Nonparametric estimation of Exact consumer surplus with endogeneity in price

Anne Vanhems*

February 7, 2009

Abstract

This paper deals with nonparametric estimation of variation of exact consumer surplus with endogenous prices. The variation of exact consumer surplus is linked with the demand function via a non linear differential equation and the demand is estimated by nonparametric instrumental regression. We analyze two inverse problems: smoothing the data set with endogenous variables and solving a differential equation depending on this data set. We provide some nonparametric estimator, present results on consistency and optimal choice of smoothing parameters, and compare the asymptotic properties to some previous works.

Keywords: Nonparametric regression, Instrumental variable, Inverse problem

JEL classifications: Primary C14; secondary C30

*Toulouse Business School, 20 Boulevard Lascrosses, BP 7010, F-31068 Toulouse, France, a.vanhems@esc-toulouse.fr

1 Introduction

In structural econometrics, interest parameters are often defined implicitly by a relation derived from the economic context and depending on the law of distribution of the data set. Such problems require to explicit the link between the parameter of interest and the law of data set and can be considered as inverse problems. Depending on the regularity properties of the relation to solve, they are either well-posed (ie there exists a unique stable solution) or ill-posed.

This work analyzes two mixed inverse problems and is motivated by a particular economic relation, the link between the variation of exact consumer surplus associated to some price variation and the observed demand function. Such a framework was studied in particular in Hausman and Newey (1995). Their objective is to measure the impact on the consumer welfare of a price change for one good. One way to proceed is to calculate the variation of exact consumer surplus, which is a monetary way of measuring the change in welfare. To do so, consider one consumer, define y his income, q the demand in good and p^1 the price of a unique good. Assume that there exists a price variation from p to p^1. The variation of exact consumer surplus for an income level y, denoted by S_y, represents the cost to pay to the consumer so that his welfare does not change for a price change (see Varian (1992)).

The link between the interest parameter S_y and the demand function q is given by the following nonlinear relation:

$$\begin{cases} S'_y\left(p\right) & = -q(p, y - S_y\left(p\right)) \\ S_y\left(p^1\right) & = 0 \end{cases} \tag{1.1}$$

The demand function q is not known and can be estimated using some econometric model. Consider (Q, P, Y) a random vector defining demand, price and income, and a sample $(Q_i, Y_i, P_i)_{i=1,\dots,n}$ of observations. The demand function q can be approximated by the function g estimated by a nonparametric regression:

$$\begin{cases} Q & = g\left(P, Y\right) + U \\ E\left(U\left|P, Y\right.\right) & = 0 \end{cases}$$

In their paper, Hausman and Newey (1995) analyze gasoline consumption using data from the U.S. Department of Energy. They estimate semiparametrically the demand function, with a nonparametric estimation of g, and a parametric part including several exogenous variables like the year of survey, the city state of the household. They assume that the identification assumption $E\left(U\left|P, Y\right.\right) = 0$ is fulfilled.

The motivation for our work derives from the endogeneity of price in the analysis of demand function. In this case, the identification condition $E\left(U\left|P, Y\right.\right) = 0$ is no more satisfied and the conditional mean does not identify the structural demand relationship. To identify our interest parameter, we introduce some random variable W, called an instrument, such that $E(U|Y,W) = 0$. The underlying function g is then defined through a second equation:

$$E(Q - g(P,Y)|Y,W) = 0 \tag{1.2}$$

Solutions of this second linear problem have been extensively studied, in parametric as well as in nonparametric settings. The analysis of endogenous regressors, and more generally of simultaneity, has a great impact in structural econometrics. Since the earliest works of Amemiya (1974) and Hansen (1982), extensions to nonparametric and semiparametric models have been considered. Identification and estimation of g have been the subject of many recent economic studies (Darolles, Florens, and Renault (2002), Newey and Powell (2003), Hall and Horowitz (2005), Gagliardini and Scaillet (2007), Blundell and Horowitz (????), Blundell, Chen, and Kristensen (2007) to name but a few). In particular, the application to Gasoline demand is studied in Blundell, Horowitz, and Parey (2008). In what follows, we use Hall and Horowitz (2005) methodology to estimate g.

Our purpose in this work is to mix both problems (1.1) and (1.2) in a nonparametric setting. We plug some nonparametric instrumental regression estimator into the differential equation and study the asymptotic properties of the associated estimated solution. We apply our procedure to the gasoline consumption database used in Hausman and Newey (1995).

The paper proceeds in the following way. In the next section, we set the notations, the main equations to solve and the link with inverse problems theory. We then present our nonparametric estimator and recall the theoretical properties of each inverse problem. In section 4, we study the asymptotic behavior of our estimator.

