Endogeneity in non separable models. Application to treatment models where the outcomes are durations

J.P. Florens *

First Draft: December 2004
This version: January 2005
Preliminary and Incomplete Version

Abstract

This paper considers non separable models with endogenous variables. An example is given by treatment models where the outcomes are durations which lead to a model : $\tau = \Phi^Z(u)$ where τ is a duration and U is not independent from Z but is assumed to be independent from some instruments W. This independence condition may be expressed by a non linear integral equation from which local identification and estimation are derived. As in the separable case IV non parametric estimation is an ill posed inverse problem requiring a regularization and this paper is concentrated on Tikhonov regularization.

Keywords and phrases. non parametric instrumental estimation, Urysohn integral equation, Tikhonov regularization.
JEL classifications.

1 Introduction

This paper analyses non separable models containing endogenous explanatory variables. A general form of this family of model is :

$$Y = m(Z, U)$$

where Y and Z are observed as random elements and U is non observable. Usual "regression type" non separable models assume Z and U independent and U uniformly generated (see Maskin ( ), Horowitz ( ), or Chesher ( )). In that case the so called "non separable models" may be viewed as a way to represent conditional probability of Y given Z. Actually this representation is extremely similar to computer simulations of conditional probability measures. Relevant questions in that framework are identification and estimation of some functional transformations of m. Our objective is to consider the same questions in a case where independence between U and Z is replaced by the independance between U and W, a set of instrumental variables. Several recent papers have considered non parametric instrumental variables in additive models (see Florens ( ), Newey and Powell ( ), Darolles, Florens and Renault ( ), Carrasco, Florens and Renault ( ) and Hall and Horowitz ( )) under an assumption of mean independence. They consider models : $Y = \varphi(Z) + U$ under $E(U|W) = 0$.

*University of Toulouse and IUF (GREMAQ)

The estimation of the φ function becomes, in that case, a particular example of an inverse problem defined by an equation : $A(\varphi, F) = 0$, where φ is the parameter of interest and F the distribution of the observed variables. Equivalently φ is defined as the solution of an equation parameterized by F and estimator is usually obtained by replacing F by a suitable estimator. Additive models lead to linear integral equations (Fredholm type I equations more precisely) which are known to be ill posed inverse problems characterized in our framework by a non continuity of the solution wrt F. Non separable models require more than a mean independence condition and we will only consider full independence between the error U in the equation and the instruments. This independence property generates a non linear integral equation : from the distribution of Y and Z given W we construct the distribution of U given W assumed to be equal to the marginal known distribution of U. We develop, in this paper, a local analysis of non linear inverse problem based on implicit function theorem in functional spaces under Frechet differentiability conditions (see for a similar approach applied to auction models or game theoretic models, Florens, Protopopescu and Richard ( )). The non linear inverse problems is then locally replaced by a linear one used in particular for local identification conditions and for local definition of regularity spaces required for the study of speed of convergence. One of the main applications of non separable models is provided by duration models where the m function becomes the inverse of the integrated hazard function. In this type of model the marginal distribution of U is naturally an exponential probability measure. In this example, the monotony assumption is naturally satisfied. Duration models with endogenous cofactors are obviously motivated by treatment models where the outcome is durations. Then, all the paper will be concentrated to this type of models but the methodology and the results are obviously applicable to general non separable models.

2 Model and Identification

Let us denote by ζ (continuous or discrete) the level of a treatment and by $(\tau^\zeta)_\zeta$ the counterfactual process of outcomes. For each $\zeta \in \Theta \subset \mathbb{R}$ τ^ζ is a duration, i.e. a positive random variable. Actually this process is degenerate in the sense that the τ^ζ are all functions of a single random element. We assume :

$$\tau^\zeta = (\Lambda^\zeta)^{-1}(U) = \Phi^\zeta(U) \tag{2.1}$$

where U is an exponential (1) distribution. For each ζ, Λ is the integrated hazard function of τ^ζ and is assumed to be strictly increasing from $\mathbb{R}^+$ to $\mathbb{R}^+$. We denote by Φ^ζ its inverse function and, assuming smoothness conditions,

$$\lambda^\zeta(t) = \frac{\partial}{\partial t}\Lambda^\zeta(t) \tag{2.2}$$

is the hazard rate of τ^ζ. The parameter of interest of this model is functional and equal to $\Lambda^\zeta(t)$(or $\lambda^\zeta(t)$) or to its inverse $\Phi^\zeta(u)$. This counterfactual process is completed by an assignment mechanism. We consider a joint distribution of a vector $((\tau^\zeta)_\zeta, Z, W)$ where W is a vector of instrumental variables and Z is the level of treatment. We assume:

H1 : $(\tau^\zeta)_\zeta$ has a marginal distribution characterized by (1)

H2 : the instruments are not time dependent. (This assumption is introduced in order to simplify the presentation but is not essential for your arguments.)

