Nonparametric estimation of Exact consumer surplus with endogeneity in price

Anne Vanhems*

March 31, 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 is 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:

$$\begin{align*}
S_y'(p) &= -q(p, y - S_y(p)) \\
S_y(p_1) &= 0
\end{align*}$$

(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 $(Q_i, Y_i, P_i)_{i=1,...,n}$ of observations. The demand function $q$ can be approximated by the function $g$ estimated by a nonparametric regression:

$$\begin{align*}
Q &= g(P, Y) + U \\
E(U | P, Y) &= 0
\end{align*}$$

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 $E(U | P, Y) = 0$ 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 $E(U | P, Y) = 0$ 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:

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

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 (2007), 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 $g$.

To summarize, the purpose of 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 $[0; 1]^3$. 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.

---

1 This assumption is not very restrictive since we study solutions of differential equations that are defined locally in a neighborhood of the initial condition $S_y(p^1) = 0$
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 \( y \in [0,1] \) \( t_y(p_1,p_2) = \int f_{PYW}(p_1,y,w)f_{PYW}(p_2,y,w)dw \) 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:

\[
(T_yg)(p,y) = f_Y(y)E_W|Y\{E(Q|Y = y,W)f_{PYW}(p,y,W)|Y = y\}
\]

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.

Consider a price value \( p^1 \in ]0,1[ \). Our interest functional parameter \( S_y \) is solution of the differential equation (1.1) depending on \( m \), which can be rewritten:

\[
\begin{cases}
    S'_y(p) = -g(p,y - S_y(p)) \\
    S_y(p_1) = 0
\end{cases}
\]

or equivalently:

\[
S_y(p) = \int_p^{p_1} g(t; y - S_y(t))dt
\]

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 \( S_y \) 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 \( A_y(g, S_y) = S'_y + g(., y - S_y) \). 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 \( \Phi_y \) 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

\footnote{We fix a price value \( p^1 \) in the interior of \([0,1[\) so that a neighborhood of \( p^1 \), as definition set of \( S_y \), can also be included in \([0,1] \).}
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 $f_{P|Y,W}$ 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 for example Hall and Horowitz (2005), Carrasco, Florens, and Renault (2008) or Johannes, Van Bellegem, and Vanhems (2009) 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) that will be necessary 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 $\hat{T}_y$. A modification of the inversion is always necessary and in what follows, we consider the Tikhonov regularization and replace $\hat{T}_y^{-1}$ by $(\hat{T}_y + aI)^{-1} = \hat{T}_y^+$ where $I$ is the identity operator and $a > 0$.

**3.1.1 Estimation**

The function $g$ is estimated using kernel estimation. Consider $K$ a kernel function of one dimension, centered and separable, $h > 0$ the bandwidth parameter and $K_h(u) = (1/h)K(u/h)$. In order to get rid of edge effects, following Hall and Horowitz (2005), we
can introduce some generalized kernel function $K_h(.,.)$ such that if $t$ is not close to either 0 or 1 then $K_h(u,t) = K_h(u)$. However, in order to simplify the formulas and notations, in what follows, we simply denote it by $K_h(u)$.

To construct an estimator of $g(p,y)$, let $h_p, h_y > 0$ two bandwidth parameters and define:

$$\hat{f}_{PYW}(p, y, w) = \frac{1}{n} \sum_{i=1}^{n} K_{h_p}(p - P_i)K_{h_y}(y - Y_i)K_{h_p}(w - W_i),$$

$$\hat{f}_{PYW}^{(r-1)}(p, y, w) = \frac{1}{(n-1)} \sum_{j=1, j\neq i}^{n} K_{h_p}(p - P_i)K_{h_y}(y - Y_i)K_{h_p}(w - W_i),$$

$$\hat{t}_y(p_1, p_2) = \int \hat{f}_{PYW}(p_1, y, w)\hat{f}_{PYW}(p_2, y, w) dw,$$

$$(\hat{T}_y\psi)(p, y) = \int \hat{t}_y(\xi, p)\psi(\xi, y) d\xi.$$ 

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

$$\hat{g}(p,y) = \frac{1}{n} \sum_{i=1}^{n} (\hat{T}_y\hat{f}_{PYW}^{(r-1)})(p, y, W_i)Q_i K_{h_p}(y - Y_i).$$