2 Model Specification.

In this section, we set the notations and link our model with inverse problems theory.

2.1 The linear equation model.

The objective of this part is to set the econometric model defining the demand function q. We follow the modelization of Hall and Horowitz (2005). Consider (Q,P,Y,W,U) a random vector with all scalar random variables (to fit with the empirical application). We assume that P, Y and W are supported on $[0;1]$[1]. Let (Q_i,P_i,Y_i,W_i,U_i), for $i \geq 1$, be independent and identically distributed as (Q,P,Y,W,U). P and Y are endogenous and exogenous explanatory variables, respectively. Data (Q_i,P_i,Y_i,W_i), for $1 \leq i \leq n$, are observed.

Let f_{PYW} denote the density distribution of (P,Y,W), and f_Y the density of Y. Following Hall and Horowitz (2005) notations, we define for each $y \in [0,1]$ $t_y(p_1,p_2) =$

[1] This assumption is not very restrictive since we study solutions of differential equations that are defined locally

$\int f_{PYW}(p_1, y, w) f_{PYW}(p_2, y, w)dw$ and the operator T_y on $L_2[0,1]$ by $(T_y\psi)(p,y) = \int t_y(\xi,p)\psi(\xi,y)d\xi$. The solution g of equation (1.2) satisfies:

$$T_y g(p,y) = f_Y(y) E_{W|Y}\{E(Q|Y=y,W) f_{PYW}(p,y,W)|Y=y\} \tag{2.1}$$

where $E_{W|Y}$ denotes the expectation operator with respect to the distribution of W conditional on Y. Then, for each y for which T_y^{-1} exists, it may be proved that $g(p,y) = f_Y(y) E_{W|Y}\{E(Q|Y=y,W)(T_y^{-1} f_{PYW})(p,y,W)|Y=y\}$.

2.2 The nonlinear equation model.

Our interest functional parameter S_y is solution of the differential equation (1.1) depending on m, which can be rewritten:

$$\begin{cases} S_y'(p) & = -g(p, y - S_y(p)) \\ S_y(p^1) & = 0 \end{cases} \tag{2.2}$$

or equivalently:

$$S_y(p) = \int_p^{p^1} g(t, y - S_y(t))dt \tag{2.3}$$

The function S_y is depending on g depending itself on the law of distribution of (Q,P,Y,W). These two problems (2.1) and (2.3) can be considered as particular cases of inverses problems.

2.3 Link with inverse problems theory

Studying our interest parameter S_y is equivalent to solving both inverse problems (2.1) and (2.3).

Let start with the relation (2.3). The function S is defined by an implicit nonlinear relation (there is no restrictive assumption on the form of the function g). Denote by A_y the operator defined by $A_y(g,S) = S_y^{'} + g(.,y-S_y)$. Solving (2.3) is equivalent to inverting the operator A_y under the initial condition $S_y(p^1) = 0$. Under regularity assumptions on g, following Vanhems (2006), there exists a unique solution: $S_y(p) = \Phi_y[g](p)$, where Φ_y is continuous with respect to g. This nonlinear inverse problem is well-posed and defines a unique stable solution.

The function g itself is solution of a second linear problem (2.1). As recalled in introduction, this model is the foundation of many economic studies. Solving equation (2.1) is equivalent to trying to invert the operator T_y. Even when the probability distribution of (P,Y,W) is known, the calculation of a solution g from equation (2.1) is an ill-posed inverse problem. However f_{PYW} is unknown in general and has to be estimated from an iid sample of (P,Y,W). Two steps are necessary in order to obtain an estimator of g. The first step is to stabilize equation (2.1), the second step is to solve the stabilized equation where T_y is replaced by its estimator. Under regularity assumptions on the function g and the operator T_y, there exists a unique solution g (see Hall and Horowitz (2005) or Johannes, Van Bellegem, and Vanhems (2007) for a general overview).

Remark 2.1. The best methodology would have been to try and solve both problems in one step and invert one operator instead of two. Contrary to the operator A_y which is deterministic, T_y also depends on the law of data set and has to be estimated. Therefore, it turns out to be impossible to write our model into a single inverse problem to solve. We use a methodology in two steps to study our interest parameter S_y.

In the next section, we recall the estimation procedure and theoretical properties of both functions g and S_y separately, before mixing both inverse problems.

3 Estimation and identification

In this section, we present the nonparametric methodology used as well as the issues of identification and overidentification for both inverse problems separately. We briefly recall the results in Hall and Horowitz (2005) and Vanhems (2006) in order to prove the asymptotic properties of the final estimated functional parameter S_y.