The observed data are (τ, Z, W) where $\tau = \tau^z$ i.e. the value of outcome at the assigned level of treatment $Z = z$. As in most of the treatment models, the dependence between τ and Z come from two factors: the treatment effect described by Λ^z and the assignment bias captured by the dependence between U and Z. Without any supplementary assumption, these two effects may not be separated and the parameter of interest $\Lambda^\varsigma(t)$ is not identified. We may consider three independence or conditional independence conditions which determine identification.

A1 : $Z \perp U$

A2 : $Z \perp U|W$

A3 : $U \perp W$

The main interesting case is the last one but we first briefly discuss the first two conditions.

- The condition A1 defines the pure randomization case. In this case W may be neglected and the conditional integrated hazard function of τ given Z is precisely $\Lambda^z(t)$. This parameter is then identified and may be estimated by usual methods (see....)

- Condition A2 is a conditional randomization condition from which the identification follows from the following argument. Let us denote by $S(t|z, w)$ the conditional survivor function of τ, i.e.

$$\begin{aligned} S(t|z,w) &= P(\tau \geq t|Z = z, W = w) \\ &= P(U \geq \Lambda^z(t)|Z = z, W = w) \\ &= S_U(\Lambda^z(t)|W = w) \end{aligned} \tag{2.3}$$

  where $S_U(u|W = w) = P(U \geq u|W = w) = P(U \geq u|Z = z, W = w)$ from A2. Moreover :

$$P(U \geq u|W = w) = \frac{P(U \geq u)p(w|U \geq u)}{p(w)} \tag{2.4}$$

  where p represents both marginal and conditional density of W. Then:

$$\begin{aligned} S(t|z,w) &= \frac{e^{-\Lambda^z(t)}}{S(t)} \frac{S(t)p(w|z \leq t)}{p(w)} \\ &= \frac{e^{-\Lambda^z(t)}S(t|w)}{S(t)} \end{aligned} \tag{2.5}$$

  using obviously S for the marginal and the conditional survivor functions of τ. Then $\Lambda^z(t)$ is identified.

We are essentially interested in the case 3. Condition A3 may be interpreted as an instrumental variables condition in a strong sense because we assume an independence condition between U and W and not only a mean independence. Let F be the distribution function of the r.v. (τ, Z, W) and denote by $f(t, z, w)$ the density wrt. the Lebesgue measure. We will write, for example, $f(., z, w)$ for the marginal distribution of the r.v. (Z, W) wrt. F. Define the conditional survivor function:

$$S(t, z|w) = \int_t^\infty \frac{f(\tilde{t}, z, w)}{f(., ., w)} d\tilde{t}. \tag{2.6}$$

Note that $S(t, z|w)$ is the joint probability of $\tau \geq t$ and $Z = z$ given $W = w$ satisfying

$$P(\tau \geq t, Z \geq z|W = w) = -\int_z^\infty S(t, \tilde{z}|w)d\tilde{z}. \quad (2.7)$$

This function is identified and estimable by the observed data. From this distribution, let us deduce the joint distribution of U and Z :

$$P(U \geq u, z|w) = S(\Phi^z(u), z|w) \quad (2.8)$$

The A3 independence and $U \sim Exp(1)$ implies :

$$\int S(\Phi^z(u), z|w)dz = e^{-u} \quad (2.9)$$

In this non linear integral equation (Urysohn type I equation) S may be estimated and (2.9) needs to be solved in Φ. Identification of Λ^z in that case equivalent to unicity of the solution of (2.9). As the model is non linear, global identification is not warranted and we look for local identification based on implicit function theorem.

In the following we give sufficient conditions for a local identification of the parameter $\Phi^z(u)$ considered in a fixed point u. Then the parameter $\phi(z) = \Phi^z(u)$ can be seen as function only of z. Locally identification of Λ^z is than equivalent to unicity of the solution of following integral equation:

$$[T\phi](W) := \int S(\phi(z), z|W)dz = c, \quad (2.10)$$

where c denotes a positive constant.

