent (i.e. have the same null sets) to the Lebesgue measure on this compact. The functions of Random vectors we consider are continuous functions of their arguments. Then all the a.s. equalities becomes equalities everywhere.

the relation between Z and Y it is sufficient to know φ up to functions of X (see the literature on "Average treatment effect" (ATE) : Imbens and Angrist (1994), Heckman and Vytlacil (1999). Under regularity assumptions this means that the partial derivatives of φ w.r.t. z are identified.

Identification issue is then to describe models in which $\mathcal{N}(K_F^{IV})$ reduces to $L^2(X)$. This condition is equivalent to the property "any function of (Z, X) whose expectation given (X, W) is zero is in $L^2(X)$". This kind of condition was introduced in the analysis of relations between sufficient and ancilary statistics. Connection with identification of IV models and interpretation of this condition is a rank condition were pointed out by Newey and Powell (1989) and Darolles, Florens and Renault (2000). Extensive analysis of this concept, under the name "Strong identification" can be found in Mouchart, Rolin (1984) and in Florens et al (1990).

ii) LIV assumption

If we assume differentiability of φ and of conditional expectations we consider, φ satisfies LIV hypothesis if :

$$E\left(\frac{\partial\varphi}{\partial z_j}(z,x)|X=x,W=w\right)=\frac{\frac{\partial}{\partial w_\ell}E(Y|X=x,W=w)}{\frac{\partial}{\partial W_\ell}E(Z_j|X=x,W=w)} \qquad (10)$$

$\forall j = 1, ..., p \quad \ell = 1, ..., q.$

This definition extends naturally the linear case and can be interpreted easily. Discrete z was considered originally by Heckman and Vytlacil (1999) and discrete z and variations of w (instead of derivatives) was introduced by Imbens and Angrist (1994) and called LATE (Local Average Treatment Effect).

This equation introduces an overidentification constraint because the r.h.s. must be identical for any $l = 1, ..., q$. This condition is satisfied if $E(Y|X, W) = E(Y|X, m(X, W))$.

The function φ is the solution of a linear inverse problem where $K_F = T_F D$ with $D\varphi$ is the vector of partial derivatives of φ w.r.t the coordinates of Z and T_F is the conditional expectation operator ($\lambda(Z, X) \to T_F\lambda = E(\lambda(Z, X)|X, W)$).

This operator K_F cannot be one to one and under a regularity condition[3], it contains all the function of X. Conversely if Z is strongly identified by Wgiven X, T_F is one to one and the null set of K_F reduces to $L^2(X)$.

[3]The distribution of (Z, X) must be such that the derivative w.r.t. z_j of a function a.s. equal to a function of X must be 0, or equivalently if a function of Z is a.s. equal to a function of X if and only it is a.s. constant : this define Z and X measurably separated. (see Florens et al (1990))

iii) CF assumption

We assume there exists a function $V(Z,X,W)$ such that the information contained by Z,X,W and by V,X,W are identical (e.g. $V = Z - m(X,Z))$) and

$$E(U|Z,X,W) = E(U|V,X).$$

Consequently if $h(V,X) = E(U|V,X)$ one has :

$$E(Y|Z,X,W) = \varphi(Z,X) + h(V,X). \tag{11}$$

This assumption was used in several parametric contexts (see Heckman (1979)) and was systematically analyzed by Newey, Powell and Vella (1999).

This model is an additive regression model which implies that φ is a solution of the set of equations :

$$\begin{aligned} \varphi(Z,X) + E(h(V,X)|Z,X) &= E(Y|Z,X) \\ E(\varphi(Z,X)|V,X) + h(V,X) &= E(Y|V,X). \end{aligned}$$

Then φ is solution of :

$$\begin{aligned} &\varphi(Z,X) - E(E(\varphi(V,X)|Z,X) \\ &\quad = E(Y|Z,X) - E(E(Y|V,X)|Z,X). \end{aligned} \tag{12}$$

Equation (12) can be rewritten $K_F\varphi = \psi_F$ where $K_F = I - A_F^* A_F (A_F : L^2(Z,X) \ni \lambda \to E(\lambda|V,X) \in L^2(V,X)$ and $A_F^* : L^2(V,X) \ni \mu \to E(\mu|Z,X) \in L^2(Z,X))$

The operator K_F cannot be one to one because here also its null space contains the functions of X.

A pointed out by Newey, Powell and Vella (1999) $\mathcal{N}(K_F)$ contains only function of X if V and Z are measurably separated given X (see Florens et al (1990)), i.e. if any function of V and X a.s. equal to a function of Z and X is a.s. equal to a function of X. This condition is not always satisfied and can also be interpreted as a rank condition.

Remark : If F is dominated by the Lebesgue measure we have seen that IV assumption implies that φ satisfies a Fredholm type I equation. In the LIV case $D\varphi$ is also solution of this type of equations :

$$\int \frac{\partial\varphi}{\partial z_j}(z,x) f(z,x|x,w)dz = \psi_F(x,w),$$

where ψ_F is the r.h.s. of (10).

In the CF approach φ is solution of Fredholm type II equation :

$$\varphi(z,x) - \int \varphi(z,x)k(\xi,z,x) = \psi_F,$$

where now ψ_F is the r.h.s. of (12) and

$$k(\xi,z,x) = \int f(z,x|v,x)f(v,x|z,x)d\xi.$$

As we will see in the next section the properties of the solution are very different in this last case then in the first two cases.