### 3.1.2 Theoretical properties

In order to derive rates of convergence for $\hat{g}(p,y)$ 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 $\{\phi_{y1}, \phi_{y2}, \ldots\}$ the orthonormalized sequence of eigenvectors and $\lambda_{y1} \geq \lambda_{y2} \geq \ldots > 0$ 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:

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

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 $b_{yj}$, $\lambda_{yj}$ and $d_{yjk}$ depending on constants $\alpha$ and $\beta$, it is proved that $\hat{g}(p,y)$ converges to $g(p,y)$ in mean square at the rate $n^{-\tau\frac{2\beta-1}{2\beta+1}}$ with $\tau = \frac{2r}{2r+1}$. In particular, the constants $\alpha$ and $\beta$ are defined such that, for all $j$, $|b_{yj}| \leq C j^{-\beta}$; $j^{-\alpha} \leq C \lambda_{yj}$ and $\sum_{k \geq 1} |d_{yjk}| \leq C j^{-\alpha/2}$, $C > 0$, 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 $\hat{S}_y(p)$ is defined as solution of the estimated system:
\[
\begin{align*}
\hat{S}_y(p) &= -\hat{g}(p, y - \hat{S}_y(p)) \\
\hat{S}_y(p^1) &= 0
\end{align*}
\]

### 3.2.1 Estimation

It has been proved (see Vanhems (2006)) that under some regularity assumptions on \( g \), following Cauchy-Lipschitz theorem, for each \( y \in [0, 1] \) \(^3\), there exists a unique solution \( S_y \) defined in a neighborhood of the initial condition \( (p^1, 0) \). Again, under regularity conditions on \( \hat{g} \), following Cauchy Lipschitz theorem, there exists a unique solution \( \hat{S}_y \) defined on a neighborhood of the initial condition \( (p^1, 0) \).

The estimated solution \( \hat{S}_y \) can be 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 discretised version where \( \hat{g}_h \) is an approximation of \( \hat{g} \):

\[
\begin{align*}
\hat{S}_{y(i+1)} &= \hat{S}_{yi} - h\hat{g}_h(p_i, y - \hat{S}_{yi}) \\
\hat{S}_{y0} &= 0.
\end{align*}
\] (3.3)

In the particular case of Euler algorithm, \( \hat{g}_h = \hat{g} \). As recalled in Vanhems (2006), numerical approximation of \( \hat{S}_y \) 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

The existence and uniqueness of both solutions \( S_y \) and \( \hat{S}_y \) is proved under the following assumptions. Let first fix an income value \( y \in [0, 1] \). Denote \( I = [p^1 - \varepsilon_1, p^1 + \varepsilon_1] \), for \( \varepsilon_1 > 0 \) a closed neighborhood of \( p^1 \), \( J = [y - \varepsilon_2, y + \varepsilon_2] \) with \( \varepsilon_2 > 0 \) and \( D_y = I \times J \). The regularity conditions required for \( g \) are the following:

- \([i]\) \( \max_{(p, \tilde{y}) \in D_y} |g(p, \tilde{y})| < \varepsilon_2/\varepsilon_1 \)
- \([ii]\) \( |g(p, y_2) - g(p, y_1)| \leq k|y_2 - y_1|, \forall (p, y_i) \in D_y \) such that \( c = k\varepsilon_1 < 1 \)

Note that condition \([i]\) is not restrictive if the parameter \( \varepsilon_1 \) is chosen small enough. Condition \([ii]\) imposed \( g \) to be continuous on \( D_y \) and to satisfy the Lipschitz condition. A sufficient condition on \( g \) to satisfy this assumption is to be continuously differentiable of order 1 on \( D_y \). In the next section, in order to derive rates of convergence for the estimated solution \( \hat{S}_y \), we impose this last stronger condition.