Assumption A2.1 *Suppose that the joint density $f(t, s, w)$ satisfies:*
(a) The marginal density of Z and W satisfies a Hilbert-Schmidt condition, ie.

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

(b) $\int \sup_z \left|\frac{f(0,z,w)}{f(.,.,w)}\right| dw < \infty$

(c) The joint density $f(t, s, w)$ is continuous and continuously differentiable with respect to t for all z, w and that $\int \sup_{t,z} \left|\frac{f'(t,z,w)}{f(.,.,w)}\right| dw < \infty$, where $f'(t, z, w)$ denotes the partial derivative of f wrt. the first component.

Proposition 2.1 *Let A2.1 be statisfied. Then, the operation T maps the space L^2_Z into the space L^2_W, is Fréchet differentiable at every point $\phi_0 \in L^2_Z$, where the Fréchet derivative K_{ϕ_0} in ϕ_0 is given by*

$$[K_{\phi_0}(\phi)](w) = \int \phi(z)\frac{f(\phi_0(z), z, w)}{f(., ., w)}dz. \quad (2.11)$$

Note that K_{ϕ_0} is a linear continuous integral operator transforming functions of z into functions of w. The local unicity of the solution of (2.9) is warranted if K_{ϕ_0} is a one to one linear operator. A natural condition is the following :

Assumption CSI *The random element (τ, Z) is strongly identified by (τ, W) conditionally on τ.*

The concept of strong identification extends the statistical definition of "boundely complete" statistics (see Florens, Mouchart, and Rolin 1990) and has been previously used as a sufficient condition of identification of instrumental variables regression. In this context, an extension of this concept to conditional strong identification is required in order to formalize the time dependence in equation (10). Assumption CSI means that :

$$\forall h(\tau, z) \quad E(h(\tau, z)|\tau, w) = 0 \Rightarrow h = 0, \tag{2.12}$$

where all the equalities between random variables are almost sure equalities *w.r.t.* the DGP. Intuitively this assumption means that Z and W are sufficiently dependent for any τ.
Let us now complete the argument and consider the equation $K_{\phi_0}\phi = 0$. Equivalently :

$$\int \Phi^z(\wedge^z(t))f(z|t, w)dz = 0 \tag{2.13}$$

which implies $\Phi^z = 0$ under the CSI assumption. We summarize this analysis by the following theorem :

Theorem 2.2 *Let a model with counterfactual: $((\tau^z), Z, W))$ where $\tau^z = \Phi^z(U)$ and $U \sim Exp(1)$. If:*

1. *$U \perp W$*

2. *Z is strongly identified by W conditionally on $\tau = \tau^z$*

Then : Φ^z is locally identified by the observed model which generated $\tau = \tau^z$ and Z given W.

3 Estimation

The estimation of $\Phi^z(u)$ is based on the resolution of equation (2.9) where S is replaced by a non parametric estimation. Actually resolution of integral equations of type I leads to ill posed inverse problems requiring regularization methods.

As a first simplication, we consider the problem with u fixed. Then the parameter $\phi(z) = \Phi^z(u)$ can be seen as function only of z. Obviously the analysis may be conducted for many values of u. Elements of the problems are now functions of Z and of W only. The set of functions of Z is the set of square integrable function w.r.t. the true distribution on Z and the set of functions of W is also L^2 relative to the true distribution. We keep an intuitive context to our presentation by not explicitly assuming all the integrability conditions required by our computations. Let us recall the integral operation of interest:

$$[T(\phi)](w) = \int S(\phi(z), z|w)dz. \tag{3.14}$$

T is a non linear operator transforming functions of Z into functions of W.

Solving the operator equation $T(\phi_0) = e^{-u}$ is ill-posed, i.e. the solution ϕ_0 does not depend continuously from the rhs. Different regularization methods may be used to obtain a stable problem. We first propose a Thikonov regularization, a (simplified) Newton Kantorowitch (NK) method and a Landweber-Fridman iteration.