It is easy to verify that if (Y,Z,W) are jointly normal this three problems give identical (linear) solutions. In non linear models this equivalence is no longer true and one can easily construct a model where the solutions are different (see Florens et al (2000) for example and equalities conditions)

6 Instrumental variables estimation

In order to simplify the argument we concentrate our analysis to the specific case where no exogenous variables appear in the function φ. Then, I.V. assumption becomes $E(U|W) = $ constant and φ can only be identified up to constant term. It is natural in this context to assume that $E(U) = 0$ in order to eliminate this identification problem and the case we consider now assumes :

$$E(Y - \varphi(Z)|W) = 0. \tag{13}$$

We complete this assumption by the following hypothesis on the joint probability measure on (Z,W). This hypothesis is fundamental for our spectral decomposition approach (see for a different point of view of spectral decomposition of the conditional expectation operator see Chen et al (2000)).

Assumption 6.1 *The joint distribution of (Z,W) is dominated by the product of its marginal probabilities and its density is square integrable w.r.t. the product measure.*

In the case of a probability measure dominated by the Lebesgue measure this condition is equivalent to

$$\int \frac{f^2(z,w)}{f(z)f(w)} dzdw < \infty.$$

■

Let us now denote by T_F and T_F^* its dual operator, the two conditional expectation operators :

$$T_F: \; L^2(Z) \rightarrow L^2(W) \quad T_F\varphi = E(\varphi|W) \quad \varphi \in L^2(Z)$$

$$T_F^* L^2(W) \rightarrow L^2(Z) \quad T^*\psi = E(\psi|Z) \quad \psi \in L^2(W).$$

The original problem may be denoted $T_F\varphi = r_F$ where $r_F = E(Y|W) \in L^2(W)$.

Under the assumption 6.1, T_F is a compact operator (see Breiman and Friedman (1985)) and the analysis developed in section 4 applied. I.V. estimation is an ill posed inverse problem and need a regularization procedure.

The same argument applied to LIV estimation. Take as parameter of interest the vector of partial derivatives $D\varphi$. This vector of functions is also solution to an ill posed inverse problem, $T_F D\varphi = \psi_F$ where ψ_F is defined in equation (10) and where the linear operator is compact.

Under an assumption $m(Z,V)$ analogous to the assumption on (Z,W), CF estimation leads to a well posed inverse problem and don't need a regularization. Indeed φ is solution of $(I - A_F^* A_F)\varphi = \psi_F$ (see 12). The function ψ_F is in the domain of $(I - A_F^* A_F)$ and the inverse operator is bounded and then continuous. This can be seen by using a spectral decomposition of $A_F^* A_F$ whose eigen values are denoted μ_j^2 and eigen vecteurs ε_j. Then

$$\varphi = \sum_{j=1}^{\infty} \frac{1}{1-\mu_j^2} < \psi_F, \varepsilon_j > \varepsilon_j.$$

The sum start at $j = 1$ because ε_0 is the constant function equal to 1 and $\langle \psi_F, \varepsilon_F \rangle = 0$ because ψ_F is a zero mean vector.

This serie converges in norm L^2 because

$$\sum_{j=1}^{\infty} \left(\frac{1}{1-\mu_j^2} \right)^2 \langle \psi_F, \varepsilon_j \rangle^2 \leq \left(\frac{1}{1-\mu_1^2} \right)^2 \sum_{j=1}^{\infty} \langle \psi_F, \varepsilon_j \rangle^2 \leq \left(\frac{1}{1-\mu_1^2} \right)^2 ||\psi_F||^2.$$

Finally the $Sup||(I - A_F^* A_F^*)^{-1}\psi_F||$ (where $||\psi|| \leq 1$ and $\psi \in$ Domain $(I - A_F^* A_F^*)^{-1}$ = set of zero mean vector) is smaller than $\left| \frac{1}{1-\mu_1} \right|$ which means that the inverse operator is continuous.

We conclude this section by a short description of the practical implementation of the estimation of φ in the case of I.V. assumption. The sample is $(y_n, z_n, w_n)_{n=1,\ldots,N}$ and the equation $(\alpha_N I + T_{\hat{F}_N}^* T_{\hat{F}_N})\varphi = T_{\hat{F}_N}^* r_{\hat{F}_N}$ may be simplified into :

$$\alpha_N\varphi(z) + \frac{1}{\sum_{\ell=1}^{N} H_N(z-z_\ell)} \sum_{\ell=1}^{N} \frac{\sum_{n=1}^{N} \varphi(z_n) H_N(w_\ell - w_n)}{\sum_{n=1}^{N} H_N(w_\ell - w_n)} H_N(z-z_\ell)$$

$$= \frac{1}{\sum_{\ell=1}^{N} H_N(z-z_\ell)} \sum_{\ell=1}^{N} \frac{\sum_{n=1}^{N} y_n H_N(w_\ell - w_n)}{\sum_{n=1}^{N} H_N(w_\ell - w_n)} H_N(z-z_\ell), \tag{14}$$

where H_N are usual smoothing kernel (conventionally the same letter is used for different kernels applied to the w's or the z's). This functional equation gives $\varphi(z)$ for any z knowing $\varphi(z_n)n = 1, ..., N$. Then in a first step rewrite equation (14) for $z = z_1, ...z_N$. This provides a $N \times N$ linear system which can be solved in order to obtain the $\varphi(z_n)$. The choice of α_N parameter is very important and we will see in the next section what are the contraints on its speed of convergence and how can be a choice of this parameter.