3.1 The linear inverse problem

We first consider the nonparametric instrumental regression defined in equation (2.1). It is a Fredholm equation of the first kind and generates an ill-posed inverse problem. For the purpose of estimation, we need to replace the inverse of T_y by a regularized version. Indeed, it is well-known that the ill-posedness of this equation implies that a consistent estimator of g is not found by a simple inversion of the estimated operator $\widehat{T_y}$. A modification of the inversion is always necessary and in what follows, we consider the Tikhonov regularization and replace $\widehat{T}_y^{-1}$ by $(\widehat{T}_y + aI)^{-1} = \widehat{T}_y^{+}$ where I is the identity operator and $a > 0$.

3.1.1 Estimation

Consider K a kernel function of one dimension, centered and separable, $h > 0$ the bandwidth parameter and $K_h(u) = (1/h)K(u/h)$.[2] To construct an estimator of $g(p, y)$, let $h_p, h_y > 0$

[2] Note that we could have introduced some generalized kernel function to overcome edge effects, as in Hall and Horowitz (2005). It is not necessary in our context since we intent to estimate a local solution S_y of the differential equation in the neighborhood of the initial condition.

two bandwidth parameters and define:

$$\begin{aligned}
\widehat{f}_{PYW}(p,y,w) &= \frac{1}{n}\sum_{i=1}^{n} K_{h_p}(p-P_i)K_{h_y}(y-Y_i)K_{h_p}(w-W_i),\\
\widehat{f}^{(-i)}_{PYW}(p,y,w) &= \frac{1}{(n-1)}\sum_{j=1,j\neq i}^{n} K_{h_p}(p-P_i)K_{h_y}(y-Y_i)K_{h_p}(w-W_i),\\
\widehat{t}_y(p_1,p_2) &= \int \widehat{f}_{PYW}(p_1,y,w)\widehat{f}_{PYW}(p_2,y,w)dw,\\
(\widehat{T}_y\psi)(p,y,w) &= \int \widehat{t}_y(\xi,p)\psi(\xi,y,w)d\xi.
\end{aligned}$$

The nonparametric estimator of $g(p,y)$ is defined by:

$$\widehat{g}(p,y) = \frac{1}{n}\sum_{i=1}^{n}(\widehat{T}_y^{+}\widehat{f}^{(-i)}_{PYW})(p,y,W_i)Q_iK_{h_y}(y-Y_i). \tag{3.1}$$

3.1.2 Theoretical properties

In order to derive rates of convergence for Hall and Horowitz (2005) estimator, it is necessary to impose regularity conditions on the operator T_y. Assume that for each $y \in [0,1]$, T_y is a linear compact operator and note $\{\phi_{y1}, \phi_{y2}, ...\}$ the orthonormalized sequence of eigenvectors and $\lambda_{y1} \geq \lambda_{y2} \geq ... > 0$ the respective eigenvalues of T_y. Assume that $\{\phi_{yj}\}$ forms an orthonormal basis on $L_2[0,1]$ and consider the following decompositions on this orthonormal basis:

$$\begin{cases}
t_y(p_1,p_2) &= \sum_{j=1}^{\infty}\lambda_{yj}\phi_{yj}(p_1)\phi_{yj}(p_2),\\
f_{PYW}(p,y,w) &= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} d_{yjk}\phi_{yj}(p)\phi_{yk}(w),\\
g(p,y) &= \sum_{j=1}^{\infty} b_{yj}\phi_{yj}(p).
\end{cases} \tag{3.2}$$

Under regularity conditions on the density f_{PYW} and the kernel K (f_{PYW} has r continuous derivatives and K is of order r), on the function $g(p,y)$, and on the rate of decrease of the coefficients b_{yj}, λ_{yj} and d_{yjk} depending on constants α and β, it is proved that $\widehat{g}(p,y)$ converges to $g(p,y)$ in mean square at the rate $n^{-\tau\frac{2\beta-1}{2\beta+\alpha}}$ with $\tau = \frac{2r}{2r+1}$. In particular, the constants α and β are defined such that, for all j, $|b_{yj}| \leq Cj^{-\beta}$, $j^{-\alpha} \leq C\lambda_{yj}$ and $\sum_{k\geq 1}|d_{yjk}| \leq Cj^{-\alpha/2}$, $C>0$, uniformly in $y \in [0,1]$.

3.2 The nonlinear inverse problem

Consider now the second inverse problem defined by equation (2.3).