3.1 Thikonov regularization

Let us first consider general considerations about resolution of equation (2.9) using Tikhonow regularization. A regularized solution can be defined as a minimizer of :

$$\min_{\phi} \|T(\phi) - e^{-u}\|^2 + \alpha\|\phi\|^2 \tag{3.15}$$

The first order condition derived from (3.15) is :

$$\langle T(\phi) - e^{-u}, K_{\phi_0}\tilde{\phi}\rangle = \alpha\langle \phi(z), \tilde{\phi}(z)\rangle \tag{3.16}$$

for any $\tilde{\phi}$, where $K_{\phi_0} = \int \tilde{\phi}(z) f(\phi_0(z), z|w)dz$ denotes the Fréchet derivative of T in ϕ_0. The adjoint operator of K_{ϕ_0}, relative to the considered Hilbert spaces, is :

$$K^*_{\phi_0}\tilde{\psi} = \int \tilde{\psi}(w)\frac{f(\phi_0(z), z, w)}{f(., z, .)}dw. \tag{3.17}$$

Then condition (3.16) is equivalent to :

$$< K^*_{\phi_0}T(\phi) - K^*_{\phi_0}e^u, \phi >= \alpha < \phi, \tilde{\phi} > \tag{3.18}$$

for any $\tilde{\phi}$. Then we consider a solution ϕ_α of

$$\alpha\phi - K^*_{\phi_0}T(\phi) = K^*_{\phi_0}e^{-u} \tag{3.19}$$

or equivalently

$$\alpha\phi(z) - \int\left\{\int S(\phi(\tilde{z}), \tilde{z}|w)d\tilde{z}\right\}\frac{f(\phi(z), z, w)}{f(., z, .)}dw = e^{-u}\frac{f(\phi(z), z, .)}{f(., z, .)} \tag{3.20}$$

as regularized solution. The estimation of ϕ_α follows from (3.20). The two functions S and f may be replaced by their non parametric estimators $\hat{S}$ and $\hat{f}$ (using for example kernel smoothing estimators) and the non linear integral equation may be solved.

3.2 Newton Kantorowitch method

We propose a (simplified) Newton Kantorowitch (NK) method obtained by the following argument. We have two possible strategies :

I the first one consists in applying NK algorithm to (2.13). This method leads to solve, at each step, a linear Fredholm equation of the first kind, which is an ill posed problem and then which must be solved using, e.g. a Tikhonov regularization.

II The second strategy consists in starting from (3.20) and to use NK algorithm which will give, at each step, a well posed linear problem.

Actually, these two methods are not equivalent and for simplicity, we only present the second one. This second method may be viewed as a simplification of the first one where one part of a derivative is not introduced. (See appendix for this computation). Let us start with (2.13) written again $T(\phi) = e^{-u}$. Let ϕ_{m-1} the solution obtained at step $m-1$. We consider :

$$T(\phi) - e^{-u} \simeq T(\phi_{m-1}) - e^{-u} - K_{\phi_{m-1}}(\phi - \phi_{m-1}) \tag{3.21}$$

where $K_{\phi_{m-1}}$ is the Fréchet derivative of T in ϕ_{m-1}. Then ϕ_m is obtained as solution of :

$$K_{\phi_{m-1}}(\phi - \phi_{m-1}) = T(\phi_{m-1}) - e^{-u}. \tag{3.22}$$

Let us now introduce a Tikhonov regularization and ϕ^{α}_{m-1} the regularized approximation of ϕ at step $m-1$. Then ϕ^{α}_{m} will be defined as the regularized solution of

$$K_{\phi_{m-1}}\phi = K_{\phi_{m-1}}\phi^{\alpha}_{m-1} + T(\phi^{\alpha}_{m-1}) - e^{-u} \tag{3.23}$$

Then :

$$\phi^{\alpha}_{m} = (\alpha I + K^{*}_{\phi_{m-1}}K_{\phi_{m-1}})^{-1}K^{*}_{\phi_{m-1}}(K_{\phi_{m-1}}\phi^{\alpha}_{m-1} + T(\phi^{\alpha}_{m-1}) - e^{-u}). \tag{3.24}$$

3.3 Landweber-Fridman iteration

Let ϕ_{m-1} the solution obtained at step $m-1$.Then ϕ_m is given by:

$$\phi_m := \phi_{m-1} - cK^{*}_{\phi_{m-1}}[T(\phi_{m-1}) - e^{-u}]. \tag{3.25}$$