\(^3\)we fix \( y \) in the interior of \([0, 1]\) for convenience, to make sure that \( y - S_y(p) \) still belongs to \([0, 1]\).
We can also introduce the parameters \( \varepsilon_{1n} \) and \( \varepsilon_{2n} \), define the neighborhoods \( I_n \) and \( D_{yn} \) such that \( \hat{g} \) satisfies the two following assumptions:

1. \( \max_{(p,\tilde{y}) \in D_{yn}} |\hat{g}(p, \tilde{y})| < \varepsilon_{2n}/\varepsilon_{1n} \)
2. \( |\hat{g}(p, y_2) - \hat{g}(p, y_1)| \leq k_n|y_2 - y_1|, \forall (p, y_i) \in D_{yn} \), such that \( c_n = k_n \varepsilon_{1n} < 1 \)

In order to define both solutions \( S_y \) and \( \hat{S}_y \) on the same neighborhood \( D_y \), we need an additional assumption of convergence of the Lipschitz factor \( k_n \) to \( k \). In other words, under the condition that \( \partial_{y_2} \hat{g} \) (i.e. the derivative of \( \hat{g} \) with respect to the second variable) converges uniformly to \( \partial_{y_2} g \), both solutions can be defined on a common subset \( I \) and the inverse problem is stable and well-posed (see Vanhems (2006) for more details).

The next step to derive rates of convergence is 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 methodology used to transform the nonlinear equation into a linear problem is closely related to functional delta method and similar to in Hausman and Newey (1995) and Vanhems (2006). Then, under the assumptions of existence uniqueness and stability for \( \hat{S}_y \) and \( S_y \), it can be proved that:

\[
\forall p \in I, \hat{S}_y (p) - S_y (p) = I(p, y) + R_n(p, y)
\]

where \( R_n(p, y) = o_P(\|\hat{g} - g\|) \) and \( \|\hat{g} - g\|^2 = \int_{D_y} (\hat{g} - g)^2(a, b)dadb \). The first term \( I(p, y) \) is linear in \( \hat{g} - g \) and will be detailed in the next section. Note that all the asymptotic results will be given using the \( L_2 \) norm which will be written \( \|\cdot\| \). The different other norms will be clearly specified.

Introducing this expansion enables us to transform the nonlinear problem into a linear one, up to a residual term. Under the condition that both terms converge, our estimator is consistent. More precisely, we can analyze the behavior of each term:

- 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 set-
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 ⁴:

\[ m_y(p,t) = 1_{[p_1,p]}(t) e \left[ \int_{\mathbb{R}} \frac{\partial}{\partial e} g(u,y - S_y(u)) du \right] = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} c_{yjk} \phi_{yj}(p) \phi_{yk}(t) \]

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 \(C\) and \(E(Q^2) < +\infty\). The function \(g\) is continuously differentiable of order \(2\) on \([0, 1]\).

- **[A3]** The constants \(\alpha, \beta, \nu\) satisfy \(\beta > 1/2, \nu > 1/2, \alpha > 1\) and \(\max(\beta + \nu - 1/2; 2\nu - 1) < \alpha < \min(2\nu; 2\beta; 3 + \nu)\). Moreover, \(|b_{yj}| \leq C j^{-\beta}, j^{-\alpha} \leq C \lambda_{yj}, \sum_{k \geq 1} |d_{yjk}| \leq C j^{-\alpha/2} \) and \(\sum_{k \geq 1} |c_{yjk}| \leq C j^{-2\nu}\) uniformly in \(y\), for all \(j \geq 1\).

- **[A4]** The parameters \(a, h_p, h_y\) satisfy \(a \asymp n^{-\alpha\tau/(2\beta+\alpha)}, h \asymp n^{-1/(2r+1)}\) 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 \(\phi_{yj}\) form an orthonormal basis for \(L_2[0, 1]\) and \(\sup_p \sup_{y} \max_{j} |\phi_{yj}(p)| < \infty\).

- **[A7]** \(\forall y \in [0, 1]\), \(\sup_{D_y} |\frac{\partial}{\partial e} \tilde{g}(p, \tilde{y}) - \frac{\partial}{\partial e} g(p, \tilde{y})|\) converges in probability to 0.

