Nonparametric study of solutions of differential equations

Anne Vanhems*
ESC Toulouse
20, bd Lascrosses
BP 7010 31068 Toulouse Cedex 7, France
e-mail: a.vanhems@esc-toulouse.fr

March 23, 2005

*I would like to thank Jean-Pierre Florens, Richard Blundell, Eric Renault, Christine Thomas-Agnan, for stimulating conversations, suggestions and advice. I am very grateful to Oliver Linton, and two anonymous referees for most helpful comments.

Nonparametric study of solutions of differential equations

Anne Vanhems

ESC Toulouse
20, bd Lascrosses
BP 7010 31068 Toulouse Cedex 7, France
e-mail: a.vanhems@esc-toulouse.fr

Abstract

The solution of differential equations lies at the heart of many problems in structural economics. In econometrics the general nonparametric analysis of consumer welfare is historically the most obvious application, but there are also many in finance and other fields. This work considers the general nonparametric form for these problems and identification conditions. It derives a kernel based estimator and shows consistency and asymptotic normality.

In particular, the link with inverse problems allows to define it in terms of well-posed inverse problem and to stress the regularity properties of the estimated solution.

Keywords: Inverse problem, nonparametric kernel estimation, differential equations

1 Introduction

Consider independent identically distributed observations which admit an unknown cumulative distribution function F. Structural econometrics considers implicit transformation of F. Of course, the theory developed depends on the nature and properties of the transformation considered. For example, there exists a wide literature about integral transformations, like additive models (Hastie and Tibshirani 1990), single index models (Ichimura 1987) (Horowitz 1998) or instrumental variables theory (Blundell and Powell 2000, Darolles, Florens, Renault 2000), (Newey, Powell, Vella 2000).

This paper is concerned with studying differential transformations of F, and by extension of the conditional expectation. Such a purpose can be justified first by the numerous applications in economics and econometrics. In econometrics the general nonparametric analysis of consumer welfare is historically the most obvious application, and there are also many in finance and other fields. Hausman and Newey (1995) studied the variation of exact consumer surplus associated to a price change by solving a differential equation. Ait-Sahalia (1996) studied the solution of a partial differential equation in order to define the price of an interest-rate derivative security. Many examples can also be quoted in physics (for example Florens and Vanhems 2001).

This work considers the general nonparametric form for these problems and identification conditions. It derives a kernel based estimator and shows consistency and asymptotic normality.

Basically, it deals with the study of the following system:.

$$\begin{cases} \lambda'(x) & = m(x, \lambda(x)) \\ \lambda(x_0) & = \lambda_0 \end{cases} \tag{1}$$

The function m is assumed to be continuous and is defined by $m(x, y) = I\!\!E\,(Z\,|X = x, Y = y)$ where (X, Y, Z) are 3 real-valued random variables. The parameters x_0 and λ_0 are given by the underlying economic background.

Let briefly detail an example of application in microeconomics. It is taken from an article by Hausman and Newey (1995). The objective is to measure the impact on the consumer welfare of a price change for one good. Therefore, consider one consumer; define y its income, q the demand in good and p^1 the price of a unique good. Assume that there exists a price variation from p^0 to p^1. A way to capture the impact on the consumer is to calculate the variation of exact consumer surplus λ: it represents the cost to pay to the consumer so that his welfare does not change for a price change. It is a monetary measure of the variation of utility.

The first objective is to find a relation that links the functions λ and q. For that purpose, consider a price path $p(t), t \in [t_0, t_1]$ where $p(t_0) = p^0$ and $p(t_1) = p^1$ and $\lambda(p(t))$ is the variation of exact consumer surplus between $p(t)$ and $p(t_1)$. Then, it is possible to derive the following relation between the interest parameter λ and the demand function q:

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

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

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

This is clearly a particular case of differential equation of order one. Compared to the general equation (1), the random vector (X, Y, Z) represents the price of a good (X), the income of an individual (Y) and the quantity consumed for the associated good (Z). The function m is defined as the observed demand function and the solution λ measures the variation of consumer surplus associated to the variation of prices from x_0 to x. A similar problem could also be studied for the variation of firm profit and the variation of demand factor.

The solution λ is then defined implicitely through the previous equation (1) and in order to study its properties, it will be necessary to explicit the link between m and λ. This kind of problems belongs to the general class of inverse problems and this paper will in particular stress the properties of well-posed inverse problem of equation (1).

Moreover, if a solution λ exists, it depends on the unknown distribution function of (X, Y, Z), F and it is necessary to introduce some nonparametric method to estimate it. Let $\widehat{m}_n(x, y)$ be a nonparametric kernel estimator of the conditional expectation and $\widehat{\lambda}_n$ the associated estimated solution. Therefore, the main contribution of this work is to derive a detailed analysis of this estimated solution, with results on consistency, optimality (in comparison with Hausman and Newey's work), and asymptotic normality. In particular, the link with inverse problems allows to define it in terms of well-posed inverse problem and to stress the regularity properties of the estimated solution as a functional of nonparametric kernel estimators. Indeed, integrating a nonparametric kernel estimator will improve its asymptotic properties and this general result can also be verified on $\widehat{\lambda}_n$.

The paper is organized as follows. Section 2 sets up the main problem to solve and presents the identification and overidentification conditions. Section 3 gives the numerical aspect and implementation technics related to differential equations. Section 4 is devoted to asymptotic results (consistency, optimality and asymptotic normality). The proofs are collected in the last section.

2 Differential equations and nonparametric kernel estimators

This section provides a first description of the functions λ and $\widehat{\lambda}_n$ defined by (1), in particular the properties of identification and overidentification.

2.1 Presentation of the problem

Let $S = (X, Y, Z)$ be a $\mathbb{R}^3$-valued random vector. The data available consist of $s_1, ..., s_n$ realizations of $S_1, ..., S_n$ independent identically distributed as $f_S(., ., .)$, an unknown density function with associated cumulative distribution function $F_S(., ., .)$. Let assume that the third coordinate of S, Z, is square integrable that is $\mathbb{E}Z^2 < +\infty$.

Define $f(x, y) = \displaystyle\int f_S(x, y, z)dz$, $\varphi(x, y) = \displaystyle\int z f_S(x, y, z)dz$; and the conditional expectation: $m(x, y) = \mathbb{E}\left[Z \,|X = x, Y = y\right] = \frac{\varphi(x,y)}{f(x,y)}$ on $f(x, y) > 0$.

Then, the initial system to be studied is the following, where the function m is defined on a subset of $\mathbb{R}^2$ and is at least continuous:

$$\begin{cases} \lambda'(x) & = m(x, \lambda(x)) \\ \lambda(x_0) & = \lambda_0 \end{cases} \tag{2}$$

The first objective is to detail the properties of the solution (if it exists) of this system. Moreover, since it depends on an unknown parameter, it has to be estimated through a nonparametric method.

The estimator of the function $m(x, y)$ is calculated using nonparametric kernel estimation. Let K denote a kernel function of one dimension, centered and separable, and σ_n a bandwidth parameter. Consequently, the kernel estimator of m is defined on $\hat{f}_n(x, y) > 0$ as

$$\hat{m}_n(x, y) = \frac{\overset{\wedge}{\varphi}_n(x, y)}{\hat{f}_n(x, y)} = \frac{\sum\limits_{i=1}^{n} z_i K\left(\frac{x - x_i}{\sigma_n}\right) K\left(\frac{y - y_i}{\sigma_n}\right)}{\sum\limits_{i=1}^{n} K\left(\frac{x - x_i}{\sigma_n}\right) K\left(\frac{y - y_i}{\sigma_n}\right)}$$

where $\hat{f}_n(x, y)$ and $\overset{\wedge}{\varphi}_n(x, y)$ are the respective estimators of $f(x, y)$ and $\varphi(x, y)$.

Remark 2.1 : *This expression could have been written with a kernel of dimension* 2*, which corresponds to the* 2*-dimensional random vector* (X, Y)*. Nevertheless, in order to simplify the calculus and especially the proofs, we assume that the kernel function is centered and separable, in order to decompose it into* 2 *independent kernel functions of dimension* 1 *(See Wand and Jones 1995).*

Remark 2.2 : *Moreover, a unique bandwidth parameter* σ_n *is considered for all the variables. A more general definition could have been used with two different smoothing parameters* σ_n^X, σ_n^Y*, which are approximately equal up to a multiplicative constant term. For sake of simplicity, it is assumed that* $\sigma_n^X = \sigma_n^Y = \sigma_n$

Therefore, the associated estimated system (3), where $\widehat{m}_n(x, y)$ is the kernel estimator of $m(x, y) = \mathbb{E}\left(Z \,|X = x, Y = y\right)$, is defined by:

$$\begin{cases} \lambda'(x) & = \widehat{m}_n(x, \lambda(x)) \\ \lambda(x_0) & = \lambda_0 \end{cases} \tag{3}$$

Before studying the convergence of two possible solutions, let check whether or not there exists a unique solution for both systems and also if the solutions are stable under small perturbations of initial conditions. In other words, the objective of the next part is to study the well-posedness of a particular inverse problem.

2.2 Link with inverse problems theory

Denote by A an operator which depends on two arguments, m and λ, such that:

$$A\left(m,\lambda\right)=\lambda'-m\left(.,\lambda\right)$$

Therefore, solving equation (2) is equivalent to solving:

$$\left\{\begin{array}{ll} A\left(m,\lambda\right) & =0 \\ \lambda(x_0) & =\lambda_0 \end{array}\right. \tag{4}$$

The objective is then to express the solution λ as a function of m, that is: $\lambda=\Phi(m)$. Such a kind of problem obviously belongs to the inverse problem theory (Florens 2000). The properties of the solution λ will in particular depend on the regularity of the initial operator A.

A classical class of inverse problems deals with integral operators. Inversion of some integral operators has already been studied, like Fredholm integral (Groetsch 1984), Abel type equation (Garza, Hall and Ruymgaart 1999). These kind of operators are usually linear and compact on Hilbert spaces ($\mathbb{L}^2$) (Van Rooij and Ruymgaart 1999). The differences between these examples and (4) are mainly twofold. First, without imposing restrictions on the form of the function m, the relation (4) defines a nonlinear equation to solve. Therefore, one important step of this analysis will be to try to linearize this relation in order to study its properties. Further, the operator to invert is a differential one, and then, contrary to integral operators, has not the good regularity properties expected for the previous ones. For example, A will not be continuous with respect to the uniform norm, but for some Sobolev norm (detailed later).