Example: Discrete treatment

We consider as an example the case where $\zeta \in \{0, 1\}$. In that case no regularization is needed because the set of functions of ζ is finite dimensional. The equation (18) becomes the set of ywo non linear equations in Φ^0 and Φ^1:

$$\int [S(\Phi^0(u), 0|w) + S(\Phi^0(u), 1|w)] + f(\Phi^0, 0, w)dw = e^{-u}f(\Phi^0(u), 0)$$

$$\int [S(\Phi^0(u), 0|w) + S(\Phi^0(u), 1|w)] + f(\Phi^1(u), 1, w)dw = e^{-u}f(\Phi^1(u), 1)$$

4 Statistical bias in the resolution of an Urysohn type I equation

The objective of this section is to compute a bound for the speed of convergence of the difference between $\hat{\Phi}$ the regularized estimator solution of a non linear integral equation and Φ^{α} the regularized solution of the true equation. We just present a sketch of the computation and the details are left to the reader. Let us first recall the notations :

- We have defined the $\Phi^z(u)$ function in the first section as the inverse of the integrated hazard function in the counterfactual model. As all our analysis is conducted for a given u, we denote by $\varphi(z)$ the function $\Phi^z(u)$. This function is an element of $L^2(z)$ and the equation to be solved is $T(\varphi) = r$ where in that case r is a known constant equal to e^{-u}. The operator T is valued into e^{-u}.
- We have previously defined two compact operators K_{φ} and K^{*}_{φ}, respectively equal to the derivative of T and to its adjoint.
- The operator T is unknown and $\hat{T}$ denotes an estimator of T to which $\hat{K}_{\varphi}$ and $\hat{K}^{*}_{\varphi}$ are associated. Actually $\hat{T}$, $\hat{K}_{\varphi}$ and $\hat{K}^{*}_{\varphi}$ are defined by :

$$\hat{T}(\varphi) = \int \hat{S}(\varphi(z), z|w))dz \tag{4.26}$$

$$\hat{K}_\varphi(\hat{\varphi}) = -\int \hat{\varphi}(z)\hat{f}(\varphi(z), z|w)dz \tag{4.27}$$

$$\hat{K}^*_\varphi(\tilde{\psi}) = -\int \tilde{\psi}(w)\frac{\hat{f}(\varphi(z), z, w)}{f(\hat{z})}dw \tag{4.28}$$

Note that $\hat{K}^*_\varphi\tilde{\psi}$ is not the adjoint of $\hat{K}_\varphi$ in $L^2(Z)$.
As we have seen before the estimator $\hat{\varphi}$ verified :

$$\alpha\hat{\varphi} + \hat{K}^*_\varphi\hat{T}(\hat{\varphi}) = \hat{K}^*_{\hat{\varphi}}r \tag{4.29}$$

and φ^α is defined :

$$\alpha\varphi^\alpha + K^*_{\varphi^\alpha}T(\varphi^\alpha) = K^*_{\varphi^\alpha} \tag{4.30}$$

Then :

$$\alpha\left(\hat{\varphi} - \varphi^\alpha\right) + \hat{K}^*_{\hat{\varphi}}\hat{T}(\hat{\varphi}) - K^\alpha_{\varphi^\alpha}T(\varphi^\alpha) = \hat{K}^*\hat{\varphi}r - K^*\varphi^\alpha r \tag{4.31}$$

and

$$\begin{aligned} \alpha\left(\hat{\varphi} - \varphi^\alpha\right) &+ \hat{K}^*_{\hat{\varphi}}\hat{T}(\hat{\varphi}) - \hat{K}^*_{\hat{\varphi}}\hat{T}(\varphi^\alpha) \\ &= -(\hat{K}^*_{\hat{\varphi}}\hat{T}(\varphi^\alpha) - \hat{K}^*_{\varphi^\alpha}T(\varphi^\alpha)) + \hat{K}^*_{\varphi^\alpha r} - K^*_{\varphi^\alpha r} \end{aligned} \tag{4.32}$$

First we consider the approximation of :

$$\hat{K}^*_{\hat{\varphi}}\hat{T}(\hat{\varphi}) - \hat{K}^*\hat{\varphi}\hat{T}(\varphi^\alpha) \;\; by \;\; \hat{K}^*_{\hat{\varphi}}\hat{K}^*_{\hat{\varphi}}\left(\hat{\varphi} - \varphi^\alpha\right) \tag{4.33}$$