Let comment all these assumptions. We need to control for the convergence of the residual term \(\tilde{g} - g\) with Hall and Horowitz (2005) assumptions ([A1] – [A6]). We also

⁴ the notation of \(m_y(p, t)\) with \(y\) as a subscript is arbitrary, in order to follow the initial notation of the operator \(T_y\). We could as well have written \(m(p, t, y)\).
need to ensure existence and uniqueness of \( S_y \) and \( \widehat{S}_y \) and the stability of the inverse problem with Vanhems (2006) assumptions ([A1] – [A2], [A5], [A7]). At last, we control for the regularization introduced by solving the differential equation with the parameter \( \nu \) in assumption [A3]. We recall at last that the notation of the kernel function \( K \) is used to simplify the expansions in the proofs but in order to prevent edge effects, a generalized kernel function has to be used.

4.2 Asymptotic mean square properties

**Theorem 4.1.** Consider assumptions [A1] – [A7]. Then we can prove the following results.

- **Existence, Uniqueness and Stability of solutions.** For each \( y \in [0, 1[ \), there exist unique solutions \( S_y \) and \( \widehat{S}_y \) defined on a neighborhood \( I \) of \( p^1 \). We define a common neighborhood \( D_y \) of the initial condition \( (p^1, 0) \).

- **Linear decomposition.** Consider \( y \in [0, 1[ \). For any \( p \in I \),

\[
\widehat{S}_y (p) - S_y (p) = - \int (\widehat{g} - g)(t, y - S_y(t)) m_y(p,t) dt + R_n(p,y) \tag{4.1}
\]

\[
= I(p,y) + R_n(p,y) \tag{4.2}
\]

with \( R_n(p,y) \) the residual term introduced in equation (3.4)

- **Convergence in mean square.** Under the additional property:

\[
\sup_{y \in [0, 1]} E\{I(p,y)\}^2 dp \leq \sup_{y \in [0, 1]} E\{\int (\widehat{g} - g)(t, y) m_y(p,t) dt\}^2 dp \tag{4.3}
\]

we can prove that:

\[
\sup_{y \in [0, 1]} E(\|I(\cdot, y)\|^2) = O(n^{-2(\beta+\nu-1)}) \tag{4.4}
\]

**Remark 4.1.**

- The linear term \( I(p,y) \) can be written using the scalar product in \( L_2[0, 1] \): \( I(p,y) = (\widehat{g} - g)(\cdot, y); m_y(p, .) \). Compared to the results in Hall and Horowitz (2005), we study the scalar product of the estimator \( \widehat{g} \) with a smooth function. Moreover solving the differential equation improves the regularity of the initial estimator \( \widehat{g} \) and the rate of convergence for \( \widehat{S}_y \) is expected to be quicker. It is faster than \( n^{-\beta/2} \), which is the rate obtained by Hall and Horowitz (2005). The rate of convergence depends on the parameter \( \nu \) which can be interpreted as the regularity induced by solving the differential equation. However, compared to the result obtained in Vanhems (2006), we are not able to identify a gain in regularity and a gain in dimension.

That may be due to the complexity of the initial estimator \( \widehat{g} \).

- The condition (4.3) is quite natural in economics, it means that we neglect the compensated income in the surplus equation. This condition is very simple to prove as
the income value $y$ is initially fixed in $]0,1[$. Since $S_y$ takes values in a neighborhood of 0 and $\hat{g} - g$ are continuous functions on $[0,1]^2$, we can conclude that $y - S_y(p)$ also varies in $]0,1[$, which proves equation (4.3).

A Proofs

• Existence, uniqueness and stability.

Proof. Existence and uniqueness of solutions $S_y$ and $\hat{S}_y$ is proved using Cauchy Lipschitz theorem, under the sufficient condition that both functions $g$ and $\hat{g}$ are continuously differentiable of order 1, which is assumed in $[A2]$ and $[A5]$. Moreover, under assumption $[A7]$ of uniform convergence, we can define a common lipschitz factor $k$ for both functions $g$ and $\hat{g}$ and common neighborhoods $I$ and $D_y$ (see Vanhems (2006), proof of Lemma 2.2 on page 150, for details).

• Linear decomposition.