Let now study the well-posedness of our problem: if it admits a unique solution stable under perturbation of the function m (when m is replaced by a kernel estimator $\widehat{m}_n$).

2.3 Identification and overidentification

The study of identification and overidentification is needed for both systems (2) and (3).

2.3.1 Study of the true differential equation

Consider first the initial system (2). The proof for existence and uniqueness of a solution λ is given using a well-known analysis result on ordinary differential equations: the Cauchy-Lipschitz theorem (Schwartz 1970).

Let $I = [x_0 - a, x_0 + a]$ for $a > 0$ and $C(I)$ be the Banach space of continuous functions $x(t)$ on I with the sup norm $\|x\|_\infty = \max_{t\in I} |x(t)|$.

In what follows, without loss of generality, assume that $x_0 = \lambda_0 = 0$, so that $I = [-a, a]$.

Define $D = \{(x, y)\,; |x| \leq a, |y| \leq b\}$ and

$C_{b,0}(I) = \{u \in C(I); u(0) = 0, \|u\|_\infty \leq b\}$. The following assumptions are made:

$[a]$: $\max_{(x,y)\in D} |m(x, y)| < b/a$

$[b]$: $|m(x, y_1) - m(x, y_2)| \leq k\,|y_1 - y_2|$, $\forall\,(x, y_i) \in D$ such that $c = ka < 1$.

Lemma 2.1 : *Under assumptions* $[a]$ *and* $[b]$, *there exists a unique solution to* (2), $\lambda \in C_{b,0}(I)$

This result comes from the Cauchy-Lipschitz theorem; the mathematical points are detailed in the mathematical appendix.

Both assumptions $[a]$ and $[b]$ lead to some comments on the regularity imposed on m and f that are detailed below.

Remark 2.3 : *First, the important condition for proving existence and uniqueness of a solution is assumption* $[b]$. *Indeed, assumption* $[a]$ *is just imposed by the local definition of our solution on* D. *Moreover, it is not very restrictive since the parameter* a *will usually be chosen small enough; in particular, if the function* m *is continuous, this assumption can be satisfied for a well-chosen* a.

Remark 2.4 : *Assumption* $[b]$ *imposes* m *to be continuous on* D *and to satisfy the Lipschitz condition uniformly on the second argument; in particular a sufficient condition on* m *to verify this assumption is to be continuously differentiable of order* 1 *on* D *(or that* m *is continuously differentiable of order* 2 *on* D, *as needed for the mean square convergence).*

Remark 2.5 : *Thanks to Cauchy-Lipschitz theorem, a unique solution* λ *can be defined inside a compact neighborhood of the initial condition* $(0, 0)$. *It could be interesting to find the optimal interval* I *and a maximal solution defined on this optimal interval, or in other words, to check if the solution can be extended to a larger interval. Theoretically, under the same assumptions, it can be proved that it exists a unique maximal solution. That means concretely to piece together local solutions if the intersection of their definition intervals is not empty.*

Remark 2.6 : *Note that these assumptions* $[a]$ *and* $[b]$ *are in particular satisfied under standard regularity assumptions on* f, *like continuity and continuous derivability, and it is not necessary to restrict the definition domain of* f *to compact support (contrary to standard assumptions in nonparametric estimation), or to worry about kernel estimation in the boundary of finitely supported random variables. In other words, in principle, there is no need here for imposing compact supports on the variables* X, Y, Z *or for bounding* f *from below. The restriction to a compact neighborhood around*

$(0,0)$ *will be imposed in a second step when defining a unique solution to the differential equation. Then, construct the conditional expectation* $m(x,y)$ *(and its associated estimator* $\widehat{m}_n(x,y)$*) on every point where* $f(x,y) > 0$ *and just restrict it to a compact support* D *only to define the solution of the differential equation.*

However, in order not to rule out simple cases like uniform random variables, we don't define φ *and* f *on* $I\!R^2$ *but consider that any compact subset* Θ *of* $I\!R^2$ *containing* D *is a good condidate for the definition domain of these functions.*

Remark 2.7 *: Moreover, to be sure that the function* m *is defined on* D*, we implicitely assume that* f *is strictly positive on* D*. By continuity, under the assumption that* $f(0,0) > 0$*, it will always be possible to define a neighborhood around* $(0,0)$ *where* m *is well defined.*

To summarize, here are the assumptions imposed on φ and f:

$[A_1(s_1)]$ (i) φ and f are continuous bounded functions on a compact subset $\Theta \subset I\!R^2$, such that $D \subset int(\Theta)$ (interior of Θ), and continuously differentiable of order s_1 on D

(ii) $f(0,0) > 0$.

This assumption depends on one parameter s_1 which captures the regularity of the function f.

2.3.2 Study of the estimated differential equation

To calculate $\widehat{m}_n(x,y)$, introduce some nonparametric kernel estimator with a kernel function K. As mentioned before, K is a one dimension kernel for sake of simplicity. Here are the assumptions required for K:

$[A_2(r,s_2)]$ (i) K is a bounded and Lebesgue integrable function of $I\!R$ into $I\!R$. $\int_{-\infty}^{+\infty} K(x)dx = 1$.

(ii) K is of order $r \geq 2$, that is: $\forall 1 \leq p < r$, $\int_{-\infty}^{+\infty} x^p K(x)dx = 0$, $\int_{-\infty}^{+\infty} x^r K(x)dx \neq 0$ and $\int_{-\infty}^{+\infty} |x|^r |K(x)| dx < +\infty$, where $x \in I\!R$.

(iii) K is continuously differentiable up to order s_2 and its derivatives of order up to s_2 are in $L^2(I\!R)$.

This assumption depends on two parameters r and s_2, which correspond to the order and the regularity of the function K. When $r = 2$, K is of order 2, which is a common used assumption. Further restrictions may be imposed on K, as in the next section, with the notion of kernel of order 3 (which is of course no longer symmetric). Such an additional assumption may be useful when solving the inverse problem, in particular when the operator to inverse regularizes the initial estimator $\widehat{m}_n$.

This point will be explored later, let just remark that solving this differential equation allows to obtain a function λ whose regularity properties are increased compared to m. This smoothing property can obviously also be observed for the associated estimated functions $\widehat{\lambda}_n$ and $\widehat{m}_n$.

As noted in remark (2.4), assume that m is continuously differentiable of order 1 on D; moreover thanks to assumption $[A_2(r, s_2 = 1)]$, the estimator $\widehat{m}_n$ is also at least continuously differentiable of order 1 on D. Therefore, Cauchy-Lipschitz theorem proves that for all n there exists a unique local solution $\widehat{\lambda}_n$ to the system (3). In particular:

$$\forall n \geq 0, \exists k_n, \forall (x, y_i) \in D, |\widehat{m}_n(x, y_1) - \widehat{m}_n(x, y_2)| \leq k_n |y_1 - y_2|$$

Moreover, under the assumption of uniform convergence a.s of $\frac{\partial}{\partial y}\widehat{m}_n$ to $\frac{\partial}{\partial y}m$ (the notation $\frac{\partial}{\partial y}$ (resp. $\frac{\partial}{\partial x}$) represents the derivative with respect to the 2nd (resp.1st) variable), it is possible to define a global lipschitz factor for the functions $\widehat{m}_n$, $n \geq 0$, and m. This will be useful to check the stability of the inverse problem. Let first assume the following properties to ensure the uniform convergence of $\frac{\partial}{\partial y}\widehat{m}_n$ to $\frac{\partial}{\partial y}m$:

$[A_3]$ (i) The derivative of K, K', is bounded and lipchitzian

(ii) For $n \to +\infty$, $\sigma_n \to 0$ and $\frac{n\sigma_n^3}{\log n} \to +\infty$

Lemma 2.2 *: Under the assumptions* $[A_1 (s_1 = 1)]$, $[A_2 (r, s_2 = 1)]$ *with* $r \geq 2$ *,and* $[A_3]$*, there exists almost surely a real* $k > 0$ *which defines a unique Lipschitz factor for all the functions* $\widehat{m}_n$*,* $n \geq 0$*, and* m *on a compact neighborhood of* $(0, 0)$ *.*

2.3.3 Conclusion on our inverse problem

The last point to check is the stability of our solutions λ and $\widehat{\lambda}_n$, using the result of the previous lemma. Introduce another notation, the Picard operator associated to the equation (2):

$$T : \begin{cases} C_{b,0}(I) & \to C_{b,0}(I) \\ u & \to T(u) \end{cases}$$

such that:

$$\forall x \in I, T(u)(x) = \int_0^x m(s, u(s))ds$$

and define the estimated associated Picard operator:

$$\left(\widehat{T}_n u\right)(x) = \int_0^x \widehat{m}_n\left(s, u(s)\right) ds$$

As a consequence of the Cauchy-Lipschitz theorem, λ is a fixed point of T $(T(\lambda) = \lambda)$ and $\widehat{\lambda}_n$ is a fixed point of $\widehat{T}_n$ $(\widehat{T}_n\left(\widehat{\lambda}_n\right) = \widehat{\lambda}_n)$. Then, Lemma (2.2) allows to derive the following result:

$$d_\infty(\widehat{\lambda}_n,\lambda) \leq \frac{1}{1-c} d_{\Pi_k(I)}(\widehat{T}_n, T) \qquad (5)$$

where d_∞ and $d_{\Pi_k(I)}$ are uniform norms defined respectively on the subsets I and $C_{b,0}(I)$:

$$\begin{cases} d_\infty(u,v) & = \sup_{t\in I} |u(t)-v(t)| \\ d_{\Pi_k(I)}(T_1,T_2) & = \sup_{y\in C_{b,0}(I)} d_\infty\left(T_1 y, T_2 y\right) \end{cases}$$

and c is a constant that does not depend on n. (See appendix for details).

This inequality is quite interesting since it links known operators $\widehat{T}_n$ and T to unknown functions $\widehat{\lambda}_n$ and λ. Indeed, $\widehat{T}_n$ represents the integral of the kernel regression and it is easier to study the convergence of $\widehat{T}_n$ to T. Therefore, this inequality gives a first result of convergence for the interest parameter λ, under the condition of uniform convergence of $\widehat{T}_n$ to T. This is a first easy result to obtain without calculating explicitly the function λ. The next section will give results on the asymptotic rate of convergence of $\widehat{\lambda}_n$ to λ in mean square and in law. So, this relation is a first step in exploiting the link between m and λ, but of course, this first analysis needs to be investigated.