The r.h.s of (31) then becomes :

$$\left(\alpha I + \hat{K}^*_{\hat{\varphi}}\hat{K}_{\hat{\varphi}})\right)\left(\hat{\varphi} - \varphi^\alpha\right) \tag{4.34}$$

and its norm is smaller than $||\hat{\varphi} - \varphi^\alpha||$. Indeed

$$\begin{aligned} &||(\alpha I + \hat{K}^*_{\hat{\varphi}}, \hat{K}_\varphi)\ (\hat{\varphi} - \varphi^\alpha)||^2 = \\ \alpha^2 < \hat{\varphi} - \varphi^\alpha, \hat{\varphi} - \varphi^\alpha > &+2\ \alpha < \hat{\varphi} - \varphi^\alpha, \hat{K}^*_{\hat{\varphi}}\hat{K}_{\hat{\varphi}}\left(\hat{\varphi} - \varphi^\alpha\right) > \\ &< \hat{K}^*_{\hat{\varphi}}\hat{K}_{\hat{\varphi}}\left(\hat{\varphi} - \varphi^\alpha\right), \hat{K}^*_{\hat{\varphi}}\hat{K}_{\hat{\varphi}}\left(\hat{\varphi} - \varphi^\alpha\right) > \end{aligned} \tag{4.35}$$

The third term is obviously positive and the second is equal to

$$2\alpha < \hat{K}_{\hat{\varphi}}\left(\hat{\varphi} - \varphi^\alpha\right), \hat{K}_{\hat{\varphi}}\left(\hat{\varphi} - \varphi^\alpha\right) > \text{and is then also positive.} \tag{4.36}$$

i)The r.h.s. of (31) may be re written :

$$
\begin{aligned}
&- [\hat{K}^*_{\hat{\varphi}}\hat{T}(\varphi^\alpha) - \hat{K}^*_{\varphi^\alpha}T(\varphi^\alpha) - \hat{K}_{\hat{\varphi}}\hat{T}(\varphi) + K^*_{\varphi^\alpha}T(\varphi)] \\
&\quad + \hat{K}^*_{\hat{\varphi}}T(\varphi) - K^*_{\varphi^\alpha}\ e^{-u} - \hat{K}^*_{\hat{\varphi}}\hat{T}(\varphi) + K^*_{\varphi^\alpha}\hat{T}(\varphi)
\end{aligned}
\tag{4.37}
$$

After simplification, this expression may be written as $-(A+B)$ where :

$$
A = \hat{K}^*_{\hat{\varphi}}\left(\hat{T}(\varphi^\alpha) - \hat{T}(\varphi)\right) - K^*_{\hat{\varphi}}\left(T(\varphi^\alpha) - T(\varphi)\right) \tag{4.38}
$$

We use Taylor approximations of $\hat{T}$ and T and we replace $K^*_{\hat{\varphi}}$, $K_{\hat{\varphi}}$, $K^\alpha_{\varphi^\alpha}$ by K^*_φ and K_φ. Then A is approximated by :

$$
(\hat{K}^*_\varphi\hat{K}_\varphi - K^*_\varphi K^*_\varphi)(\varphi^\alpha - \varphi) \tag{4.39}
$$

Finally :

$$
B = \hat{K}^*_{\hat{\varphi}}\left(\hat{T}(\varphi) - T(\varphi)\right) \tag{4.40}
$$

which has the same behavior as :

$$
K^*_\varphi\left(\hat{T}(\varphi) - T(\varphi)\right) \tag{4.41}
$$

ii) using the previous computations we get :
<u>Theorem : 5.1</u>

$$
||\hat{\varphi} - \varphi^\alpha|| \leq \frac{1}{\alpha}\Big\{||\hat{K}^*_\varphi\hat{K}^*_\varphi - K^*_\varphi||||\varphi^\alpha - \varphi|| + ||K^*_\varphi(\hat{T}(\varphi) - T(\varphi))||\Big\} \tag{4.42}
$$

We are now facing a problem which is extremely similar to the linear analysis of IV estimation. In our case, under suitable Kernel smoothing estimation we have :

$$
\begin{aligned}
||\hat{K}^*_\varphi\hat{K}_\varphi - K^*_\varphi K_\varphi||^2 &= 0_p\left(\frac{1}{nh_n^p} + h_n^{2\rho}\right) \\
||\hat{K}^*_\varphi(\hat{T}(\varphi) - T(\varphi)||^2 &= 0_p\left(\frac{1}{n} + h_n^{2\rho}\right)
\end{aligned}
\tag{4.43}
$$

where n is the sample size, h_n the bandwith, p the dimension of the vector z and ρ a smoothness index. Then :
<u>Theorem 5.2</u>

$$
||\hat{\varphi} - \varphi^\alpha||^2 = 0_p\left[\left(\frac{1}{\alpha^2 nh_n^p}\frac{h_n^{2\rho}}{\alpha^2}\right)||\ \varphi^\alpha - \varphi||^2 + \frac{1}{\alpha^2 n} + \frac{h_n^{2\rho}}{\alpha^2}\right] \tag{4.44}
$$