Proof. This proof is directly adapted from Vanhems (2006) (proof of proposition 4.1, page 151). Under the assumptions of existence and uniqueness, for any $y \in ]0,1[$ there exists a unique solution to (2.2) $S_y(p) = \Phi_y[g](p)$. The objective is to try and characterize the functional $\Phi_y$ that is the exact dependence between $S_y$ and $g$. Consider the operator $A_y$ defined as follows:

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

where $C(I)$ is the space of continuous functions defined on $I$ and $C^1(D_y)$ the space of functions defined on $D_y$ and continuously differentiable of order 1. We consider also the space $C^1_{\varepsilon,0}(I)$ the space of continuous functions defined on $I$ and satisfying both assumptions $[i]$ and $[ii]$ of section 3.2.1. The space $C^1_{\varepsilon,0}(I)$ stands for continuously differentiable functions of order 1 belonging to $C_{\varepsilon,0}(I)$.

Note that both spaces $(C^1(D_y), \|\cdot\|)$ and $(C(I), \|\cdot\|)$ are Banach spaces. Moreover we define the following norm:

$$\|\cdot\|' = \max \left( \|v\|, \|v'\| \right)$$

on $C^1_{\varepsilon,0}(I)$. We can easily see that $\left(C^1_{\varepsilon,0}(I), \|\cdot\|' \right)$ is a Banach space. The use of such a norm allows us to have the continuity and linearity of the following function:

$$D : \begin{cases} \left(C^1_{\varepsilon,0}(I), \|\cdot\|' \right) & \to (C(I), \|\cdot\|) \\ f & \mapsto f' \end{cases}$$

10
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 \( C^1(D_y) \times C^1_{c,0}(I) \) 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 \( d_2 A_y(g, S_y) \). Moreover, we have:

\[
\forall h \in C^1_{c,0}(I), \forall p \in I, d_2 A_y(g, S_y)(h)(p) = h'(p) + \frac{\partial}{\partial e_2} g(p, y - S_y(p)).h(p)
\]

We have to prove that \( d_2 A_y(g, S_y) \) is a bijection. Let us show first the surjectivity:

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

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

\[
\forall p \in I, h(p) = -\int_{p_1}^{p} \left( v(s).e^{\int_{p_1}^{s} \frac{\partial}{\partial e_2} g(t, y - S_y(t))dt} \right) ds
\]

Therefore, \( d_2 A_y(g, y - S_y) \) is surjective. Let us now demonstrate the injectivity, that is

\[
Ker (d_2 A_y(g, y - S_y)) = \{0\}
\]

We are going to solve \( d_2 A_y(g, y - S_y)h = 0, h \in C^1_{b,0}(I) \). 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 \( d_2 A_y(g, S_y) \) is bijective. Let us now demonstrate the bi-continuity of \( d_2 A_y(g, S_y) \). 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 \( d_2 A_y(g, S_y) \) 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:
∀u ∈ U, A_y(u, v) = 0 has a unique solution in V

Let us note: v = Φ_y [u] this unique solution for u ∈ U.

Now we are going to differentiate the relation: A_y(u, Φ [u]) = 0, ∀u ∈ U and apply it in (g, S_y = Φ_y [g]). Let us first differentiate A_y: ∀h ∈ C^1(D_y) × C^1_{ε,z,0}(I),

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{∂}{∂ε}g(p, y - S_y(p))dS(h)(p)

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) = ˆg - g and dS_y(h) = dΦ_y [g] (ˆg - g) in order to find:

dΦ_y [g] (ˆg - g)'(p) = -\frac{∂}{∂ε}g(p, y - Φ_y [g] (p))d(ˆg - g)(p) - (ˆg - g) (p, y - Φ_y [g] (p))

Solving it leads us to:

\[ dΦ_y [g] (ˆg - g)(p) = - \int_{p}^{P} ((ˆg - g)(t, y - Φ_y [g] (t)).e^{\int_{p}^{t} (g(u, y - Φ_y [g] (u))du)dt} dt \]

So the statement is proved.