Moreover, this inequality allows to conclude that the inverse problem is stable, since the perturbation involved in the solution is controlled by the initial perturbation. Therefore, the system (2) defines a well-posed inverse problem, and there exists a unique solution which is stable.

The last result proves the uniform consistency in probability of our estimator. Indeed, following equation (5), we have:

$$\sup_{x\in I}\left|\widehat{\lambda}_n(x)-\lambda(x)\right| \leq C. \sup_{(x,y)\in D} |\widehat{m}_n(x,y)-m(x,y)|$$

where $C = \frac{a}{1-c}$.

Proposition 2.1 *: Assume the following properties* $[A_1\,(s_1=1)]$, $[A_2\,(r,s_2=1)]$, $[A_3]$, $I\!E\left(Z^2\right) < +\infty$. *Then,*

$$\sup_{x\in I}\left|\widehat{\lambda}_n(x)-\lambda(x)\right| \xrightarrow{P} 0$$

The next sections deal with numerical implementation, which is useful in practice, and results on the asymptotic convergence of the estimated solution $\widehat{\lambda}_n$ to the true one λ.

3 Numerical implementation

The numerical aspect of this method is quite important in order to implement a numerical approximation of the solution and of the error of estimation in concrete cases. Indeed, the system to solve is the following:

$$\begin{cases} \lambda'(x) & = \dfrac{\sum\limits_{i=1}^{n} Z_i K\left(\frac{x-X_i}{\sigma_n}\right) K\left(\frac{\lambda(x)-Y_i}{\sigma_n}\right)}{\sum\limits_{i=1}^{n} K\left(\frac{x-X_i}{\sigma_n}\right) K\left(\frac{\lambda(x)-Y_i}{\sigma_n}\right)} \\ \lambda(x_0) & = \lambda_0 \end{cases}$$

This expression is intentionally simplified by using a 1-dimensional kernel function. The idea is to find some intuition whether or not the estimated system could transform the initial problem (2) into an easier one to compute. Unfortunately, restrictions on the form of K to obtain an explicit estimated solution are not obvious, so the objective is now to stick to a more traditional approach and use some algorithm to approximate the solution.

Under Cauchy-Lipschitz assumptions, the Euler-Cauchy algorithm, which is a particular one step method of implementation, can be used. This is not the most impressive method but it is first a very easy method to implement and moreover as its speed of convergence is the slowest one, it will give an good idea of the performance that can be reached. Let briefly recall the methodology. The idea is to subdivide the interval $I = [-a, a]$ with p equidistant points $x_1, ..., x_p$ where $x_{i+1} = x_i + h$ and $h = 2a/p$ (assuming that $x_0 = \lambda_0 = 0$). Then, to each x_i is associated a λ_i defined recurrently by the following equation:

$$\begin{cases} \lambda_{i+1} & = \lambda_i + h m_h(x_i, \lambda_i) \\ \lambda_0 & = 0 \end{cases} \tag{6}$$

In the particular case of Euler algorithm, $m_h = \widehat{m}_n$. The solution approximated will be the polygonal map that joins the points (x_n, λ_n), and it converges to the exact solution, under the assumption that m_h follows the Lipschitz assumption. Let study more precisely the convergence of this solution. Define $e_i = \lambda_i - \lambda(x_i)$ the approximation error. If $\widehat{m}_n$ is a continuously differentiable function, then it can be proved that:

$$e_p = O(h) = O\left(\frac{1}{p}\right)$$

Therefore, denote by $\widehat{\lambda}_{nh}$ the solution of the approximated system above (6). Using an triangular inequality:

$$\left|\widehat{\lambda}_{nh} - \lambda\right| \leq \underbrace{\left|\widehat{\lambda}_{nh} - \widehat{\lambda}_n\right|}_{A} + \underbrace{\left|\widehat{\lambda}_n - \lambda\right|}_{B}$$

The second term B will be studied in the next section. Moreover, assume that $h = O(\frac{1}{n})$ (it is possible since there is no restriction in the choice of h). Then, $A = O\left(\frac{1}{n}\right) = o\left(\frac{1}{\sqrt{n}}\right)$. The result in the next section will also show that the rate of convergence of B to 0 in mean square is less than $\frac{1}{\sqrt{n}}$.

So, using the Euler approximation, when $\begin{cases} n \to +\infty \\ p \to +\infty \end{cases}$, $\widehat{\lambda}_{nh}$ converges to λ with the rate of convergence of B.

That was an example of numerical resolution. In fact, the Lipschitz assumption will allow to find various consistent methods, like Heun's method, Runge -Kutta's method (see Sibony & Mardon 1984). These more subtle methods differ by the use of different functions m_h, but of course the final rate of convergence will not change since they have a higher speed of convergence.

4 Asymptotic behavior of the estimated solution

This section gives theorems that establish conditions for convergence in mean square based on some linearization in probability and asymptotic normality. In what follows, it will always be assumed that the inverse problem is well-posed.

As noted previously, the system to solve is a nonlinear one that is why the first result concerns the linearization of the difference $\widehat{\lambda}_n - \lambda$ in order to study next its asymptotic properties. The methodology used is closely related to some delta method for functionals of the nonparametric kernel estimator. In a way, our objective is to use an extension of the classical framework as presented in van der Vaart and Wellner (1996), and derive properties for functionals of the conditional expectation kernel estimator.

4.1 Linearization of the problem

The following preliminary result tries to transform the nonlinear equation into a linear problem.

Proposition 4.1 *: (i) Under the assumptions of proposition 2.1, it can be proved that:*

$$\forall x \in I, \widehat{\lambda}_n(x) - \lambda(x) = \int_0^x \left((\widehat{m}_n - m)(t, \lambda(t)).e^{\left[\int_t^x \frac{\partial}{\partial y} m(u,\lambda(u))du\right]} \right) dt + R_{1,n}(x)$$

where $R_{1,n}(x) = o_P\left(\|\widehat{m}_n - m\|_\infty\right)$ *and* $\|\widehat{m}_n - m\|_\infty = \sup_{(x,y)\in D} |\widehat{m}_n(x,y) - m(x,y)|$.

(ii) Assume moreover that m *and* K *are continuously differentiable of order* 2*, then the previous decomposition can be transformed:*

$$\forall x \in I, \widehat{\lambda}_n(x) - \lambda(x) = \int_0^x \left((\widehat{m}_n - m)(t, \lambda(t)).e^{\left[\int_t^x \frac{\partial}{\partial y} m(u,\lambda(u))du\right]} \right) dt + R_{2,n}(x)$$

where $R_{2,n}(x) = O_P\left(\|\widehat{m}_n - m\|_2^2\right)$ *and* $\|\widehat{m}_n - m\|_2^2 = \iint_D (\widehat{m}_n - m)^2(x,y)\,dxdy$

This preliminary result is crucial for the rest of the paper. Indeed, our problem is splitted into two parts: a linear one that behaves as the integral of the conditional expectation, and a residual term that converges to zero. The next step is to study the convergence of each term.

Remark 4.1 : *Both residual terms $R_{1,n}(x)$ and $R_{2,n}(x)$ depend on the argument x, as well as the functions $\widehat{\lambda}_n(x)$ and $\lambda(x)$. The results we demonstrate in what follows are pointwise and not uniform with respect to x. More precisely, $\|\widehat{m}_n - m\|_\infty = \sup_{(x,y)\in D} |\widehat{m}_n(x,y) - m(x,y)|$ and*

$\|\widehat{m}_n - m\|_2^2 = \iint_D (\widehat{m}_n - m)^2(x,y)\,dxdy$ are calculated on the compact neighborhood D strictly included in the definition domain Θ. These norms do not depend on any argument x or y. Both residual terms can be expressed the following way:

$$\begin{aligned} R_{1,n}(x) &= o_P\left(\|\widehat{m}_n - m\|_\infty\right) \\ &= \|\widehat{m}_n - m\|_\infty\, o_P(1) \end{aligned}$$

$$\begin{aligned} R_{2,n}(x) &= O_P\left(\|\widehat{m}_n - m\|_2^2\right) \\ &= \|\widehat{m}_n - m\|_2^2\, O_P(1) \end{aligned}$$

and both $o_P(1)$ and $O_P(1)$ are functions of one argument x, like the functions λ and $\widehat{\lambda}_n$.

Remark 4.2 : *In order to stick to the most general case, assume that $\frac{\partial}{\partial x}m \neq 0$ and that $\frac{\partial}{\partial y}m \neq 0$. Indeed, if $\frac{\partial}{\partial y}m = 0$, the solution would simply be:*

$$\lambda(x) = \int_0^x m(t)dt$$

which will be much easier to study. Moreover, when $\frac{\partial}{\partial x}m = 0$, consider the particular autonomous equation, that is:

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

This is also a direct application of integration of nonparametric kernel estimators, with a $\sqrt{n}$ rate of convergence, the same result as expected for the case $\frac{\partial}{\partial y}m = 0$.

The following preliminary result allows us to get rid off the residual term in probability.

$[A_4]$ For $n \to +\infty$, $\sqrt{n}\sigma_n^{4+1/2} \to 0$

Proposition 4.2 : *Assume the following properties: $[A_1(s_1 = 2)]$, $[A_2(r, s_2 = 2)]$, $[A_3]$, $[A_4]$ and $I\!E\left(Z^2\right) < +\infty$. Then,*

$$\sqrt{n\sigma_n}\,\|\widehat{m}_n - m\|_2^2 \xrightarrow{P} 0$$

The next step is to concentrate on the asymptotic properties of the first term in the decomposition of proposition 4.1, that is mean square convergence and optimal choice of the bandwidth. The previous result will be useful at last to conclude for the asymptotic normality of our estimator, since the second part can be neglected in probability. Let us now concentrate on the first part of our decomposition.

4.2 Asymptotic mean square properties

The next result can be demonstrated with a second order kernel. This is a usual context for studying kernel estimators, and it is sufficient to prove the asymptotic consistency of our estimated solution. Moreover, we assume from now on that m is a continuously differentiable of order 2 function. Thanks to the assumption $[A_2(r, s_2 = 2)]$, the associated estimator is also continuously differentiable of order 2. We first derive a convergence in mean square for the linear part $\widehat{I}_n(x)$ of the following expansion:

$$\begin{aligned}\int_0^x \left((\widehat{m}_n - m)(t,\lambda(t)).e^{\left[\int_t^x \frac{\partial}{\partial y} m(u,\lambda(u))du\right]} \right) dt &= \int_0^x \left(\left(\frac{\widehat{\varphi}_n - m.\widehat{f}_n}{f} \right)(t,\lambda(t)).e^{\left[\int_t^x \frac{\partial}{\partial y} m(u,\lambda(u))du\right]} \right) dt \\ &\quad + R_{3,n}(x) \\ &= \widehat{I}_n(x) + R_{3,n}(x)\end{aligned}$$

Our purpose is to linearize the initial estimator with respect of the density function in order to study the asymptotic mean square error. We derive some asymptotic results on $\widehat{I}_n(x)$ and show that the residual term $\sqrt{n\sigma_n}R_{3,n}(x)$ converges in probability to zero.