If, moreover, $\varphi \in \Psi_\beta$ (see following section),

$$
||\hat{\varphi} - \varphi||^2 = 0\left(\frac{1}{\alpha^{2-\beta}nh_n^p} + \frac{h_n^{2\rho}}{\alpha} + \alpha^\beta\right) \tag{4.45}
$$

Then if $\frac{p}{2\rho} \leq \frac{\beta}{2+\beta}$, the optional choice of h_n is proportional to $n^{-\frac{1}{2\rho}}$ and α should be taken proportionally to $n^{-\frac{1}{2\beta}}$. Then, $||\hat{\varphi} - \varphi||^2 = 0\left(n^{-\frac{\beta}{\beta+2}}\right)$

As in the separable case, let us underline that this speed of convergence is only a bound even if it has been proved to be minmax (see Hall and Horowitz) ( )). However under particular assumptions, it has been shown that better rate of convergence may be reached (see Carrasco and Florens ( )).

5 Regularization bias

Let us denote by φ^α the solution of the equation :

$$\alpha\varphi + K_\varphi^* T(\varphi) = K^\alpha \varphi r \tag{5.46}$$

This equation is a non linear integral equation which may be solved by a NK algorithm. The objective of this paragraph is to study the speed of convergence of $||\varphi^\alpha - \varphi||^2$ when φ is the true value solution of $T(\varphi) = r$. Indeed :

$$\begin{aligned} \varphi^\alpha - \varphi &= \frac{1}{\alpha}\left[K_{\varphi^\alpha}^* r - K_{\varphi^\alpha}^* T(\varphi^\alpha)\right] - \varphi \\ &= \frac{1}{\alpha}\left[K_{\varphi^\alpha}^* \left(T(\varphi) - T(\varphi^\alpha)\right)\right] - \varphi \end{aligned} \tag{5.47}$$

Using a Taylor expression of T around φ we may replace $T(\varphi^\alpha) - T(\varphi)$ by $K_\varphi\left(\varphi^\alpha - \varphi\right)$ plus terms which may be neglected because they are of higher order in $||\varphi^\alpha - \varphi||$.

Then :

$$\varphi^\alpha - \varphi = -\frac{1}{\alpha}\left[K_{\varphi^\alpha}^* K_\varphi\left(\varphi^\alpha - \varphi\right)\right] - \varphi \tag{5.48}$$

The other step is to replace $K_{\varphi^\alpha}^*$ by K_φ^*. Using Taylor expansion this replacement only generates high order terms in $||\varphi^\alpha - \varphi||$. Then :

$$\left(I + \frac{1}{\alpha}K_\varphi^* K_\varphi\right)\left(\varphi^\alpha - \varphi\right) = -\varphi \tag{5.49}$$

and

$$\left(\varphi^\alpha - \varphi\right) = -\alpha\left(\alpha I + {K^*}_\varphi K_\varphi\right)^1 \varphi \tag{5.50}$$

We are now facing a standard question about regularization of linear operator. The operator K_φ is a compact operator with a singular decomposition φ_j, λ_j (see...) and the true value φ should be assumed to be in :

$$\Phi_\beta = \left\{\varphi / \sum_{j=0}^{\propto} \frac{< \varphi_1 \varphi_j >^2}{\lambda_{j^2}\beta} < \propto \right\} \tag{5.51}$$

for some $\beta \in]0, 2[$. For that case it has been shown that $||\hat{\varphi} - \varphi||^2 = 0\left(\alpha^\beta\right)$. The main specificity of the non linear case is that K_φ depends on φ. This point links in a more complex way the two function φ and the operator T in order to get an order of convergence of $||\hat{\varphi} - \varphi||^2$. In particular the order of regularity of φ is only locally defined and the strength or the weakness of the instruments characterized by the decline of the λ_j depend on φ.