**Convergence in mean square.**

*Proof.* We analyze the following term: \( \int (ˆg - g)(t, y)m_y(p, t)dt \). The objective is to prove that:

\[ \sup_{y∈[0,1]} E\{ \int (ˆg - g)(t, y)m_y(p, t)dt \}^2 dp = O(n^{-\frac{2(3+α−1)}{2β+α}}) \]

The sketch of the proof is very similar to the demonstration in Hall and Horowitz (2005). We decompose the difference \( \int (ˆg - g)(t, y)m_y(p, t)dt \) into four terms and
analyze the convergence of each one. Define:

\[
D_{ny}(p) = \int \{ \int g(x, y) f_{PYW}(x, y, w) T_y^+ (\hat{f}_{PYW} - f_{PYW})(t, y, w) dx dw \} m_y(p, t) dt
\]

\[
A_{n1y}(p) = \frac{1}{n} \sum_{i=1}^{n} \int (T_y^+ f_{PYW})(t, y, W_i) Q_i K_{h_p}(y - Y_i) m_y(p, t) dt,
\]

\[
A_{n2y}(p) = \frac{1}{n} \sum_{i=1}^{n} \int \{(\hat{T}_y^+ - T_y^+) f_{PYW}\}(t, y, W_i) Q_i K_{h_p}(y - Y_i) m_y(p, t) dt - D_{ny}(p),
\]

\[
A_{n3y}(p) = \frac{1}{n} \sum_{i=1}^{n} \int \{(\hat{T}_y^+ - T_y^+) f_{PYW}\}(t, y, W_i) Q_i K_{h_p}(y - Y_i) m_y(p, t) dt + D_{ny}(p),
\]

\[
A_{n4y}(p) = \frac{1}{n} \sum_{i=1}^{n} \int \{(\hat{T}_y^+ - T_y^+) f_{PYW}\}(t, y, W_i) Q_i K_{h_p}(y - Y_i) m_y(p, t) dt.
\]

Then \(\int \hat{g}(t, y) m_y(p, t) dt = A_{n1y}(p) + A_{n2y}(p) + A_{n3y}(p) + A_{n4y}(p)\) and the theorem will follow if we prove that:

\[
E \| A_{n1y} - \int g(t, y) m_y(p, t) dt \|^2 = O(n^{-\frac{2(\lambda + \mu)}{2\beta + \alpha}}), \tag{A.1}
\]

\[
E \| A_{n2y} \|^2 = O(n^{-\frac{2(\lambda + \mu)}{2\beta + \alpha}}), \text{ for } j = 2, 3, 4. \tag{A.2}
\]

We will then carefully detail the proof for equation (A.1) and very briefly indicate the way to prove equations (A.2) following Hall and Horowitz (2005).

To derive (A.1), we first decompose the bias term.

\[
EA_{n1y}(p) - \int g(t, y) m_y(p, t) dt = I_1 + I_2
\]

With

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

\[
I_2 = O(h_y^3) \int \int (T_y^+ f_{PYW})(t, y, w) q \frac{\partial}{\partial y} f_{QY}(q, w, y) dq dw \right] m_y(p, t) dt
\]

Therefore, \(\|EA_{n1y}(p) - \int g(t, y) m_y(p, t) dt\|^2 \leq 2(\|I_1\|^2 + \|I_2\|^2)\) and

\[
\|I_1\|^2 = \sum_k \left( a \sum_j b_{yj} c_{yjk}(\lambda_j + a)^{-1} \right)^2 
\]

\[
\leq C^2 \left( a \sum_j |b_{yj}| j^{-2\beta}(\lambda_j + a)^{-1} \right)^2
\]
Using Cauchy-Schwartz inequality, we get:

\[
\|I_1\|^2 \leq C^2 a^2 \left( \sum_j j^{-2\nu} \right) \left( \sum_j |b_{yj}|^2 j^{-2\nu} (\lambda_j + a)^{-2} \right) \\
\leq \text{const.} a^2 \left( \sum_j |b_{yj}|^2 j^{-2\nu} (\lambda_j + a)^{-2} \right)
\]

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

\[
\|I_1\|^2 = O(n^{-\frac{2(\beta+\nu)-1}{2\beta+\alpha}}).
\] (A.3)

Consider now the second term \(I_2\) the statistical bias. We have:

\[
I_2 \leq \text{const.} h_y^2 \left( E_{W|Y} \left( (T_y^+ f_{PYW})(., y, W)|Y = y\right); m_y(p, .) \right) \\
\leq \text{const.} h_y^2 \sum_{j,k,l} d_{yjk} c_{ylj} \lambda_{yj} + a \phi_y(p)
\]