Theorem 4.1 : *Assume the following properties:* $[A_1(s_1 = 2)]$, $[A_2(r = 2, s_2 = 2)]$, $[A_3]$, $[A_4]$ *and* $\mathbb{E}\left(Z^2\right) < +\infty$. *Assume moreover that the partial derivatives of* f *are bounded. Then:*

$$\begin{aligned}\forall x \in I, \mathbb{E}\left[\widehat{I}_n(x)\right]^2 &= \tfrac{1}{n\sigma_n}\|K\|_2^2 \int_0^x \tfrac{v(t,\lambda(t))}{f(t,\lambda(t))}\gamma(x,t)^2 dt.\left(1+o\left(1\right)\right) \\ &+ \tfrac{\sigma_n^4}{4}\left(\int u^2K(u)du\right)^2 \left[\begin{array}{c} \displaystyle\int_0^x \gamma(x,t) \int \left(\tfrac{z-m(t,\lambda(t))}{f(t,\lambda(t))}\right). \\ \left(\tfrac{\partial^2 f_S}{\partial x^2}(t,\lambda(t),z) + \tfrac{\partial^2 f_S}{\partial y^2}(t,\lambda(t),z)\right) dzdt \end{array} \right]^2 .\left(1+o\left(1\right)\right)\end{aligned}$$

where

$$\gamma(x,t) = e^{\left[\int_t^x \frac{\partial}{\partial y} m(u,\lambda(u))du\right]}$$

and

$$v\left(t,\lambda(t)\right) = \int \left(z - m\left(t,\lambda\left(t\right)\right)\right)^2 \frac{f_S(t,\lambda(t),z)}{f(t,\lambda(t))}dz$$

is a version of $Var\left(Z \,|X = t, Y = \lambda(t)\right)$

Proposition 4.3 : *Under the assumptions of proposition 4.2,*

$$\forall x \in I, \sqrt{n\sigma_n}R_{3,n}(x) \xrightarrow{P} 0$$

This first theorem shows the consistency in mean square of $\widehat{I}_n(x)$.

Some comparisons can already be derived with the initial nonparametric kernel estimator $\widehat{m}_n$. Indeed, recall the expansion of the asymptotic quadratic error for the kernel regression estimator. The asymptotic variance is of order $\frac{1}{n\sigma_n^2}$ and the square bias is of order σ_n^4 (Bosq 1998).

The difference is in the variance term and it comes from the fact that there is a gain in dimension by applying an integral operator to the initial kernel regression. Indeed, the solution λ is in dimension 1 whereas the function m is in dimension 2 .

4.3 Asymptotic optimal choice

The objective of this sub-section is to study more carefully the asymptotic optimal choice of the bandwidth parameter by taking into account the particular smoothing properties of our inverse problem. As it has been proved earlier, solving the differential equation allows to transform a continuous differentiable function of order 2, the function m, into a continuous differentiable function of order 3, the function λ. Therefore, the minimax properties of our estimated interest parameter, $\widehat{\lambda}_n$, should be studied within this second class of functions.

The next result simply derives the classical optimal bandwidth, obtained by minimizing the asymptotic mean square error with respect to σ_n for a fixed n. We will then extend this result and show that we can derive a smaller optimal bandwidth, more adapted to the C^3-class of functions.

Let first present the result of optimality with a second order kernel.

Corollary 4.1 *: Under the previous assumptions,*
if $\left[\int_0^x \gamma(x,t) \int \left(\frac{z-m(t,\lambda(t))}{f(t,\lambda(t))}\right) \cdot \left(\frac{\partial^2 f_S}{\partial x^2}(t,\lambda(t),z) + \frac{\partial^2 f_S}{\partial y^2}(t,\lambda(t),z)\right) dzdt\right] \neq 0$, *the value of* σ_n *which minimizes the asymptotic mean square error for a fixed* n *is:*

$$\sigma_n^* = \left(\frac{4\|K\|_2^2 \int_0^x \frac{v(t,\lambda(t))}{f(t,\lambda(t))}\gamma(x,t)^2 dt}{n\left[\int u^2 K(u)\right]^2 \left[\int_0^x \gamma(x,t) \int \left(\frac{z-m(t,\lambda(t))}{f(t,\lambda(t))}\right) \cdot \left(\frac{\partial^2 f_S}{\partial x^2}(t,\lambda(t),z) + \frac{\partial^2 f_S}{\partial y^2}(t,\lambda(t),z)\right) dzdt\right]^2} \right)^{1/5}$$

Our next issue is to take into account the properties of the solution λ to achieve a better rate of convergence, that is the minimax rate of convergence for a given set of functions. Indeed, since m is assumed to be continuously differentiable of order 2, λ is then continuously differentiable of order 3. Therefore, the optimality will not reached with a kernel of order 2, but of order 3.

The next results show that, without assuming that m is C^3 for the two arguments, but only for the second one, a better rate of convergence can be achieved, by taking into account the properties of the integral operator applied to the kernel regressor. Let now present the result of convergence of the quadratic error with the 3rd order kernel.

Theorem 4.2 : *Assume the following properties:* $[A_1\,(s_1=2)]$, $[A_2\,(r=3,s_2=2)]$, $[A_3]$, $[A_4]$ *and* $I\!E\left(Z^2\right)<+\infty$. *Assume moreover that* f *is continuously differentiable of order* 3 *with respect to the second argument and that the partial derivatives of* f *are bounded. Then, the following expansion can be derived:*

$$\forall x \in I, I\!E\left[\widehat{I}_n(x)\right]^2 = \frac{1}{n\sigma_n}\|K\|_2^2\int_0^x \frac{v(t,\lambda(t))}{f(t,\lambda(t))}\gamma(x,t)^2dt.\left(1+o\left(1\right)\right)$$
$$+\gamma_1^2(x)\frac{\sigma_n^6}{36}\left(\int u^3K(u)du\right)^2$$
$$.\left[\int\left\{\begin{array}{c}\int_0^x\left(\alpha(t,z).\frac{\partial}{\partial y^3}f_S(t,\lambda\left(t\right),z)\right)dt-\alpha(x,z)\left[\frac{\partial^2}{\partial x^2}f_S\left(x,\lambda\left(x\right),z\right)-\frac{\partial^2}{\partial x^2}f_S\left(0,0,z\right)\right]\\+\int_0^x\frac{\partial}{\partial x}\alpha(t,z)\left[\frac{\partial^2}{\partial x^2}f_S\left(x,\lambda\left(x\right),z\right)-\frac{\partial^2}{\partial x^2}f_S\left(0,0,z\right)\right]dt\end{array}\right\}dz\right]^2$$
$$.\left(1+o\left(1\right)\right)$$

where

$$\left\{\begin{array}{l}\gamma(x,t)=e^{\left[\int_0^x\frac{\partial}{\partial y}m(u,\lambda(u))du-\int_0^t\frac{\partial}{\partial y}m(u,\lambda(u))du\right]}=\gamma_1(x).\gamma_2(t)\\ \alpha(t,z)=\frac{z-m(t,\lambda(t))}{f(t,\lambda(t))}\gamma_2(t)\end{array}\right.$$

The 3-order kernel reduces the bias term. This difference can be explained by a gain in regularity for the estimated solution, compared to the bias for the kernel regressor. Indeed, as mentioned before, the initial function m was continuously differentiable of order 2 and the solution λ is then continuously differentiable of order 3.

Corollary 4.2 *Under the same assumptions,*

$$\text{if } A=\int\left\{\begin{array}{c}\int_0^x\left(\alpha(t,z).\frac{\partial}{\partial y^3}f_S(t,\lambda\left(t\right),z)\right)dt-\alpha(x,z)\left[\frac{\partial^2}{\partial x^2}f_S\left(x,\lambda\left(x\right),z\right)-\frac{\partial^2}{\partial x^2}f_S\left(0,0,z\right)\right]\\+\int_0^x\frac{\partial}{\partial x}\alpha(t,z)\left[\frac{\partial^2}{\partial x^2}f_S\left(x,\lambda\left(x\right),z\right)-\frac{\partial^2}{\partial x^2}f_S\left(0,0,z\right)\right]dt\end{array}\right\}dz\neq 0,$$

the value of σ_n *which minimizes the asymptotic mean square error for a fixed* n *is:*

$$\sigma_n^*=\left(\frac{36\|K\|_2^2\int_0^x\frac{v(t,\lambda(t))}{f(t,\lambda(t))}\gamma_2(t)^2dt}{n\left[\int u^3K(u)\right]^2A^2}\right)^{1/7}$$

Again, there is difference with the result for the kernel regression, since the optimal bandwidth was of order $\left(\frac{1}{n}\right)^{1/6}$.

The interest of this result is to take benefit from all the features of our inverse problem in order to obtain some better optimal results. Indeed, it has already been proved that the rate of convergence for the regression kernel is the best attainable in a minimax sense. Moreover, the optimal bandwidth $\sigma_n^* \sim n^{-1/7}$ leads to the minimax speed of convergence for the continuously differentiable function of order 3 in dimension 1. The estimated solution then achieves the optimal convergence of order $n^{6/7}$.

4.4 Asymptotic normality

Let now derive an asymptotic normality theorem. This result is less new and has already been proved by Hausman and Newey (1995). In particular, the faster convergence rate for the solution is analogous to the results on averaging kernel estimators over some arguments that are given in Linton and Nielsen (1995) and Newey (1994).

Two theorems are presented, with the two different assumptions for the kernel order.

Theorem 4.3 *: Assume the following properties:* $[A_1\,(s_1=2)]$, $[A_2\,(r=2, s_2=2)]$, $[A_3]$, $[A_4]$, *and* $\mathbb{E}\left(Z^4\right) < +\infty$. *Assume moreover that the partial derivatives of* f *are bounded, that* $n\sigma_n^5 \to 0$.