The function $\widehat{S}_y(p)$ is defined as solution of the estimated system:

$$\begin{cases}
\widehat{S}'_y(p) &= -\widehat{g}(p, y-\widehat{S}_y(p))\\
\widehat{S}_y(p^1) &= 0
\end{cases}$$

3.2.1 Estimation

The estimated solution $\widehat{S}_y$ is approximated using numerical implementation. Various classical algorithms can be used to calculate a solution, like Euler-Cauchy algorithm, Heun's method, Runge Kutta method. Hausman and Newey (1995) use a Buerlisch-Stoer algorithm from Numerical recipes. Let briefly recall the general methodology. Consider a grid of equidistant points $p_1, ..., p_n$ where $p_{i+1} = p_i + h$ and $p_1 = p^1$. The differential equation (2.2) is transformed into a discretized version:

$$\begin{cases} \widehat{S}_{y(i+1)} & = \widehat{S}_{yi} - h\widehat{g}_h(p_i, y - \widehat{S}_{yi}) \\ \widehat{S}_{y0} & = 0. \end{cases} \tag{3.3}$$

Where $\widehat{g}_h$ is an approximation of $\widehat{g}$. In the particular case of Euler algorithm, $\widehat{g}_h = \widehat{g}$. As recalled in Vanhems (2006), numerical approximation of $\widehat{S}_y$ does not impact the theoretical properties of the estimator since they have a higher speed of convergence than nonparametric estimation methods.

3.2.2 Theoretical properties

It has been proved (see Vanhems (2006)) that under some regularity assumptions on g, following Cauchy-Lipschitz theorem, for each $y \in [0,1]$, there exists a unique solution S_y defined in a neighborhood of the initial condition $(p^1, 0)$. Again, under regularity conditions on $\widehat{g}$, following Cauchy-Lipschitz theorem, there exists a unique solution $\widehat{S}_y$ defined on a neighborhood of the initial condition $(p^1, 0)$. The stability of the inverse problem (2.3) is fulfilled if the estimator $\widehat{S}_y$ is consistent, that is if $\frac{\partial}{\partial e_2}\widehat{g}$ (i.e. the derivative of $\widehat{g}$ with respect to the second variable) converges uniformly to $\frac{\partial}{\partial e_2}g$. Under the condition that $\|\frac{\partial}{\partial e_2}\widehat{g} - \frac{\partial}{\partial e_2}g\|_\infty \to 0$, the estimator $\widehat{S}_y$ converges almost surely to S_y and the nonlinear inverse problem is well-posed. (see Vanhems (2006) for more details). In order to derive rates of convergence, we need to explicit the link between the solution S_y and the function g. The main issue of this differential inverse problem is its nonlinearity. The following preliminary result transforms the nonlinear equation into a linear problem. The methodology used is closely related to functional delta method and close to result used in Hausman and Newey (1995) and Vanhems (2006). Denote $I = [p^1 - \varepsilon_1, p^1 + \varepsilon_1]$, for $\varepsilon_1 > 0$ a closed neighborhood of p^1 and $D = \{(p, y); p \in I, |y| \le \varepsilon_2\}$, for $\varepsilon_2 > 0$.

Proposition 3.1. : *(i) Under the assumption of consistency of $\widehat{S}_y$ to S_y, it can be proved that:*

$$\forall p \in I, \widehat{S}_y(p) - S_y(p) = -\int_{p^1}^{p} \left((\widehat{g} - g)(t, y - S_y(t)).e^{\left[\int_p^t \frac{\partial}{\partial e_2} g(u, y - S_y(u)) du\right]} \right) dt + R_{1,n}(p, y)$$

where $R_{1,n}(p) = o_P(\|\widehat{g} - g\|_\infty)$ *and* $\|\widehat{g} - g\|_\infty = \sup_{(a,b)\in D} |\widehat{g}(a,b) - g(a,b)|$.

(ii) Assume moreover that g and K are at least continuously differentiable of order 2 (ie $r \geq 2$), then the previous decomposition can be transformed:

$$\forall p \in I, \widehat{S}_y(p) - S_y(p) = -\int_{p^1}^{p} \left((\widehat{g} - g)(t, y - S_y(t)).e^{\left[\int_p^t \frac{\partial}{\partial e_2} g(u, y - S_y(u)) du\right]} \right) dt + R_{2,n}(p, y)$$

where $R_{2,n}(p) = O_P\left(\|\widehat{g} - g\|^2\right)$ *and* $\|\widehat{g} - g\|^2 = \int\int \mathbf{1}_D (\widehat{g} - g)^2 (a, b)\, dadb$

Introducing this expansion enables us to transform the nonlinear problem into a linear one, up to a residual term. Hence the rate of convergence of $\widehat{S}_y(p)$ towards $S_y(p)$ can be deduced from both terms.