This approach avoids any computation of eigen values or eigen vectors but they are implicitely present in the resolution of the linear system. Using the same methodology than in Darolles, et al (2002) one can check that the estimator we have defined may rewrite :

$$\varphi_{\hat{F}_N}^{\alpha N} = \sum_{j=0}^{N-1} \frac{\hat{\lambda}_{j\hat{F}_N}}{\alpha_N + \hat{\lambda}^2_{\hat{F}_N}} (\frac{1}{N} \sum_{n=1}^{N} y_n \varphi_{j\hat{F}_N}(z_n)) \varphi_{j\hat{F}_N}(z), \tag{15}$$

where $\lambda^2_{j\hat{F}_N}$ are the N non null eigenvalues of $T^*_{\hat{F}_N} T_{\hat{F}_N}$ and $\varphi_{j\hat{F}_N}$ their corresponding eigenvectors.

7 Asymptotic theory for Tikhonov regularization of ill posed linear inverse problems

In this section, we concentrate our presentation on new questions raised by the linear inverse problem $K_F\varphi = \psi_F$ where K_F is a compact operator. We will then assumed asymptotic behavior of the elements of the equation (which can be difficult to verify in particular models) and we will show how their are transformed by the resolution. As announced before, we will develop an Hilbert space approach, both on consistency and on asymptotic normality.

Let φ_{F_0} be the unique solution of $K_{F_0}\varphi = \psi_{F_0}$ where F_0 is the true DGP which is an element of $\mathcal{F}_0$.

We denoted by $\varphi^{\alpha}_{F_0}$ the solution of :

$$(\alpha I + K^*_{F_0}K_{F_0})\varphi = K^*_{F_0}\psi_{F_0} = K^*_{F_0}K_{F_0}\varphi_0,$$

for any $\alpha > 0$. Given a sample $(s_1, ..., s_N)$ $\hat{F}_N$ is a estimator of F and $K_{\hat{F}_N}$ and $\psi_{\hat{F}_N}$ the corresponding estimation of K_F and ψ_F.

The properties of this estimation mechanism are given by the following assumptions :

Assumption 7.1 *$\exists a_N$ sequence in $\mathbb{R}$ $a_n \to \infty$ such that*

$$\|K^*_{\hat{F}_N}K_{\hat{F}_N} - K^*_F K_F\| \sim O\left(\frac{1}{a_N}\right).$$ [4]

■

In this assumption the norm of an operator A from Φ to Φ is defined by $sup_{\|\varphi\|\leq 1}\|A\varphi\|$ and the norm on Φ is the Hilbert norm possibly dependent on F_0.

Assumption 7.2 *$\exists b_N$ sequence in $\mathbb{R}$ $b_N \to \infty$ such that*

$$\|K^*_{\hat{F}_N}\psi_{\hat{F}_N} - K^*_{\hat{F}_N}K_{\hat{F}_N}\varphi_0\| \sim O(\frac{1}{b_N}).$$

■

This assumption replace assumption on $\psi_{\hat{F}_N}$. Intuitively $\psi_{\hat{F}_N}$ converges to ψ_{F_0} equal to $K_{F_0}\varphi_0$ but as K^*_F is a compact operator taking the image of $\psi_{\hat{F}_N} - K_{\hat{F}_N}\varphi_0$ by $K^*_{\hat{F}_N}$ regularizes the estimation and may improve the speed of convergence.

Assumption 7.3 *$\alpha_N \to 0$, $\frac{1}{\alpha_N a_N} \sim O(1)$ and $\alpha_N b_N \to \infty$.*

■

Theorem 7.1 *Under assumptions 7.1, 7.2 and 7.3 $\|\varphi^{\alpha_N}_{\hat{F}_N} - \varphi\| \to 0$ in probability.*

[4]All the equivalence are in probability w.r.t. the DGP. Almost sure equivalences will give a.s. convergence in theorem 7.1.

■

Proof : This proof is standard if the operator K_F is known and where the only error is on ψ_F (see Groetsch (1984) or Kress (1999)). Extension to estimation error on K_F generalizes the arguments developed in Carrasco and Florens (2002) and in Darolles et al (2000). The main steps of the proofs are the following :

i)
$$\|\varphi_{\hat{F}_N}^{\alpha_N} - \varphi_{F_0}\| \leq \|\varphi_{\hat{F}_N}^{\alpha_N} - \varphi_{F_0}^{\alpha_N}\| + \|\varphi_{F_0}^{\alpha_N} - \varphi_{F_0}\|,$$

and $\|\varphi_{F_0}^{\alpha_N} - \varphi_{F_0}\| \to 0$ if $\alpha_N \to 0$ (see any of the above reference).

ii)
$$\varphi_{\hat{F}_N}^{\alpha_N} - \varphi_{F_0}^{\alpha_N} = (\alpha_N I + K_{\hat{F}_F}^* K_{\hat{F}_N})^{-1} K_{\hat{F}_N} \psi_{\hat{F}_N}$$

$$-(\alpha_N I + K_{F_0}^* K_{F_0})^{-1} K_{F_0}^* K_{F_0} \varphi_0$$

$$= (\alpha_N I + K_{\hat{F}_N}^* K_{\hat{F}_N})^{-1} (K_{\hat{F}_N}^* \psi_{\hat{F}_N} - K_{\hat{F}_N}^* K_{\hat{F}_N} \varphi_{F_0})$$

$$+\alpha_N \left[(\alpha_N I + K_{\hat{F}_N}^* K_{\hat{F}_N})^{-1}) - (\alpha_N I + K_{F_0}^* K_{F_0})^{-1} \right] \varphi_{F_0}.$$