Therefore, $\sqrt{n\sigma_n}\left(\widehat{\lambda}_n - \lambda\right)(x)$ *converges in law to a centered Gaussian distribution with an asymptotic variance term equal to:*

$$V = \|K\|_2^2 \int_0^x \frac{v\left(t, \lambda(t)\right)}{f(t, \lambda(t))} \gamma(x,t)^2 dt$$

Remark 4.3 *: Under the assumptions of the previous theorem, the asymptotic distribution of* $\sqrt{n\sigma_n}\left(\widehat{\lambda}_n - \lambda\right)(x)$ *is centered to zero. An asymptotic bias equal to zero would have been obtained by undersmoothing the estimator and so by assuming that* σ_n *converges faster to zero (than the optimal one). Under the assumption of optimal bandwidth choice, the asymptotic bias term is then equal to:*

$$\frac{1}{2}\left(\int u^2 K(u)du\right) \left[\begin{array}{c} \int_0^x \gamma(x,t) \int \left(\frac{z-m(t,\lambda(t))}{f(t,\lambda(t))}\right) \cdot \\ \left(\frac{\partial^2}{\partial x^2} f_S\left(t, \lambda\left(t\right), z\right) + \frac{\partial^2}{\partial y^2} f_S\left(t, \lambda\left(t\right), z\right)\right) dz dt \end{array} \right]$$

Theorem 4.4 *: Assume the following properties:* $[A_1\,(s_1=2)]$, $[A_2\,(r=3, s_2=2)]$, $[A_3]$, $[A_4]$, *and* $\mathbb{E}\left(Z^4\right) < +\infty$. *Assume moreover that* f *is continuously differentiable of order* 3 *with respect to the second argument and that the partial derivatives of* f *are bounded. Under the additinal assumption* $n\sigma_n^7 \to 0$, *it can be proved that* $\sqrt{n\sigma_n}\left(\widehat{\lambda}_n - \lambda\right)(x)$ *converges in law to a centered Gaussian distribution with an asymptotic variance term equal to:*

$$V = \|K\|_2^2 \int_0^x \frac{v\left(t, \lambda(t)\right)}{f(t, \lambda(t))} \gamma(x,t)^2 dt$$

Remark 4.4 *: Under the assumption of optimal bandwidth choice, the asymptotic bias term is then* *equal to:*

$$\frac{\gamma_1(x)}{6}\left(\int u^3K(u)du\right)$$
$$.\left[\int\left\{\begin{array}{c}\int_0^x\left(\alpha(t,z).\frac{\partial}{\partial y^3}f_S(t,\lambda(t),z)\right)dt-\alpha(x,z)\left[\frac{\partial^2}{\partial x^2}f_S(x,\lambda(x),z)-\frac{\partial^2}{\partial x^2}f_S(0,0,z)\right]\\ +\int_0^x\frac{\partial}{\partial x}\alpha(t,z)\left[\frac{\partial^2}{\partial x^2}f_S(x,\lambda(x),z)-\frac{\partial^2}{\partial x^2}f_S(0,0,z)\right]dt\end{array}\right\}dz\right]$$

The asymptotic variance given previously $V=\|K\|_2^2\int_0^x\frac{v(t,\lambda(t))}{f(t,\lambda(t))}\gamma(x,t)^2dt$ is very clearly depending on x through the second integral. In particular, when x is very close to zero, the variance is also close to zero. This is a logical property since the solution at point 0 is known and equal to 0. Therefore, around this point, there is less incertitude about the estimated solution.

The asymptotic variance estimation has also been discussed by Hausman and Newey (1995) and Newey (1994). It is indeed useful for inference based on the asymptotic normality result and construction of confidence intervals. The asymptotic variance term can of course be estimated by be replacing the unknown parameters by kernel estimators in order to study the consistence of this estimated variance (like Newey 1994). An alternative approach to variance estimation, or confidence intervals, is the bootstrap technique. The idea is then to sample the data randomly with replacement, to estimate several times the solution $\widehat{\lambda}_n$ and calculate the empirical variance of these solutions. Thus, the bootstrap provides a way to substitute computation for mathematical analyisi if calculating the asymptotic distribution of an estimator is difficult (see for example Horowitz 2001).

Remark 4.5 *: In the classical studies of density kernel estimation and conditional expectation kernel estimation, the covariance term can be neglected since its rate of convergence to zero is faster than the variance one. In this particular problem, let analyse the covariance between* $\widehat{I}_n(x)$ *and* $\widehat{I}_n(y)$ *for* $x\neq y$*.*

$$cov\left(\widehat{I}_n(x),\widehat{I}_n(y)\right)=cov\left(\int_0^x\left(\left(\frac{\widehat{\varphi}_n-m.\widehat{f}_n}{f}\right)(t,\lambda(t)).\gamma(x,t)\right)dt,\int_0^y\left(\left(\frac{\widehat{\varphi}_n-m.\widehat{f}_n}{f}\right)(s,\lambda(s)).\gamma(y,s)\right)ds\right)$$
$$=\frac{1}{n\sigma_n^2}cov\left(\begin{array}{c}\frac{Z-m(X,\lambda(X))}{f(X,\lambda(X))}\gamma(x,X)\mathbb{1}_{0\leq X\leq x}K\left(\frac{\lambda(X)-Y}{\sigma_n}\right)\\ ,\frac{Z-m(X,\lambda(X))}{f(X,\lambda(X))}\gamma(y,X)\mathbb{1}_{0\leq X\leq y}K\left(\frac{\lambda(X)-Y}{\sigma_n}\right)\end{array}\right).(1+o(1))$$
$$=\frac{1}{n\sigma_n}\|K\|_2^2.\int_0^{\min(x,y)}\frac{\gamma(x,t)\gamma(y,t)}{f(t,\lambda(t))}v(t,\lambda(t))dt.(1+o(1))$$

Therefore, contrary to the usual results, as pointed by Horowitz (2001), the covariance term cannot be neglected in this case. An explanation of this result is that both arguments of the conditional expectation function $m(t,\lambda(t))$ *depend on the same variable* t*. Under a more general hypothesis like* $m(x,y)$*, the covariance term would have been neglected compared to the variance one.*

5 Conclusion

This paper has been concerned with estimating the solution of a differential equation depending on the law of distribution of the dataset. Some empirical examples can illustrate this problem. One of them is an article from Florens Vanhems (2000) which deals with a physical case of measuring the ionosphere thickness. Another could be an economical extension of Hausman and Newey 's work with the measure of profit variation of a firm. For example, it can be used to measure the impact of new postal tariffing over a customer firm (denoted by (C)). Let then study the impact of a factor's price variation on the surplus of (C). It can be shown that the relation between the profit variation of (C) and the function of demand factor is also a differential equation of order 1 (like the relation found by Hausman and Newey between the variation of consumer surplus and the demand function). The previous results can then be applied to obtain an asymptotic study of the estimated solution.

Other applications and extensions can also be studied. For example, the next step could be to study random differential equations, that is equations depending on a unknown function F_S and also on an error term ε which perturbed the equation.

$$\begin{cases} \lambda'(x) & = \phi(x, \lambda(x), F_S, \varepsilon) \\ \lambda(x_0) & = \lambda_0 \end{cases}$$

A natural application is to consider for example the microeconomic problem of Hausman and Newey on the variation of exact consumer surplus and to assume that the residual in the underlying econometric model contains heterogeneity between individuals. Therefore, the error term cannot be neglected anymore as a measurement error but has to be taken into account into the differential equation. In particular, this perturbation will have to be estimated before solving the equation. Moreover, an interesting issue could be to compare the unperturbed solution with an aggregated solution with respect to ε, and to link it with the theory of representative consumer.(this problem is developed in Vanhems 2001)

Another extension to study is the case of endogenous variables. This is a direct application of the general relation:

$$\begin{cases} \lambda'(x) & = m(x, \lambda(x)) \\ \lambda(x_0) & = \lambda_0 \end{cases}$$

where m will no more represent a conditional expectation but something more complex defined with instrumental variables. Such a problem has been in particular studied in Loubes Vanhems (2002).

6 Mathematical appendix

6.1 Inversion of operator and Picard contraction mapping

This section presents the mathematical tools used in the article.

A solution to the initial problem satisfies the equivalent integral equation:

$$\lambda(x) = \lambda_0 + \int_{x_0}^{x} m\left(s, \lambda(s)\right) ds$$

The Picard operator T associated with the system (2) is defined as:

$$\left(Tu\right)(x) = \lambda_0 + \int_{x_0}^{x} m\left(s, u(s)\right) ds \tag{7}$$

Let $I = [x_0 - a, x_0 + a]$ for $a > 0$ and $C(I)$ be the Banach space of continuous functions $x(t)$ on I with the sup norm $\|x\|_\infty = \max_{t \in I} |x(t)|$. Then $T : C(I) \to C(I)$. Furthermore, a solution $\lambda(t)$ to (2) is a fixed point of T.

In what follows, as before, let $x_0 = \lambda_0 = 0$, so that $I = [-a, a]$.

Recall that $D = \{(x, y)\,; |x| \leq a, |y| \leq b\}$ and $C_{b,0}(I) = \{u \in C(I); u(0) = 0, \|u\|_\infty \leq b\}$.

Define now a metric on $C_{b,0}(I)$ that is:

$$d_\infty(u, v) = \|u - v\|_\infty = \sup_{t \in I} |u(t) - v(t)|\,, \; \forall u, v \in C_{b,0}(I)$$

$(C_{b,0}(I), d_\infty)$ is a complete metric space.

Another way to write lemma 2.1 is to present the following result:

Lemma 6.1 *Under assumptions* $[a]$ *and* $[b]$, T *is a* $C_{b,0}(I)$*-valued contraction map, with c as contraction factor.*

The common proof for those two lemmas is the following:

Proof. Lemmas 6.1 and 2.1

$\forall u, v \in C_{b,0}(I)$:

$$\begin{aligned}
d_\infty(Tu, Tv) &= \sup_{x \in I} \left| \int_0^x \left(m\left(s, u(s)\right) - m\left(s, v(s)\right)\right) ds \right| \\
&\leq \sup_{x \in I} \left| \int_0^x \left|\left(m\left(s, u(s)\right) - m\left(s, v(s)\right)\right)\right| ds \right| \\
&\leq \sup_{x \in I} \left| \int_0^x k\left|u(s) - v(s)\right| ds \right| \\
&\leq a k d_\infty(u, v) \\
&\leq c d_\infty(u, v)
\end{aligned}$$

So, T is a contraction since $0 < c < 1$. ■

Thanks to Banach's fixed point theorem, there exists a unique fixed point $\lambda \in C_{b,0}(I)$, and therefore, a unique solution to (2). The compact set D is usually called the "the safety cylinder". Denote $\Pi_k(I)$ the set of all Picard operators $T : C_{b,0}(I) \to C_{b,0}(I)$ as defined in (7) where m satisfies the two assumptions above. In particular, all the functions belonging to $\Pi_k(I)$ are contractions, with the same contraction factor $c = ka < 1$.