- the linear part. The rate of convergence of the estimated solution of the differential equation (2.2) is expected to be greater than the rate of convergence of the estimator of the function g since there is a gain in regularity. Moreover, we also expect a gain in dimension since we transform a function of two arguments into a function of one argument.

- the residual term, which is the counterpart in the Taylor expansion. This term converges to zero by definition and we will neglect it in what follows. Rather we obtain an approximation rate up to this remainder term, controlled in probability.

4 Asymptotic behavior of the estimated solution

In this section, we aim at giving the asymptotic behavior of the solution of the differential equation obtained after estimating the regression function observed in an endogenous setting. Note first that all the asymptotic results will be given using the L^2 norm which will be written $\|.\|$. The different other norms will be clearly specified.

4.1 Assumptions

Here are the assumptions required for the consistency and mean square convergence. In particular we provide rates of decay for the generalized fourier coefficients defined in equations (3.2). We also introduce the following decomposition:

$$\begin{aligned} m_y(p, t) &= 1_{[p^1, p]}(t).e^{\left[\int_p^t \frac{\partial}{\partial e_2} g(u, y - S_y(u)) du\right]} \\ &= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} c_{yjk} \phi_{yj}(p) \phi_{yk}(t) \end{aligned}$$

We then make the following assumptions, mostly adapted from Hall and Horowitz (2005) and Vanhems (2006).

- [A1] The data (Q_i, P_i, Y_i, W_i) are independent and identically distributed as (Q, P, Y, W), where P, Y, W are supported on $[0, 1]$.

- [A2] The distribution of (P, Y, W) has a density f_{PYW} with $r \geq 2$ derivatives, each derivative bounded in absolute value by $C > 0$, uniformly in p and y. The functions $E(Q^2|Y = y, W = w)$ and $E(Q^2|P = p, Y = y, W = w)$ are bounded uniformly by C.

- [A3] The constants α, β, ν satisfy $\beta > 0$, $\nu > 0$, $\beta + \nu > 1/2$, $\alpha > 1 - 2\nu$, and $(\beta + \nu) - 1/2 \leq \alpha < 2(\beta + \nu)$. Moreover, $|b_{yj}| \leq Cj^{-\beta}$, $j^{-\alpha} \leq C\lambda_{yj}$, $\sum_{k\geq1} |d_{yjk}| \leq Cj^{-\alpha/2}$ and $\sum_{k\geq1} |c_{yjk}| \leq Cj^{-\nu}$ uniformly in y, for all $j \geq 1$.

- [A4] The parameters a, h_p, h_y satisfy $a \asymp n^{-\alpha\tau/(2\beta+\alpha)}$, $h_p \asymp n^{-\gamma}$, $h_y \asymp n^{-1/(2r+1)}$ as n goes to infinity, where $\tau = 2r/(2r + 1)$.

- [A5] The kernel function K is a bounded and Lebesgue integrable function defined on $[0, 1]$. $\int K(u)du = 1$ and K is of order $r \geq 2$. Moreover, K is continuously differentiable of order r with derivatives in $L_2([0, 1])$.

- [A6] For each $y \in [0, 1]$, the function ϕ_{yj} form an orthonormal basis for $L_2[0, 1]$ and $sup_p sup_y max_j |\phi_{yj}(p)| < \infty$.

- [A7] $sup_{p,y}|\frac{\partial}{\partial e_2}\widehat{g}(p, y) - \frac{\partial}{\partial e_2}g(p, y)|$ converges in probability to 0.

Remark on assumption $[A3]$ that allows to control the regularity of T_y, g and m_y.

4.2 Consistency

Proposition 4.1. *Under assumptions $[A1] - [A7]$, our estimator $\widehat{S}_y(p)$ is consistent and converges in probability to $S_y(p)$.*

Then we can apply the result of Proposition 3.1 and write:

$$\begin{aligned}
\widehat{S}_y(p) - S_y(p) &= -\int (\widehat{g} - g)(t, y - S_y(t)).m_y(p, t)dt + R_{2,n}(p, y) \\
&= I(p, y) + R_{2,n}(p, y)
\end{aligned}$$

4.3 Asymptotic mean square properties

Theorem 4.2. *Consider assumptions $[A1] - [A7]$ and the following property:*

$$sup_{y\in[0,1]} \int E\{I(p, y)\}^2 dp \leq sup_{y\in[0,1]} \int E\{\int (\widehat{g} - g)(t, y)m_y(p, t)dt\}^2 dp \tag{4.1}$$

Then, we can prove that:

$$sup_{y\in[0,1]} E(\|I(., y)\|^2) = O(n^{-\tau\frac{2(\beta+\nu)-1}{2\beta+\alpha}}) \tag{4.2}$$

Remark 4.1. • Note that the rate of convergence depends on the parameter ν which can be interpreted as the regularity induced by solving the differential equation. It is faster than $n^{-\tau\frac{2\beta-1}{2\beta+\alpha}}$, which is the rate obtained by Hall and Horowitz (2005).