The last equality follows from the identity

$$(\alpha I + A)^{-1} A = I - \alpha(\alpha I + A)^{-1}.$$

Then $\|\varphi_{\hat{F}_N}^{\alpha_N} - \varphi_0^{\alpha_N}\| \leq I + II$ where I and II are defined and analyzed separately.

iii)
$$I = \|(\alpha_N I + K_{\hat{F}_N}^* K_{\hat{F}_N})^{-1} (K_{\hat{F}_N}^* \psi_{\hat{F}_N} - K_{\hat{F}_N}^* K_{\hat{F}_N} \varphi_{F_0})\| \leq$$

$$\|(\alpha_N I + K_{\hat{F}_N}^* K_{\hat{F}_N})^{-1}\| \|K_{\hat{F}_N}^* \psi_{\hat{F}_N} - K_{\hat{F}_N}^* K_{\hat{F}_N} \varphi_{F_0}\|.$$

The first term is majored by $\frac{1}{\alpha_N}$ (see Groetsch (1984)) and the second is $0(\frac{1}{b_N})$ by assumption 7.2. By assumption 7.3 $\alpha_N b_N \to \infty$ and $I \to 0$

iv)

$$II = \alpha_N \| \left[(\alpha_N I + K_{\hat{F}_N}^* K_{\hat{F}_N})^{-1} - (\alpha_N I + K_{F_0}^* K_{F_0})^{-1} \right] \varphi_{F_0} \|$$

$$= \|\alpha_N (\alpha_N I + K_{F_0}^* K_{F_0}^*)^{-1} \varphi_{F_0}\| \times \|K_{\hat{F}_N}^* K_{\hat{F}_N} - K_{F_0}^* K_{F_0}\| \times \|(\alpha I + K_{F_0}^* K_{F_0})^{-1}\|.$$

The first term is equal to $\|\varphi - \varphi^{\alpha_N}\|$ and has a zero limit. The second term is by assumption 7.1 is equivalent to $\frac{1}{a_N}$ and the last term is smaller than $\frac{1}{\alpha_N}$. As $\frac{1}{\alpha_N a_N} \sim O(1),\ II \to 0$.

■

Example 7.1 *Consider example 4.1. Following e.g. Carrasco and Florens (2000) a). We have* $\|K_{\hat{F}_N} - K_{F_0}\| \sim O(\frac{1}{\sqrt{N}})$. *Using the property* $K_F^* = K_F^*$ *and a first order approximation, its follows that* $\|K^2_{\hat{F}_N} - K^2_{F_0}\|$ *is also equivalent to* $\frac{1}{\sqrt{N}}$. *Moreover*

$$\|K_{\hat{F}_N}\psi_{\hat{F}_N} - K^2_{\hat{F}_N}\varphi_0\| \ \leq \ \|K_{\hat{F}_N}\| \left\{\|\psi_{\hat{F}_N} - K_{F_0}\varphi_0\| + \|\hat{K}_{\hat{F}_n} - K_{F_0}\|\|\varphi_{F_0}\|\right\}$$

which implies $b_n = \sqrt{N}$ *because* $\|\psi_{\hat{F}_N} - K_{F_0}\varphi_0\| \sim O\left(\frac{1}{\sqrt{N}}\right)$.

Then the two conditions are satisfied if $\alpha_n\sqrt{N} \to \infty$.

■

Example 7.2 *Consider the case of IV estimation. It is proved in Darolles et al (2002) that* $\frac{1}{a_N} = \frac{1}{\sqrt{Nh_N^p}} + h_N^\rho$ *where* h_N *is the bandwith of the kernel smoothing,* p *the dimension of* z *and* ρ *is the minimum between the order of the kernel and the degree of smoothness of the density of the DGP. Moreover* $\frac{1}{bN} = \frac{1}{\sqrt{N}} + h_N^\rho$. *Then the estimator is consistent if* $\frac{h_N^{2\rho}}{\alpha_N^2} \to 0$ *and* $\frac{1}{\alpha_N^2 N h_N^p} \sim O(1)$.

■

The decomposition of $\varphi_{\hat{P}_N}^{\hat{\alpha}_N} - \varphi_{F_0}$ considered in the proof of theorem 7.1 can be used to determine an optimal speed of convergence to 0 of α_N and to give a bound on the speed of convergence of $\|\varphi_{\hat{F}_N}^{\alpha_N} - \varphi_{F_0}\|$. This analyse requires an assumption of the behavior of the regularization bias $\|\varphi_{F_0}^{\alpha_N} - \varphi_{F_0}\|$ which satisfies :

$$\begin{aligned} \|\varphi_{F_0}^{\alpha_N} - \varphi_{F_0}\| &= \alpha_N(\alpha_N I + K_{F_0}^* K_{F_0})^{-1}\varphi_{F_0} && (16)\\ &= \alpha_N^2 \sum_{j=0}^{\infty} \frac{1}{(\alpha_N + \lambda_{jF_0})^2}\langle\varphi_{F_0}, \varphi_{jF}\rangle\varphi_{jF_0}. && (17) \end{aligned}$$

We will assume that φ_{F_0} is such that $\|\varphi_{F_0}^{\alpha_n} - \varphi_{F_0}\|^2 \sim O(\alpha^\beta)$.