Recall now a general result that will be useful for what follows.

Theorem 6.1 : *Under [a] and [b], let $T_1, T_2 \in \Pi_k(I)$, with c as contraction factor. Define the following metric on $\Pi_k(I)$:*

$$d_{\Pi_k(I)}(T_1, T_2) = \sup_{y \in C_{b,0}(I)} d_\infty (T_1 y, T_2 y)$$

If λ_1 and λ_2 denote the respective fixed points of T_1 and T_2, then

$$d_\infty(\lambda_1, \lambda_2) \leq \frac{1}{1-c} d_{\Pi_k(I)}(T_1, T_2)$$

Proof. Indeed:

$$\begin{aligned} d_\infty(\lambda_1, \lambda_2) &= d_\infty(T_1\lambda_1, T_2\lambda_2) \\ &\leq d_\infty(T_1\lambda_1, T_1\lambda_2) + d_\infty(T_1\lambda_2, T_2\lambda_2) \\ &\leq c d_\infty(\lambda_1, \lambda_2) + d_{\Pi_k(I)}(T_1, T_2) \\ &\leq \frac{1}{1-c} d_{\Pi_k(I)}(T_1, T_2) \end{aligned}$$

Thus, the result is proved. ∎

Remark 6.1 : *Let check that the metric $d_{\Pi_k(I)}$ is well defined on $C_{b,0}(I)$. In particular let verify that:*

$$d_{\Pi_k(I)}(T_1, T_2) < +\infty, \; T_1, T_2 \in \Pi_k(I)$$

So:

$$\begin{aligned} \sup_{u \in C_{b,0}(I)} \|T_1 u - T_2 u\|_\infty &= \sup_{u \in C_{b,0}(I)} \left\| \int_0^x \left[m_1(s, u(s)) - m_2(s, u(s))\right] ds \right\|_\infty \\ &\leq \sup_{u \in C_{b,0}(I)} \max_{x \in I} \left| \int_0^x \left|m_1(s, u(s)) - m_2(s, u(s))\right| ds \right| \\ &\leq a \max_{(x,y) \in D} |m_1(x, y) - m_2(x, y)| \\ &\leq 2b \end{aligned}$$

Thus, without restriction, it is possible to define $d_{\Pi_k(I)}(T_1, T_2)$, $T_1, T_2 \in \Pi_k(I)$.

Remark 6.2 : *Another way to demonstrate this is to recall that the set of lipschitzian functions with the same lipschitz factor k defined from real compact sets to real compact sets is compact for the uniform convergence topology (see Choquet 1964 p.94).*

Remark 6.3 *: Thanks to the theorem just proved, our problem is a well-posed one under the compact set D since existence and uniqueness of a solution stable under small variations have been proved. This remark is very important since it is posssible now to try to inverse the differential operator.*

6.2 Proofs

6.2.1 Proof of Lemma 2.2

Consider the smallest Lipschitz factors for m and $\widehat{m}_n$, $n \geq 0$. Therefore:

$$\begin{aligned} \forall n, \forall (x, y_i) &\in D, |\widehat{m}_n(x, y_1) - \widehat{m}_n(x, y_2)| \leq k_n |y_1 - y_2| \\ k_n &= \max_D \left| \frac{\partial}{\partial y} \widehat{m}_n \right| \end{aligned}$$

and

$$\begin{aligned} \forall (x, y_i) &\in D, |m(x, y_1) - m(x, y_2)| \leq k_0 |y_1 - y_2| \\ k_0 &= \max_D \left| \frac{\partial}{\partial y} m \right| \end{aligned}$$

These factors k_n and k_0 exist since we maximize continuous functions on a compact set.

Moreover, since $\frac{\partial}{\partial y}\widehat{m}_n$ converges uniformly to $\frac{\partial}{\partial y}m$:

$$\forall \varepsilon, \exists n_0, \forall n \geq n_0, \forall (x, y) \in D, \left| \frac{\partial}{\partial y} \widehat{m}_n(x, y) - \frac{\partial}{\partial y} m(x, y) \right| \leq \varepsilon \; as$$

So:

$$\forall \varepsilon, \exists n_0, \forall n \geq n_0, \forall (x, y) \in D, \left| \frac{\partial}{\partial y} \widehat{m}_n(x, y) \right| \leq \varepsilon + \left| \frac{\partial}{\partial y} m(x, y) \right|$$

and

$$\forall \varepsilon, \exists n_0, \forall n \geq n_0, k_n \leq \varepsilon + k_0$$

Let now define

$$k = \max(\varepsilon + k_0, (k_j)_{j \leq n_0})$$

Then, it will always be possible to find a compact neighborhood $I = [-a, a]$ around 0, $a > 0$, such that $ak < 1$.

For example, let use:

$$a = \inf \left\{ \frac{1}{2(k + \varepsilon)}, \frac{1}{2k_n}, n \leq n_0 \right\}$$

The constraints that allow to use to uniform convergence are provided by Bosq (1998) theorem 3.2 p.73 .

Thus, $T \in \Pi_{k_0}(I) \subset \Pi_k(I)$ and $\widehat{T}_n \in \Pi_k(I)$. The result is demonstrated.

6.2.2 Proof of proposition 2.1

The result follows directly from the inequality:

$$\sup_{x\in I}\left|\widehat{\lambda}_n(x)-\lambda(x)\right|\leq C.\sup_{(x,y)\in D}|\widehat{m}_n(x,y)-m(x,y)|$$

The convergence of $\|\widehat{m}_n-m\|_\infty$ is proved using the assumption $[A_3]$ (indeed, the assumptions for the uniform convergence of $\frac{\partial}{\partial y}\widehat{m}_n$ to $\frac{\partial}{\partial y}m$ are stronger).

6.2.3 Proof of proposition 4.1

Under the previous assumptions $[A1]$ and $[A2]$, there exists a unique solution to (2) $\lambda(x)=\Phi\left[m\right](x)$. The objective is to try to characterize the functional Φ that is the exact dependance between λ and m. The methodology used is in particular very close to the well-known Delta method (as detailed for example in van der Vaart and Wellner 1996).

The objective is to study the following function:

$$A:\begin{cases} C^1(D)\times C^1_{b,0}(I) & \to C(I)\\ (u,v) & \mapsto A(u,v)\end{cases}$$

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

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

$\left(C^1(D),\|.\|_\infty\right)$ and $(C(I),\|.\|_\infty)$ are Banach spaces.

Moreover define $\|.\|_\infty^{'}=\max\left(\|v\|_\infty,\|v'\|_\infty\right)$ on $C^1_{b,0}(I)$. Then $\left(C^1_{b,0}(I),\|.\|_\infty^{'}\right)$ is a Banach space. To demonstrate it, use the uniform convergence of functions and its application to derivability. The use of such a norm allows to have the continuity and linearity of the following function:

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

So: $\forall x\in I, A(u,v)(x)=v'(x)-u(x,v(x))$. Denote by W, an open subset of $C^1(D)\times C^1_{b,0}(I)$ and $(m,\lambda)\in W$. Then A is continuous on W (it is a sum of continuous functions) and $A(m,\lambda)=0$. Let check the hypothesis of the implicit function theorem. A is in fact continuously differentiable (thanks to the same argument) so it is possible to take its derivative with the second variable $d_2A(m,\lambda)$. Moreover:

$$\forall h\in C^1_{b,0}(I),\forall x\in I, d_2A(m,\lambda)(h)(x)=h'(x)-\frac{\partial}{\partial y}m(x,\lambda(x)).h(x)$$

The next step is then to prove that $d_2A(m,\lambda)$ is a bijection. Let show first the surjectivity:

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

This is a linear differential equation that can be solved in order to find:

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

Therefore, $d_2A(m,\lambda)$ is surjective. The next step is then to demonstrate the injectivity of the function, that is $Ker\left(d_2A(m,\lambda)\right) = \{0\}$. For that purpose, let solve $d_2A(m,\lambda)(h) = 0,\ h \in C^1_{b,0}(I)$. A linear differential equation is then also obtained and its solution is equal to:

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

Therefore, $c = 0$. Thus, $d_2A(m,\lambda)$ is bijective.

Then at last, it is necessary to demonstrate the bi-continuity of $d_2A(m,\lambda)$. In the usual implicit function theorem, this assumption is not required, but here the use of infinite dimension spaces requires a more general theorem with further assumptions to satisfy. The continuity of $d_2A(m,\lambda)$ has already been proved since A is continuously differentiable. The continuity of the reversible function is given by an application of Baire Theorem: if an function is linear continuous and bijective on two Banach spaces, the reversible function is continuous.

Therefore, it is possible now to apply the implicit function theorem: $\exists U$ an open subset around m and V an open subset around λ such that:

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

Note: $y = \Phi[u]$ this unique solution for $u \in U$. By construction, Φ is continuously differentiable of order 1 on U.

Now differentiate the relation: $A(u,\Phi[u]) = 0, \forall u \in U$ and apply it in $(m, \lambda = \Phi[m])$. Then, the result is:

$\forall h \in C^1(D) \times C^1_{b,0}(I)$,

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

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

$$d\Phi[m](\widehat{m}_n - m)'(x) = \frac{\partial}{\partial y} m\left(x, \Phi[m](x)\right).d\Phi[m](\widehat{m}_n - m)(x) + (\widehat{m}_n - m)\left(x, \Phi[m](x)\right)$$

Solving it leads to:

$$d\Phi[m](\widehat{m}_n - m)(x) = \int_0^x \left(\begin{array}{c} (\widehat{m}_n - m)(t, \Phi[m](t)). \\ e^{\left[\int_0^x \frac{\partial}{\partial y} m(u,\Phi[m](u))du - \int_0^t \frac{\partial}{\partial y} m(u,\Phi[m](u))du\right]} \end{array} \right) dt$$

$$= \int_0^x \left((\widehat{m}_n - m)(t, \lambda(t)).e^{\left[\int_t^x \frac{\partial}{\partial y} m(u,\lambda(u))du\right]} \right) dt$$

$$= \int_0^x \left((\widehat{m}_n - m)(t, \lambda(t)).\gamma(x,t) \right) dt$$

where $\gamma(x,t) = e^{\left[\int_t^x \frac{\partial}{\partial y} m(u,\lambda(u))du\right]}$.