6 Exact and approximate computation of the estimator

Let c be a univariate Nadaraya-Watson kernel and $\hat{C}$ its survivor function.

$$\bar{C}(t) = \int_t^\infty c(u)du \tag{6.52}$$

We use the same letter c for multivariate kernel and h_n represents the bandwith.
Let $(t_i, z_i, w_i)_{i=1,...,n}$ an i.i.d. sample of (τ, Z, W). The estimated joint diversity is equal to:

$$\hat{f}(t,z,w) = \frac{1}{n{h_n}^{1+p+q}} \sum_{i=1}^n c(\frac{t-t_i}{h_n})\ c(\frac{z-z_i}{h_n})\ c(\frac{w-w_i}{h_n}) \tag{6.53}$$

and for example:

$$\hat{S}(t,z|w) = \frac{1}{h_n^p} \frac{\sum_{i=1}^n \bar{C}(\frac{t-t_i}{h_n})c(\frac{z-z_i}{h_n})c(\frac{w-w_i}{h_n})}{\sum_{i=1}^n c(\frac{t-t_i}{h_n})} \tag{6.54}$$

The estimator $\hat{\varphi}$ is solution of the estimated first order conditions:

$$\alpha\hat{\varphi}(z) + \hat{K}^*_{\hat{\varphi}}\hat{T}(\hat{\varphi}) = e^{-u}\frac{\hat{f}(\hat{\varphi}(z), z)}{\hat{f}(z)} \tag{6.55}$$

or, equivalently:

$$\begin{aligned}
\alpha\hat{\varphi}(z) \quad - \quad & \left\{\int\left\{\int \frac{1}{h_n^p}\frac{\sum_j \bar{C}(\frac{\hat{\varphi}(\zeta)-t(j)}{h_n})\quad c(\frac{\zeta-z_j}{h_n})\quad c(\frac{w-w_j}{h_n})}{\sum_j c(\frac{w-w_j}{h_n})}\right\}d\zeta\right. \\
& \left.\frac{1}{{h_n}^{1+q}}\frac{\sum_i c\left(\frac{\hat{\varphi}(z)-ti}{h_n}\right) c\left(\frac{z-z_i}{h_n}\right) c\left(\frac{w-w_i}{h_n}\right)}{\sum_i c\left(\frac{z-z_i}{h_n}\right)}dw\right\} \\
= \quad & -e^{-u}\frac{1}{h_n}\frac{\sum_i c\left(\frac{\hat{\varphi}(z)-ti}{h_n}\right) c\left(\frac{z-z_i}{h_n}\right)}{\sum_i c\left(\frac{z-z_i}{h_n}\right)}
\end{aligned} \tag{6.56}$$

Let us denote:

$$w_{ij} = \frac{1}{h_n^q}\int \frac{c(\frac{w-w_i}{h_n})\quad c(\frac{w-w_j}{h_n})}{\sum_j c(\frac{w-w_j}{h_n})} \tag{6.57}$$

which can be approximated by: Let us denote:

$$w^*_{ij} = \frac{c(\frac{w_i-w_j}{h_n})}{\sum_j c(\frac{w_i-w_j}{h_n})} \tag{6.58}$$

and

$$a_j(\hat{\varphi}) = \frac{1}{h_n^p}\int \bar{C}(\frac{\varphi(\zeta)-t_j}{h_n})\quad c(\frac{\zeta-z_j}{h_n})d\zeta \tag{6.59}$$

which can be approximated by:

$$a_j(\hat{\varphi})^* = \bar{C}(\frac{\hat{\varphi}(z_j)-t_j}{h_n}) \tag{6.60}$$

The estimator $\hat{\varphi}$ is the solution of:

$$\alpha\hat{\varphi}(z) = \sum_{ij}[w_{ij} \quad a_j(\hat{\varphi}) - e^{-u}] \quad \frac{\frac{1}{h_n} \quad c(\frac{\hat{\varphi}(z)-t_i}{h_n}) \quad c(\frac{z-z_i}{h_n})}{\sum_i \ c(\frac{z-z_i}{h_n})} \tag{6.61}$$

and an approximation may be obtained by replacing w_{ij} and $\alpha\hat{\varphi}(z)$ by w_{ij}^* and $a_j(\hat{\varphi})^*$. If approximation (which does not affect the speed of convergence) is adopted equation () may be solved by a fixed point method. In a first step the $\hat{\varphi}(z_j)$ has to be computed and in a second step the term between brackets is known and the value of $\hat{\varphi}$ at any point z may be also derived from a fixed point recursion.

A Appendix - Proofs