This condition associate φ_{F_0} and K_{F_0} and is basically a condition on the relative rate of decline of the Fourrier coefficients of φ_{F_0} in the basis $\varphi_{jF}(\langle\varphi_{F_0}, \varphi_{jF_0}\rangle)$ and of the eigenvalues $\lambda^2_{jF_0}$ of the operator.

Darolles et al (2002) shows that $\beta \in]0, 2]$ and gives characteristics of particular cases of β. In case of instrumental variables the β coefficient may be interpreted as a measure of the strength or of the weakness of the instruments. Then :

$$\|\varphi_{\hat{F}_N}^{\alpha_N} - \varphi_{F_0}\|^2 = 0\left(\frac{1}{\alpha_N^2 aN} + \frac{1}{\alpha_N^2 bN}\alpha_N^\beta + \alpha_N^\beta\right),$$

and an optimal choice of α_n will egalises the behavior of the first and last term and gives :

$$\alpha_N = a_N^{-\frac{1}{\beta+2}}.$$

We need to verify that under this choice, the second term converges to 0. If it is the case $a_N^{\frac{\beta}{\beta+2}}$ gives a lower bound of the speed of convergence. In the applications given above this bound is $n^{\frac{\beta}{\beta+2}}$ (under a suitable choice of the bandwith if a kernel estimation is necessary).

The last element to be consider is the asymptotic normality of our estimator. This normality follows from the next hypothesis :

Assumption 7.4

$$b_N(K_{\hat{F}_N}^*\psi_{\hat{F}_N} - K_{\hat{F}_N}^* K_{\hat{F}_N}\varphi_{F_0}) \Rightarrow N(0,\Omega).$$

This convergence is assumed to be a functional convergence in the Hilbert space Φ and Ω is a covariance operator is this space.

■

Let assume first that K_{F_0} is known and that the parameter α is kept constant. Under this two conditions one has :

$$b_n(\varphi_{\hat{F}_N} - \varphi_{F_0}^\alpha) = (\alpha I + K_{F_0}^*)^{-1}(b_n(K_{F_0}^*\psi_{\hat{F}_N} - K_F^* K_{F_0}\varphi)),$$

converges in Φ to a zero mean normal probability whose covariance operator is equal to

$$(\alpha I + K_{F_0}^* K_{F_0})^{-1}\Omega(\alpha I + K_{F_0}^* K_{F_0})^{-1}.$$

Indeed, standard matrix computation can be extended to continuous operators.

The extension of this result to the case of an unknown operator K_F, with α constant modifies this result in the following way :

Let $B_N^\alpha = \alpha\left[(\alpha I + K_{\hat{F}_N}^\alpha K_{\hat{F}_N})^{-1} - (\alpha I + K_{F_0}^\alpha K_{F_0})\right]\varphi$.

We have obviously (see part ii) of the proof of theorem 7.1)

$$b_N(\varphi_{\hat{F}_N}^\alpha - \varphi_{F_0}^\alpha - B_N^\alpha) = (\alpha I + K_{\hat{F}_N}^* K_{\hat{F}_N})^{-1} b_N(K_{\hat{F}_N}^*\psi_{\hat{F}_N}\psi_{\hat{F}_N} - K_{\hat{F}_N}^* K_{\hat{F}_N}\varphi_0),$$

and this term converges to the same normal probability measure in Φ as if K_F is known. However a bias term has been introduced in the l.h.s. term. In the proof of theorem 7.1 we have check that in the case of α fixed $\|B_N^\alpha\|$ converges to zero at speed $\frac{1}{a_n}$. The bias term can be neglected if $\frac{b_n}{a_n}$ has a zero limit, i.e. if the operator converges at a higher speed than the r.h.s. of the equation.

If $\alpha_N \to 0$ we cannot expected asymptotic normality in a functional sense. In particular the limit when α_N decreases to 0 of the covariance operator Ω is not bounded and is not a covariance operator of an Hilbert valued normal element. Then we will look for pointwise normality instead of functional normality in the following sense. Let ζ be an element of ϕ. We will analyse asymptotic normality of

$$\nu_N(\zeta)\langle\varphi_{\hat{F}_N}^{\alpha_N} - \tilde{\varphi}, \zeta\rangle,$$

where $\tilde{\varphi}$ is a suitable function and $\nu_N(\zeta) \to \infty$.

This class of results is obtained using the following methodology.

1) Let us denoted by ξ_N the random element $b_N(K_{\hat{F}_N}^*\psi_{\hat{F}_N} - K_{\hat{F}_N}^*\varphi_{F_0})$ and by ξ its limit ($\xi \sim N(0,\Omega)$). For a given N, $M_N = (\alpha_N I + K_{F_0}^* K_{F_0}^*)^{-1}$ and

$$\varepsilon = \frac{\langle M_N\xi, \zeta\rangle}{\langle\zeta, M_N\Omega M_N\zeta\rangle^{\frac{1}{2}}} \sim N(0,1) \qquad \forall N,$$

because M_N is bounded and $M_N\xi \sim N(0, M_N\Omega M_N)$.