- The condition (4.1) is quite natural in economics, it means that we neglect the compensated income in the surplus equation.

A Proofs

Proof of Proposition 3.1.

Proof. This proof is directly taken from Vanhems (2006). Under the assumptions of consistency, there exists a unique solution to (2.2) $S_y(p) = \Phi_y[g](p)$. The objective is to try and characterize the functional Φ_y that is the exact dependence between S_y and m. To prove this result, it is not necessary to impose the strong assumptions required later in the paper. Indeed, consider the operator A_y defined on the following spaces:

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

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

$$C^1_{\varepsilon_2,0}(I) = \left\{v \in C_{b,0}(I), \text{ continuously differentiable and } \left\|v'\right\|_\infty < \varepsilon_2/\varepsilon_1\right\}$$

where $D = \{(u,v)\,;|x| \le \varepsilon_1, |y| \le \varepsilon_2\}$.

$\left(C^1(D), \|.\|_\infty\right)$ and $(C(I), \|.\|_\infty)$ are Banach spaces. Moreover we define the following norm:

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

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

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

So, we have: $\forall x \in I, A_y(u,v)(x) = v'(x) + u(x, y - v(x))$. Define an open subset O of $C^1(D) \times C^1_{b,0}(I)$ and $(g, S_y) \in O$. A_y is continuous on O (it is a sum of continuous applications) and $A_y(g, S_y) = 0$. Let us check the hypothesis of the implicit function theorem. A_y is in fact continuously differentiable (thanks to the same argument) so we can take its derivative with the second variable $d_2A_y(g, S_y)$. Moreover, we have:

$$\forall h \in C^1_{b,0}(I), \forall p \in I, d_2A_y(g,S_y)(h)(p) = h'(p) + \frac{\partial}{\partial e_2} g(p, y - S_y(p)).h(p)$$

We have to prove that $d_2A_y(g,S_y)$ is a bijection. Let us show first the surjectivity:

$$\forall v \in C(I), \exists ? h \in C^1_{b,0}(I); \forall p \in I, h'(p) + \frac{\partial}{\partial e_2} g(p, y - S_y(p)).h(x) = v(p)$$

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

$$\forall p \in I, h(p) = -\int_{p^1}^{p} \left(v(s).e^{\left[\int_s^p \frac{\partial}{\partial e_2} g(t, y - S_y(t))dt\right]} \right) ds$$

Therefore, $d_2A_y(g, y - S_y)$ is surjective. Let us now demonstrate the injectivity, that is

$$Ker\left(d_2A_y(g, y - S_y)\right) = \{0\}$$

We are going to solve $d_2A_y(g, y - S_y)h = 0, h \in C^1_{b,0}(I)$. We find again a linear differential equation we can solve and find:

$$\forall p \in I, h(p) = ce^{-\int_{p^1}^{p} \frac{\partial}{\partial e_2} g(t, y - S_y(t))dt} \text{ and } h(p^1) = 0$$

Therefore, we get $c = 0$. Thus, we have demonstrated that $d_2A_y(g,S_y)$ is bijective. Let us now demonstrate the bi-continuity of $d_2A_y(g,S_y)$. In the usual implicit function theorem, this assumption is not required, but here we consider infinite dimension spaces that is why we need a more general theorem with further assumptions to satisfy. The continuity of $d_2A_y(g,S_y)$ has already been proved since A_y is continuously differentiable.

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

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

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

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

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

$$\begin{aligned} dA_y(g,S_y)(h)(p) &= d_1A_y(g,S_y)dg(h)(p) + d_2A_y(g,S_y)dS_y(h)(p) \\ &= dg(h)(p, y - S_y(p)) + (dS_y(h))'(p) + \frac{\partial}{\partial e_2} g(p, y - S_y(p))dS(h)(p) \end{aligned}$$

The differential of A_y leads to a linear differential equation in $dS_y(h)$ that we can solve. Now we apply it with $dg(h) = \widehat{g} - g$ and $dS_y(h) = d\Phi_y[g](\widehat{g} - g)$ in order to find:

$$d\Phi_y[g](\widehat{g} - g)'(p) = -\frac{\partial}{\partial e_2} g(p, y - \Phi_y[g](p)).d(\widehat{g} - g)(p) - (\widehat{g} - g)(p, y - \Phi_y[g](p))$$