At last, the definition of the differentiability of Φ gives:

$$\left(\widehat{\lambda}_n - \lambda\right)(x) = (\Phi[\widehat{m}_n] - \Phi[m])(x)$$
$$= d\Phi[m](\widehat{m}_n - m)(x) + R_{1,n}(x)$$

where $R_{1,n}(x)$ is the remainding term of the Taylor development $R_{1,n}(x) = o_P(\|\widehat{m}_n - m\|_\infty)$.

The last point (ii) is to replace the usual residual term by $O_P\left(\|\widehat{m}_n - m\|_2^2\right)$, which is a stronger requirement than the usual definition (but still weaker than second-order differentiability). Assuming that m and K are continuously differentiable of order 2 implies that the operator A and then Φ are also continuously differentiable of order 2, so it is possible to replace the usual neglected term by:

$$R_{2,n}(x) = O_P\left(\|\widehat{m}_n - m\|_\infty^2\right) = O_P\left(\|\widehat{m}_n - m\|_2^2\right)$$

6.2.4 Proof of proposition 4.2

The proof follows directly from the result in mean square convergence of the conditional expectation. Indeed, we integrate over the compact neighborhood D which is strictly included in the definition domain Θ. This property allows us to avoid the usual boundary effects, and the use of generalized kernels to correct them (for example).

6.2.5 Proof of theorem 4.1

Let us derive the convergence in mean square of the linearized term: $\widehat{I}_n(x)$.

$$\mathbb{E}\left[\widehat{I}_n(x)^2\right] = var\left(\widehat{I}_n(x)\right) + \left(\mathbb{E}\left[\widehat{I}_n(x)\right]\right)^2$$

The first term can be developed:

$$var\left[\widehat{I}_n(x)\right] = var\left[\iint \frac{(z - m(t,\lambda(t)))}{f(t,\lambda(t))} 1\!\!I_{0\le t\le x}\gamma(x,t)\widehat{f}_{Sn}(t,\lambda(t),z)\,dzdt\right]$$

and by definition:

$$\begin{aligned}
var\left[\widehat{I}_n(x)\right] &= var\left[\frac{1}{n\sigma_n^2}\sum_{i=1}^{n}\int \frac{(Z_i - m(t,\lambda(t)))}{f(t,\lambda(t))} 1\!\!I_{0\le t\le x}\gamma(x,t)K\left(\frac{t-X_i}{\sigma_n}\right)K\left(\frac{\lambda(t)-Y_i}{\sigma_n}\right)dt\right] \\
&= var\left[\frac{1}{n\sigma_n}\sum_{i=1}^{n}\frac{(Z_i - m(X_i,\lambda(X_i)))}{f(X_i,\lambda(X_i))} 1\!\!I_{0\le X_i\le x}\gamma(x,X_i)K\left(\frac{\lambda(X_i)-Y_i}{\sigma_n}\right)\right].(1+o(1)) \\
&= \frac{1}{n\sigma_n}\|K\|_2^2.\iint \frac{(z - m(t,\lambda(t)))^2}{f(t,\lambda(t))} 1\!\!I_{0\le t\le x}^2\gamma(x,t)\frac{f_S(t,\lambda(t),z)}{f(t,\lambda(t))}dtdz + o\left(\frac{1}{n\sigma_n}\right) \\
&= \frac{\|K\|_2^2}{n\sigma_n}\int_0^x \frac{v(t,\lambda(t))}{f(t,\lambda(t))}\gamma(x,t)^2dt + o\left(\frac{1}{n\sigma_n}\right)
\end{aligned}$$

where $v(t,\lambda(t)) = var(Z\,|X=t, Y=\lambda(t))$.

Therefore, the calculus gives:

$$var\left[\widehat{I}_n(x)\right] = \frac{1}{n\sigma_n}\|K\|_2^2\int_0^x \frac{v(t,\lambda(t))}{f(t,\lambda(t))}\gamma(x,t)^2dt + o\left(\frac{1}{n\sigma_n}\right)$$

Let consider now the bias term.

$$\begin{aligned}
I\!\!E\left[\widehat{I}_n(x)\right] &= I\!\!E\left[\int_0^x \left(\frac{\widehat{\varphi}_n - m.\widehat{f}_n}{f}\right)(t,\lambda(t))\,\gamma(x,t)dt\right] \\
&= \int_0^x I\!\!E\left[\left(\frac{\widehat{\varphi}_n - m.\widehat{f}_n}{f}\right)(t,\lambda(t))\right]\gamma(x,t)dt
\end{aligned}$$

Therefore, using classical results on the bias decomposition for the conditional expectation, it can be proved that:

$$\left(I\!\!E\left[\widehat{I}_n(x)\right]\right)^2 = \frac{\sigma_n^4}{4}\left(\int u^2K(u)du\right)^2\left[\begin{array}{c}\int_0^x \gamma(x,t)\int\left(\frac{z-m(t,\lambda(t))}{f(t,\lambda(t))}\right) \\ \left(\frac{\partial^2}{\partial x^2}f_S(t,\lambda(t),z) + \frac{\partial^2}{\partial y^2}f_S(t,\lambda(t),z)\right)dzdt\end{array}\right]^2 + o(\sigma_n^4)$$

6.2.6 Proof of proposition 4.3

By definition, we have:

$$\begin{aligned}
\sqrt{n\sigma_n}R_{3,n}(x) &= \sqrt{n\sigma_n}\int_0^x \left(\left(\frac{\widehat{\varphi}_n - m.\widehat{f}_n}{f^2}\right)(t,\lambda(t)).\left[\left(\widehat{f}_n - f\right)(t,\lambda(t)) + o_P\left(\left\|\widehat{f}_n - f\right\|_\infty\right)\right]\gamma(x,t)\right)dt \\
&= \sqrt{n\sigma_n}\int_0^x \left(\frac{\widehat{\varphi}_n - m.\widehat{f}_n}{f^2}\right)(t,\lambda(t)).\left(\widehat{f}_n - f\right)(t,\lambda(t))\gamma(x,t)dt \\
&\quad + \sqrt{n\sigma_n}\int_0^x \left(\frac{\widehat{\varphi}_n - m.\widehat{f}_n}{f^2}\right)(t,\lambda(t)).o_P\left(\left\|\widehat{f}_n - f\right\|_\infty\right)\gamma(x,t)dt
\end{aligned}$$

Using the results of proposition 4.2 and theorem 4.1, we can prove that both terms converge in probability to zero.

6.2.7 Proof of Theorem 4.2

The idea is to take into account the integral of $(\widehat{m}_n - m)(t, \lambda(t))$ with respect to t in order to obtain a quicker rate of convergence of the bias term.

The bias is defined as follows:

$$I\!E\left[\widehat{I}_n(x)\right] = I\!E \int\int \frac{z - m(t,\lambda(t))}{f(t,\lambda(t))}\gamma(x,t)\, 1\!I_{t\le x}\widehat{f}_{Sn}(t,\lambda(t),z)dtdz$$

where $\widehat{f}_{Sn}(t,\lambda(t),z) = \frac{1}{n\sigma_n^3}\sum_{i=1}^{n} K\left(\frac{t-X_i}{\sigma_n}\right) K\left(\frac{\lambda(t)-Y_i}{\sigma_n}\right) K\left(\frac{z-Z_i}{\sigma_n}\right)$.

Consider $(E_1, E_2, E_3) \sim h$ such that:

$$\begin{cases} E_1 &= X \\ E_2 &= \lambda(X) - Y \\ E_3 &= Z \end{cases}$$

Then, by a change of variables, the expression can be replaced by:

$$I\!E\left[\widehat{I}_n(x)\right] = I\!E \int\int \frac{z - m(t,\lambda(t))}{f(t,\lambda(t))}\gamma(x,t)\, 1\!I_{t\le x}\widehat{h}_n(t,0,z)dtdz$$

with $\widehat{h}_n(t,0,z) = \frac{1}{n\sigma_n^3}\sum_{i=1}^{n} K\left(\frac{t-E_{1i}}{\sigma_n}\right) K\left(\frac{0-E_{2i}}{\sigma_n}\right) K\left(\frac{z-E_{3i}}{\sigma_n}\right)$

So:

$$\begin{aligned} I\!E\left[\widehat{I}_n(x)\right] &= \frac{1}{\sigma_n^3}\int\int_0^x \frac{z - m(t,\lambda(t))}{f(t,\lambda(t))}\gamma(x,t) K\left(\frac{t-e_1}{\sigma_n}\right) K\left(\frac{0-e_2}{\sigma_n}\right) K\left(\frac{z-e_3}{\sigma_n}\right). \\ &\quad h(e_1,e_2,e_3)\, de_1 de_2 de_3 dtdz \\ &= \int\int_0^x \frac{z - m(t,\lambda(t))}{f(t,\lambda(t))}\gamma(x,t) K(u_1) K(u_2) \\ &\quad .\left[h(t-\sigma_n u_1, -\sigma_n u_2, z) - h(t,0,z)\right] du_1 du_2 dtdz \end{aligned}$$

Recall that $\gamma(x,t) = e^{\left[\int_0^x \frac{\partial}{\partial y} m(u,\lambda(u))du - \int_0^t \frac{\partial}{\partial y} m(u,\lambda(u))du\right]} = \gamma_1(x).\gamma_2(t)$. Moreover, in order to simplify the equations, denote by $\alpha(t,z)=\frac{z-m(t,\lambda(t))}{f(t,\lambda(t))}\gamma_2(t)$. So, the relation can be rewritten:

$$\begin{aligned}
I\!E\left[\widehat{I}_n(x)\right] &= \gamma_1(x)\int K(u_1)K(u_2)\left[\int_0^x \alpha(t,z)\left[h(t-\sigma_n u_1,-\sigma_n u_2,z)-h(t,0,z)\right]dt\right]dzdu_1du_2\\
&= \underbrace{\gamma_1(x)\int K(u_1)K(u_2)\left[\int_0^x \alpha(t,z)\left[h(t-\sigma_n u_1,-\sigma_n u_2,z)-h(t,-\sigma_n u_2,z)\right]dt\right]dzdu_1du_2}_{=A_1}\\
&\quad \underbrace{+\gamma_1(x)\int K(u_2)\left[\int_0^x \alpha(t,z)\left[h(t,-\sigma_n u_2,z)-h(t,0,z)\right]dt\right]dzdu_2}_{=A_2}
\end{aligned}$$