2) Let us first assume that K_{F_0} is known. Then

$$\frac{b_N\langle\varphi_{\hat{F}_N}^{\alpha_N} - \varphi_{F_0}^{\alpha_N}, \zeta\rangle}{\langle\zeta, M_N\Omega M_N\zeta\rangle^{\frac{1}{2}}} = \varepsilon + \frac{\langle\xi_N - \xi, M_N\zeta\rangle}{\langle\zeta, M_N\Omega M_N\zeta\rangle}.$$

Moreover

$$\frac{\langle\xi_N - \xi, M_N\rangle^2}{\langle\zeta, M_N\Omega M_N\zeta\rangle} \leq \|\xi_N - \xi\|^2 \frac{\|M_N\zeta\|^2}{\langle\zeta, M_N\Omega M_N\zeta\rangle^{\frac{1}{2}}}.$$

This term converges to zero if $\frac{\|M_N\zeta\|^2}{\langle\zeta, M_N\Omega M_N\zeta\rangle}$ is bounded because $\|\xi_N - \xi\| \to 0$ in probability. We introduce this condition as an hypothesis.

Assumption 7.5 $\zeta \in \Phi$ *is such that* $\frac{\|M_N\zeta\|^2}{\langle\zeta, M_N\Omega M_n\zeta\rangle} \sim 0(1)$

■

Remark that if ζ belongs to the finite dimensional subspace generated by $\varphi_0, ...\varphi_{N_0}$ (where $\lambda_j \neq 0 \forall j = 0, ..., N_0$) the assumption 7.5 is satisfied.

We note by

$$\nu_N(\zeta) = \frac{b_N^2}{\langle \zeta, M_N \Omega M_N \zeta \rangle},$$

the speed of convergence. And we may conclude that

$$\sqrt{\nu_N(\zeta)} \langle \hat{\varphi}_{\hat{F}_N}^{\alpha_N} - \varphi_{F_0}^{\alpha_N}, \zeta \rangle \Rightarrow N(0,1).$$

3) If K_{F_0} is not known let us consider :

$$\sqrt{\nu_N(\zeta)} \langle \varphi_{\hat{F}_N}^{\alpha_N} - \varphi_{F_0}^{\alpha_N} - B_N^{\alpha_N}, \zeta \rangle = \varepsilon + A_1 + A_2, + A_3,$$

where

$$A_1 = \frac{\langle \xi_N - \xi, M_N \zeta \rangle}{\langle \zeta, M_N \Omega M_N \zeta \rangle^{\frac{1}{2}}}, \quad A_2 = \frac{\langle \xi, (\hat{M}_N - M_N) \zeta \rangle}{\langle \zeta, M_N \Omega M_N \zeta \rangle^{\frac{1}{2}}},$$

(where $\hat{M}_N = (\alpha_N I + K_{\hat{F}_N}^* K_{\hat{F}_N})^{-1}$) and

$$A_3 = \frac{\langle \xi_N - \xi, (\hat{M}_N - M_N) \zeta \rangle}{\langle \zeta, M_N \Omega M_N \zeta \rangle^{\frac{1}{2}}}.$$

We have shown in the previous case that under Assumption 7.5 A_1 converges to zero. The term A_2 verifies has the same behavior than

$$\frac{\|\xi\| \, \|M_N\| \, \|K_{\hat{F}_N}^* K_{\hat{F}_N} - K_{F_0}^* K_{F_0}\| \, \|M_N \zeta\|}{\langle \zeta, M_N \Omega M_N \rangle^{\frac{1}{2}}}$$

$$\leq \frac{\|M_N \zeta\|}{\langle \zeta, M_N \Omega M_N \rangle^{\frac{1}{2}}} \frac{1}{\alpha_N a_N} \|\zeta\|,$$

because $\|M_N\| \leq \frac{1}{\alpha_N}$ and hypothesis 7.1

We then renforce the hypothesis 7.3 :

Assumption 7.6 $\alpha_N a_N \to \infty$.

■

This assumption implies that $A_2 \to 0$ and an analogous proof shows that $A_3 \to 0$.

Then under the previous assumptions

$$\sqrt{\nu_N(\zeta)} \langle \hat{\varphi}_{\hat{F}_N} - \varphi_{F_0}^{\alpha_N} - B_N^{\alpha_N}, \zeta \rangle \Rightarrow N(0,1).$$

4) The next step consists to find assumptions which transform the centering function. First we look for an elimination of the bias term $B_N^{\alpha_N}$.

$$\left|\sqrt{\nu_N(\zeta)}B_N^{\alpha_N}\right| = \frac{b_N\alpha_N}{\langle\zeta, M_N\Omega M_N\rangle^{\frac{1}{2}}}\langle(\hat{M}_N - M_N)\varphi_{F_0}, \zeta\rangle$$

$$\leq b_N\|\alpha_N M_N\varphi\|\|K^*_{\hat{F}_N}\hat{K}_{\hat{F}_N} - K^*_{F_0}K_{F_0}\|\frac{\|M_N\zeta\|}{\langle\zeta, M_N\Omega M_N\zeta\rangle^{\frac{1}{2}}}$$

$$\|\alpha_N M_N\varphi\| = \|\varphi_{F_0}^{\alpha_N} - \varphi_{F_0}\| \to 0.$$

We have just to impose that the product of the others terms is bounded. Using assumption 7.2 a general assumption is the following.

Assumption 7.7 $\frac{b_N}{a_N}\frac{\|M_N\zeta\|}{\langle\zeta, M_n\Omega M_N\zeta\rangle^{\frac{1}{2}}} \sim 0(1)$.