Solving it leads us to:

$$\begin{aligned} d\Phi_y[g](\widehat{g} - g)(p) &= -\int_{p^1}^{p} \left((\widehat{g} - g)(t, y - \Phi_y[g](t)).e^{\left[\int_p^s \frac{\partial}{\partial e_2} g(u, y - \Phi_y[g](u))du\right]} \right) dt \\ &= -\int_{p^1}^{p} \left((\widehat{g} - g)(t, y - S_y[g](t)).e^{\left[\int_p^s \frac{\partial}{\partial e_2} g(u, y - S_y[g](u))du\right]} \right) dt \\ &= -\int_{p^1}^{p} ((\widehat{g} - g)(t, y - S_y[g](t)).v(p, t))\, dt \end{aligned}$$

So the statement is proved. □

Proof of Proposition 4.1.

Proof. The proof is based on the same properties as in Vanhems (2006). To prove the consistency of $\widehat{S}_y(p)$, we need to prove there exists a unique solution to each differential system (2.2) and (**??**). Following Cauchy-Lipschitz theorem, g and $\widehat{g}$ satisfy the Lipschitz condition:

$$\begin{aligned} |g(p, y_2) - g(p, y_1)| &\leq k|y_2 - y_1|, \text{for all}(p, y_1, y_2), \\ |\widehat{g}(p, y_2) - \widehat{g}(p, y_1)| &\leq \widehat{k}|y_2 - y_1|, \text{for all}(p, y_1, y_2). \end{aligned}$$

Under the assumption that K and f_{PYW} are continuously differentiable of order $r \geq 2$, both conditions are satisfied. Moreover, to guarantee the stability of the inverse problem, we need to impose that the estimated Lipschitz factor $\widehat{k}$ converges in probability to k of or in other words that $\frac{\partial}{\partial e_2}\widehat{g}$ converges uniformly in probability to $\frac{\partial}{\partial e_2} g$. This condition is fulfilled under assumption $[A7]$. □

Proof of Theorem 4.2:

Proof. We analyze the following term: $\int_{p^1}^{p} (\widehat{g} - g)(t, y)dt$. The objective is to prove that:

$$sup_{y \in [0,1]} \int E\{\int_{p^1}^{p} (\widehat{g} - g)(t, y)dt\}^2 dp = O(n^{-\tau \frac{2(\beta+\nu)-1}{2\beta+\alpha}})$$

To prove the result, we follow the demonstration in Hall and Horowitz (2005) and define:

$$
\begin{aligned}
D_{ny}(p) &= \int_{p^1}^{p}\{\int g(x,y)f_{PYW}(x,y,w)T_y^{+}(\widehat{f}_{PYW}-f_{PYW})(t,y,w)dxdw\}dt\\
A_{n1y}(p) &= \frac{1}{n}\sum_{i=1}^{n}\int_{p^1}^{p}(T_y^{+}f_{PYW})(t,y,W_i)Q_iK_{h_y}(y-Y_i)dt,\\
A_{n2y}(p) &= \frac{1}{n}\sum_{i=1}^{n}\int_{p^1}^{p}\{T_y^{+}(\widehat{f}_{PYW}^{(-i)}-f_{PYW})\}(t,y,W_i)Q_iK_{h_y}(y-Y_i)dt-D_{ny}(p),\\
A_{n3y}(p) &= \frac{1}{n}\sum_{i=1}^{n}\int_{p^1}^{p}\{(\widehat{T}_y^{+}-T_y^{+})f_{PYW}\}(t,y,W_i)Q_iK_{h_y}(y-Y_i)dt+D_{ny}(p),\\
A_{n4y}(p) &= \frac{1}{n}\sum_{i=1}^{n}\int_{p^1}^{p}\{(\widehat{T}_y^{+}-T_y^{+})(\widehat{f}_{PYW}^{(-i)}-f_{PYW})\}(t,y,W_i)Q_iK_{h_y}(y-Y_i)dt.
\end{aligned}
$$

Then $\int_{p^1}^{p}\widehat{g}(t,y)=A_{n1y}(p)+A_{n2y}(p)+A_{n3y}(p)+A_{n4y}(p)$ and the theorem will follow if we prove that:

$$
E\|A_{n1y}-\int_{p^1}^{\cdot}g(t,y)dt\|^2 = O(n^{-\tau\frac{2(\beta+\nu)-1}{2\beta+\alpha}}), \tag{A.1}
$$

$$
E\|A_{njy}\|^2 = O(n^{-\tau\frac{2(\beta+\nu)-1}{2\beta+\alpha}}), \text{for} j=2,3,4. \tag{A.2}
$$