The density h is assumed to be C^2 for the first two arguments and C^3 only for the second one (as the density f). Then use a Taylor expansion to simplify this equation. Since the kernel K is of order 3, the second term becomes:

$$A_2=\frac{\sigma_n^3}{6}\left(\int u^3K(u)du\right)\int_0^x\int\frac{z-m(t,\lambda(t))}{f(t,\lambda(t))}\gamma(x,t).\left[\frac{\partial}{\partial y^3}h(t,0,z)\right]dtdz+o\left(\sigma_n^3\right)$$

Moreover, in order to study A_1, let use an integration by parts for the integral with respect to x:

$$\begin{aligned}
I &= \int K(u_1)\int_0^x \alpha(t,z)\left[h(t-\sigma_n u_1,-\sigma_n u_2,z)-h(t,-\sigma_n u_2,z)\right]dtdu_1\\
&= \int K(u_1)\left[\alpha(t,z)\left[\underbrace{\int_0^t h(s-\sigma_n u_1,-\sigma_n u_2,z)-h(s,-\sigma_n u_2,z)ds}_{J}\right]\right]_0^x du_1\\
&\quad -\int K(u_1)\int_0^x\frac{\partial}{\partial x}\alpha(t,z)\left[\int_0^t h(s-\sigma_n u_1,-\sigma_n u_2,z)-h(s,-\sigma_n u_2,z)ds\right]dtdu_1
\end{aligned}$$

Then, using the notation $\Psi(t,-\sigma_n u_2,z)=\int_0^t h(s,-\sigma_n u_2,z)ds$, the equation is simplified:

$$J=\Psi(t-\sigma_n u_1,-\sigma_n u_2,z)-\Psi(-\sigma_n u_1,-\sigma_n u_2,z)-\Psi(t,-\sigma_n u_2,z)+\Psi(0,-\sigma_n u_2,z)$$

By assumption, Ψ is continuously differentiable of order 3 with respect to the first variable, so a Taylor expansion can be used to obtain:

$$\begin{aligned}
I &= -\frac{\sigma_n^3}{6}\int u^3K(u)\,du.\alpha(x,z)\left[\frac{\partial^3}{\partial x^3}\Psi(x,-\sigma_n u_2,z)-\frac{\partial^3}{\partial x^3}\Psi(0,-\sigma_n u_2,z)\right]\\
&\quad +\frac{\sigma_n^3}{6}\int u^3K(u)\,du\int_0^x\frac{\partial}{\partial x}\alpha(t,z)\left[\frac{\partial^3}{\partial x^3}\Psi(t,-\sigma_n u_2,z)-\frac{\partial^3}{\partial x^3}\Psi(0,-\sigma_n u_2,z)\right]dt+o\left(\sigma_n^3\right)
\end{aligned}$$

Therefore:

$$A_1 = \gamma_1(x)\frac{\sigma_n^3}{6}\int u^3 K(u)\,du.\int\left\{\begin{array}{c} -\alpha(x,z)\left[\frac{\partial^3}{\partial x^3}\Psi(x,0,z)-\frac{\partial^3}{\partial x^3}\Psi(0,0,z)\right] \\ +\int\limits_0^x \frac{\partial}{\partial x}\alpha(t,z)\left[\frac{\partial^3}{\partial x^3}\Psi(t,0,z)-\frac{\partial^3}{\partial x^3}\Psi(0,0,z)\right]dt \end{array}\right\}dz + o\left(\sigma_n^3\right)$$

and at last:

$$I\!E\left[\widehat{I}_n(x)\right]^2 = \gamma_1^2(x)\tfrac{\sigma_n^6}{36}\left(\int u^3K(u)du\right)^2$$

$$.\left[\int\left\{\begin{array}{c} \int\limits_0^x\left(\alpha(t,z).\frac{\partial}{\partial y^3}h(t,0,z)\right)dt - \alpha(x,z)\left[\frac{\partial^2}{\partial x^2}h(x,0,z)-\frac{\partial^2}{\partial x^2}h(0,0,z)\right] \\ +\int\limits_0^x\frac{\partial}{\partial x}\alpha(t,z)\left[\frac{\partial^2}{\partial x^2}h(t,0,z)-\frac{\partial^2}{\partial x^2}h(0,0,z)\right]dt \end{array}\right\}dz\right]^2 + o\left(\sigma_n^6\right)$$

that is to say

$$I\!E\left[\widehat{I}_n(x)\right]^2 = \gamma_1^2(x)\tfrac{\sigma_n^6}{36}\left(\int u^3K(u)du\right)^2$$

$$.\left[\int\left\{\begin{array}{c} \int\limits_0^x\left(\alpha(t,z).\frac{\partial}{\partial y^3}f(t,\lambda(t),z)\right)dt - \alpha(x,z)\left[\frac{\partial^2}{\partial x^2}f(x,\lambda(x),z)-\frac{\partial^2}{\partial x^2}f(0,0,z)\right] \\ +\int\limits_0^x\frac{\partial}{\partial x}\alpha(t,z)\left[\frac{\partial^2}{\partial x^2}f(x,\lambda(x),z)-\frac{\partial^2}{\partial x^2}f(0,0,z)\right]dt \end{array}\right\}dz\right]^2 + o\left(\sigma_n^6\right)$$

where

$$\left\{\begin{array}{l} \gamma(x,t) = e^{\left[\int\limits_0^x \frac{\partial}{\partial y}m(u,\lambda(u))du - \int\limits_0^t \frac{\partial}{\partial y}m(u,\lambda(u))du\right]} = \gamma_1(x).\gamma_2(t) \\ \alpha(t,z) = \frac{z-m(t,\lambda(t))}{f(t,\lambda(t))}\gamma_2(t) \end{array}\right.$$

References

[1] Ait-Sahalia, Y. (1996). Nonparametric pricing of interest rate derivative securities, *Econometrica*, 64, 527-560.

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

[3] Bosq, D. (1998). *Nonparametric statistics for stochastic processes.* Berlin: Springer-Verlag, 2nd edition.

[4] Choquet, G. (1964). *Cours d'analyse tome II Topologie.* Masson et C^{ie}.

[5] Darolles, S., Florens, J.P. and Renault, E. (2000). Nonparametric instrumental regression, *Preprint.*

[6] Fan, J. and Gijbels, I (1996). *Local polynomial modelling and its applications.* Monographs on statistics and applied probability 66, Chapman and Hall.

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

[8] Florens, J.P. and Vanhems, A. (2000). Nonparametric study of the ionosphere thickness, *Preprint.*

[9] Garza, J., Hall, P. and Ruymgaart, F.H. (1999). A new method of solving Abel type equations, *Preprint.*

[10] Green, P.J. and Silverman, B.W. (1994). *Nonparametric regression and generalized linear models*, Monographs on statistics and applied probability 58, Chapman and Hall.

[11] Groetsch, C. (1984). *The theory of Tikhonov regularization for Fredholm equations of the first kind*, Pitman, London.

[12] Hastie, T.J. and Tibshirani, R. J. (1990). *Generalized additive models*, Monographs on statistics and applied probability 43, Chapman and Hall.

[13] Hausman, J. (1981). Exact consumer's surplus and deadweight loss, *American Economic Review*, 71, 662-676.

[14] Hausman, J. (1985). The econometrics of nonlinear budget sets, *Econometrica*, 53, 1255-1282.

[15] Hausman, J. and Newey, W. K. (1995). Nonparametric estimation of exact consumers surplus and deadweight loss, *Econometrica*, 63, 1445-1476.

[16] Horowitz, J. (1998). *Semiparametric methods in Econometrics*, Lecture Notes in Statistics 131, Springer Verlag.

[17] Horowitz, J. (2001). Nonparametric estimation of a generalized additive model with an unknown link function, *Econometrica*, 69, 499-513

[18] Horowitz, J. (2001). The bootstrap in *Handbook of Econometrics*, Vol 5, 3160-3228.

[19] Ichimura, H. (1987). Estimation of single index models, *doctoral dissertation*, Massachusetts Institute of Technology.

[20] Kunze, H. E. and Vrscay, E. R. (1999). Solving inverse problems for ordinary differential equations using the Picard contraction mapping, *Inverse Problems*, 15, 745-770.

[21] Linton, O. and Nielsen, J.P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration, *Biometrika*, 82, 93-100

[22] Loubes, J.M. and Vanhems, A. (2002). Differential equations with endogeneous efects, *Preprint*

[23] Newey, W.K. (1994). Kernel estimation of partial means and a general variance estimator, *Econometric Theory*, 10, 233-253.

[24] Newey, W., Powell, J., and Vella, F (1999). Nonparametric estimation of triangular simultaneous equations models, *Econometrica*, 67, 565-604.

[25] O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems, *Statist. Sci.* 4, 502-527.

[26] Pagan, A. and Ullah, A. (1999). *Nonparametric econometrics*, Themes in modern econometrics, Cambridge university press.

[27] Robinson, P.M. (1988). Root-N-Consistent semiparametric regression, *Econometrica*, 56, 931-954.

[28] Schwartz, L. (1970). *Analyse, Topologie générale et analyse fonctionnelle*, Collection enseignement des sciences.

[29] Serfling, R.J. (1980). *Approximation theorems of mathematical statistics*, Wiley series in probability and mathematical statistics.

[30] Sibony, M. and Mardon, J.-Cl. (1984). *Approximations et equations differentielles.* Hermann.

[31] Tikhonov, A. N. and Arsenin, V. Y. (1977). *Solutions of ill-posed problems.* V.H. Winston & Sons.

[32] van der Vaart, A. W. (1998). *Asymptotic statistics*, Cambridge Series in nStatistical and probabilistic mathematics

[33] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*, Springer, New York.

[34] Vanhems, A. (2001). Random differential equations and heterogeneity between consumers, *Preprint*

[35] Van Rooij, A.O. and Ruymgaart, F.H. (1999). On inverse estimation. In: *Asymptotics, Nonparametrics, and Time Series*, pp579-613. Dekker, New York.

[36] Vartia, Y. (1984). Efficient methods of measuring welfare change and compensated income in terms of ordinary demand functions, *Econometrica*, 51, 79-98.

[37] Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*, Monographs on statistics and applied probability 60, Chapman and Hall.