■

This assumption is satisfied under 7.5 if $\frac{b_N}{a_N} \sim 0(1)$ but this hypothesis could be too strong. If $\frac{b_N}{a_N} \to \infty$, more assumptions are needed in order to eliminate the bias term.

Then under 7.1 to 7.7 we get :

$$\nu_N(\zeta)\langle\hat{\varphi}_{\hat{F}_N} - \varphi_{F_0}^{\alpha_N}, \zeta\rangle \Rightarrow N(0,1).$$

5) Finally we want to replace $\varphi_{F_0}^{\alpha_N}$ by φ_{F_0}in the previous convergence. Recalling that $\|\varphi_{F_0}^{\varphi_N} - \varphi_{F_0}\| \sim 0(\alpha_N)$ the following assumption is required.

Assumption 7.8 $\alpha_N^2\nu_N(\zeta) \to 0$.

■

Under 7.1 to 7.6 and 7.8 we obtain :

$$\sqrt{\nu_N(\zeta)}\langle\varphi_{\hat{F}_N}^{\alpha_N} - \varphi_{F_0}, \zeta\rangle \Rightarrow N(0,1),$$

if K_{F_0} is known and

$$\sqrt{\nu_N(\zeta)}\langle\varphi_{\hat{F}_N}^{\alpha_N} - \varphi_{F_0}, -B_{F_0}^{\alpha_N}\zeta\rangle \Rightarrow N(0,1),$$

in the general case.

If moreover 7.7 is satisfied pointwise asymptotic normality without bias is satisfied :

$$\sqrt{\nu_N(\zeta)}\langle\varphi_{\hat{F}_N}^{\alpha_N} - \varphi_{F_0}, \zeta\rangle \Rightarrow N(0,1).$$

In the case developed in Example 4.1 and 7.1, all the Assumptions can be satisfied and this last pointwise normality is verified. In the case of instrumental variable estimation (example 7.2), assumption 7.7 is not true and a bias correction term must be introduced in order to get asymptotic normality.

8 Conclusion

This paper proposed a general framework for structural functional estimation and some results related to the linear compact case are given. Application to instrumental variable estimations motivates this analysis. Numerous questions are not considered. In particular, the choice of the regularization α_N in relation to optimal speed of convergence and to minimax estimation is not treated in this paper (some steps in that direction are made in Carrasco and Florens (2000)). Non linear inverse problems, some well posed linear problems, extension to dynamic models define natural extensions of this methodology. A deep discussion about the different definitions of relations between endogeneous variables is necessary for getting unambiguous non parametric estimations (see Blundell and Powell (1999) and Florens et al (2000)).

References

Aït-Sahalia, Y., (1995), "The Delta and Bootstrap Methods for Nonparametric Kernel Functionals", Discussion Paper, MIT.

Aït-Sahalia, Y. (1996), "Nonparametric Pricing of Interest Rate Derivative Securities", *Econometrica*, Vol. 64, 527-560.

Banon, G., (1978), "Nonparametric Identification for Diffusion Processes", *SIAM Journal of Control and Optimization*, 16, 380-395.

Basmann, R.L., (1959), "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equations", *Econometrica*, 25, 77-83.

Blundell, R. and J., Powell,(1999), "Endogeneity in Single Index Models", Manuscript, UCL.

Breiman, L. and J.H. Friedman, (1985), "Estimating Optimal Transformations for Multiple Regression and Correlation", *Journal of the American Statistical Association*, 80 (391), 580-619.

Carrasco, M. and J.P. Florens, (2000a), "Generalization of GMM to A Continuum of Moments Conditions", *Econometric Theory*, 16, 797-834.

Carrasco, M. and J.P. Florens (2000b) "Efficient GMM Estimation using the Empirical Characteristic Function", GREMAQ-University of Toulouse.

Chen, X., Hansen L.P. and J. Scheinkman, (2000), "Principal Components and the Long Run", Discussion paper, University of Chicago.

Colton, D., Engle H.W. McLaughin J.R. and W. Rundell (eds), (2000) *Surveys on Solution Methods for Inverse Problems*, Springer, Wien, New York.

Darolles, S., Florens J.P. and C. Gouriéroux, (1998), “Kernel Based Nonlinear Canonical Analysis”, Discussion Paper, CREST 9858.

Darolles, S., Florens J.P. and E. Renault, (1998), “Nonlinear Principal Components and Inference on a Conditional Expectation Operator”, Discussion Paper CREST.

Dearden, L. Ferri J. and C. Meghir, (2002), “The Effect of School Quality on Educational Attainment and Wages” forthcoming in *Review of Economics and Statistics*.

Debrath, L. and P. Mikusinski, (1999), *Hilbert Spaces with Applications*, Academic Press, London.

Dunford, N. and J. Schwartz, (1963), *Linear Operators 2*, Wiley, New York.

Engle, R.H., Hendry D.F. and J.F. Richard, (1983), “Exogeneity”, *Econometrica*, 51 (2) 277-304.

Florens, J.P., Heckman, J., Meghir C. and E. Vytlacil (2000), “ Instrumental Variables, Local Instrumental Variables and Control Functions”, Manuscript, University of Toulouse.

Florens, J.P. and M. Mouchart (1985), “Conditioning in Dynamic Models”, *Journal of Time Series Analysis*, 53 (1), 15-35.

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

Florens, J.P. and E. Sbaï, (2000), “Identification in Empirical Games”, GREMAQ-University of Toulouse.