To derive (A.1), note that $\|EA_{n1y}(p)-\int_{p^1}^{p}g(t,y)dt\|^2 \le 2(\|I_1\|^2+\|I_2\|^2)$ with $\|I_2\|^2 = O(h_y^{2r}a^{-2})$ and

$$
I_1 = -a\sum_k\sum_j b_{yj}c_{yjk}(\lambda_j+a)^{-1}\phi_{yk}(p).
$$

Therefore,

$$
\begin{aligned}
\|I_1\|^2 &= \sum_k\left(a\sum_j b_{yj}c_{yjk}(\lambda_j+a)^{-1}\right)^2\\
&\le C^2\left(a\sum_j|b_{yj}|j^{-\nu}(\lambda_j+a)^{-1}\right)^2
\end{aligned}
$$

We then divide the series up to the sum over $j \le J \asymp a^{-1/\alpha}$ and the complementary part. Following Hall and Horowitz (2005), we bound the right-hand side by $a^2\sum_{j\le J}(b_{yj}j^{-\nu}/\lambda_j)^2+\sum_{j>J}(b_{yj}j^{-\nu})^2$. Under assumptions[$A3$] and [$A4$], we prove that:

$$
\|EA_{n1y}(p)-\int_{p^1}^{p}g(t,y)dt\|^2 = O(n^{-\tau\frac{2(\beta+\nu)-1}{2\beta+\alpha}}). \tag{A.3}
$$

Using [$A2$], we deduce that

$$
nvar\{A_{n1y}(p)\} \le const.E\left[K_{h_y}^2(y-Y)\left(\int_{p^1}^{p}T_y^{+}f_{PYW})(t,y,W)dt\right)^2\right].
$$

Then we prove, from an expansion of $T_y^+ f_{PYW}$ and $1_{[p^1,p]}$ in their generalized Fourier series, that

$$\begin{aligned} \int var\{A_{n1y}(p)\}dp &\leq const. \frac{1}{nh_y} \sum_{jklpq} \frac{d_{jk} d_{lp} c_{jq} c_{lq}}{(\lambda_j + a)(\lambda_l + a)} \\ &\leq const. \frac{1}{nh_y} \sum_j \frac{\lambda_j j^{-2\gamma}}{(\lambda_j + a)^2} \end{aligned}$$

Using the same series decomposition as previously, we prove that

$$\begin{aligned} E\|A_{n1y} - EA_{n1y}\|^2 &= \int var\{A_{n1y}(p)\}dp \\ &= O\left((nh_y)^{-1} a^{-(\alpha+1-2\gamma)/\alpha}\right) \\ &= O\left(n^{-\tau \frac{2(\beta+\nu)-1}{2\beta+\alpha}}\right) \end{aligned}$$

Result (A.1) is implied by this bound and (A.3). □

References

Amemiya, T. (1974): "Multivariate regression and simltaneous equation models when the dependent variables are truncated normal," *Econometrica*, 42, 999–1012.

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

Blundell, R., and J. Horowitz (????): "A Nonparametric Test of Exogeneity," *Rev. Econ. Stud.*, 74, 1035–1058.

Blundell, R., J. Horowitz, and M. Parey (2008): "Measuring the price responsiveness of Gasoline demand.," Discussion Paper.

Darolles, S., J.-P. Florens, and E. Renault (2002): "Nonparametric Instrumental Regression," Working Paper # 228, IDEI, Université de Toulouse I.

Gagliardini, P., and O. Scaillet (2007): "A Specification Test for Nonparametric Instrumental Variable Regression," Swiss Finance Institute Research Paper No. 07-13.

Hall, P., and J. L. Horowitz (2005): "Nonparametric methods for inference in the presence of instrumental variables," *Ann. Statist.*, 33, 2904–2929.

Hansen, L. (1982): "Large sample properties of generalized method of moment estimators," *Econometrica*, 50, 1029–1054.

Hausman, J., and W. Newey (1995): "Nonparametric estimation of exact consumer surplus and deadweight loss," *Econometrica*, 63(6), 1445–1476.

Johannes, J., S. Van Bellegem, and A. Vanhems (2007): "Projection estimation in nonparametric instrumental regression," Discussion Paper, Institut de statistique, Université catholique de Louvain.

Newey, W. K., and J. L. Powell (2003): "Instrumental variable estimation of nonparametric models," *Econometrica*, 71, 1565–1578.

Vanhems, A. (2006): "Nonparametric study of solutions of differential equations," *Econometric Theory*, 22(1), 127–157.

Varian, H. (1992): *Microeconomic Analysis*. W.W. Norton, New York.