Florens, J.P., Protopopescu C. and J.F. Richard, (1997), “Identification and Estimation of a Class of Game Theoretic Models”, GREMAQ-University of Toulouse.

Florens, J.P. and A. Vanhems, (2000), “Estimation non paramétrique de l’épaisseur de la couche imosphérique : application aux mesures du satellite Topex-Poséïdon”, GREMAQ-University of Toulouse.

Frisch, R., (1934), “Statistical Confluence Analysis by Means of Complete Regression Systems”, Universitetets social okonomiste institutt, Oslo.

Frisch, R., (1938), “Statistical versus Theoretical Relations in Economic Macrodynamics”, Business Cycle of Conference at Cambridge.

Gaspar, P. and J.P. Florens, (1998), “Estimation of the Sea State Biais in Radar Altimeter Measurements of Sea Level : Results from a Nonparametric Method”, *Journal of Geophysical Research*, 103 (15), 803-814.

Groetsch, C., (1984), *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Pitman, London.

Guerre, E., Perrigne I. and Q. Vuong, (2000), "Optimal Nonparametric Estimation of First-Price Auctions", *Econometrica*, 68 (3), 525-574.

Hall, A., (1993) "Some Aspects of the Generalized Method of Moments Estimation". In GS. Maddala, C.R. and H.D. Vinod, (eds), *Handbook of Statistics*, 11, 393-417, North Holland, Amsterdam.

Hansen, L., (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029-1054.

Hardle, W. and O., Linton, (1994), *Applied Nonparametric Methods*, Handbook of Econometrics, 4, 2295-2339.

Hastie, T.J. and R.J. Tibshirani (1990), *Generalized Additive Models*, Chapman and Hall, London.

Hausman, J., (1981), "Exact Consumer's Surplus and Deadweight Loss" *American Economic Review*, 71, 662-676.

Hausman, J. (1985), "The Econometrics of Nonlinear Budget sets" *Econometrica*, 53, 1255-1282.

Hausman, J. and W.K. Newey, (1995) "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss", *Econometrica*, 63, 1445-1476.

Heckman, J., (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, 47, 153-61.

Heckman, J. and V., Vytlacil (1999), "Local Instrumental Variables", Working Paper, University of Chicago.

Imbens, G. and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, 62, 467-476.

Kitamura, Y. and M. Stutzer, M., (1997), "An Information Theoretic Alternative to Generalized Method of Moments Estimation", *Econometrica*, 65, 4, 861-874.

Lehman, E.L. and H. Scheffe, (1950), "Completeness, Similar Regions and Unbiased Tests. Part I", Sankhya, 10, 219-236.

Luenberger, D.G., (1969), *Optimization by Vectro Space Methods* Wiley, New York.

Manski, C., (1988), "Analog Estimation Methods in Econometrics", Chapman and Hall, London.

Mouchart, M. and J.M. Rolin, (1984), "A Note on Conditional Independence", *Statistica*, 45 (3), 427-430.

Nashed, M.Z. and G., Wahba, (1974), "Generalized Inverse in Reproducing Kernel Spaces : An Approach to Regularization of Linear Operator Equations", SIAM *Journal of Mathematical Analysis*, Vol 5, 6, 974-987.

Newey, W., and J. Powell, (1989), "Instrumental Variables for Nonparametric Models", MIT Discussion Paper.

Newey, W., Powell J. and F. Vella, (1999), "Nonparametric Estimation of Triangular Simultaneous Equations Models", *Econometrica*, 67, 565-604.

Pagan, A. and A. Ullah, (1999), *Nonparametric Econometrics*, Cambridge University Press.

Ogaki, M., (1993), "Generalized Method of Moments : Econometric Applications". In G.S. Maddala, C.R. Rao and H.D. Vinod, (eds), *Handbook of Statistics*, 11, 455-488. North Holland, Amsterdam.

Owen, A., (1990), "Empirical Likelihood Ratio Confidence Regions", *The Annals of Statistics*, 18, 1, 90-120.

Qin, J. and J. Lawless, (1994), "Empirical Likelihood and General Estimating Equations", *The Annals of Statistics*, 22, 1, 300-325.

Reiersol, O. (1941), "Confluence Analysis of Lag Moments and Other Methods of Confluence Analysis", *Econometrica*, 9, 1-24.

Reiersol, O. (1945), "Confluence Analysis by Means of Instrumental Sets of Variables", Arkiv for Mathematik, Astronomie och Fysik, 32.

Sargan, J.D. (1958), "The Estimation of Economic Relationship using Instrumental Variables", *Econometrica*, 26, 393-415.

Theil, H. (1953), "Repeated Least Squares Applied to Complete Equations System", The Hague : Central Planning Bureau (mimeo).

Tikhonov, A., and V. Arsenin, (1977), *Solutions of Ill-posed Problems*, Winston & Sons, Washington D.C.

Tricomi, F.G., (1985), *Integral Equations*, Dover, New York.

Van der Vaart, A.W. and J.A., Wellner, (1996), *Weak Convergence and Empirical Processes*, Springer, New York.

Wahba, G., (1973), "Convergence Rates of Certain Approximate Solutions of Fredholm Integral of the First Kind", *Journal of Approximation Theory*, 7, 167-185.

Wahba, G., (1990), *Spline Models for Observational Data*, SIAM, Philadelphia.

Vanhems, A. (2000), "Nonparametric Solutions to Random Ordinary Differential Equations of First Orders", GREMAQ-University of Toulouse.