Therefore, we get:

\[
\|I_2\|^2 \leq \text{const.} h_y^{2r} \sum_j \left( \sum_k d_{yjk} c_{glj} / \lambda_{yj} + a \right)^2 \\
\leq \text{const.} h_y^{2r} \left( \sum_j j^{-2\nu-o/2} / \lambda_{yj} + a \right)^2
\]

Again, we can use Cauchy-Schwartz inequality and divide the series up to the sum over \( J \) and the complementary part to get:

\[
\|I_2\|^2 \leq \text{const.} h_y^{2r} a^{2\nu-o-1} / a \\
= O(n^{-\frac{\beta+\nu}{2\beta+\alpha}}).
\]

and

\[
\|E A_{n_1y}(p) - \int g(t, y)m_y(p, t)dt\|^2 = O(n^{-\frac{\beta+\nu}{2\beta+\alpha}}).
\] (A.4)

Consider now the variance term. Using [A2], we deduce that

\[
\text{var}\{A_{n_1y}(p)\} \leq \text{const.} E_{W|Y} \left[ \left( \int T_y^+ f_{PYW}(t, y, W)m_y(p, t)dt \right)^2 \right].
\]
Then we prove, from an expansion of $T_y^+ f_{PW}^+$ and $m_y(p,.)$ in their generalized Fourier series, that

$$\int \text{var}\{A_{n1y}(p)\} dp \leq \text{const.} \frac{1}{nh_y} \sum_{jkl} \frac{d_{jkl} c_{yj} c_{yl}}{(\lambda_{yj} + a)(\lambda_{yl} + a)}$$

$$\leq \text{const.} \frac{1}{nh_y} \sum_{t} \left( \sum_{j} \sqrt{\lambda_{yj} c_{ylj}} \right)^2$$

$$\leq \text{const.} \frac{1}{nh_y} \left( \sum_{j} \sqrt{\lambda_{yj} n^{j-2\nu}} \right)^2$$

Using again Cauchy-Schwartz and the series decomposition as previously, we prove that:

$$E \| A_{n1y} - EA_{n1y} \|^2 = \int \text{var}\{A_{n1y}(p)\} dp$$

$$= O \left( (nh_y)^{-1} a^{-\nu_1 + 2\nu_2 - 2\nu_1} \right)$$

$$= O \left( n^{-\frac{2(\beta + \nu_1) - 1}{\beta + \alpha}} \right)$$

Result (A.1) is implied by this bound and (A.4).

We now present briefly how to handle with the other terms in (A.2). Start with $j = 2$. We introduce the additional notations:

$$D_{nyi}(p) = \int \{ \int g(x,y) f_{PW}^+(x,y,w) T_y^+(f_{PW}^+(x,y,w) - f_{PW}^+)(t,y,w) dx dw \} m_y(p,t) dt$$

$$A_{n2y1}(p) = \frac{1}{n} \sum_{i=1}^{n} \left( \sum_{i} (D_{nyi}(p) - D_{nyi}) \right),$$

$$A_{n2y2}(p) = A_{n2y1}(p) + A_{n2y2}(p)$$

We then study each term $\| A_{n2y1} \|^2$ and $\| A_{n2y2} \|^2$. It may be shown by tedious calculations that $E \| A_{n2y1} \|^2 = O \left( n^{-\frac{2(\beta + \nu_1) - 1}{\beta + \alpha}} \right)$. Moreover, write $\int A_{n2y2}(p)^2 dp$ as a double series and take the expected values of the terms one by one. We can again show that $E \| A_{n2y2} \|^2 = O \left( n^{-\frac{2(\beta + \nu_1) - 1}{\beta + \alpha}} \right)$.

Next we derive (A.2) for $j = 3$. Note $\Delta = \hat{T}_y - T_y$ and consider the following decomposition $T_y^+ - T_y^+ = -(I + T_y^+ \Delta)^{-1} T_y^+ \Delta T_y^+$. We introduce the additional
notations:

\[
\begin{align*}
A_{n3y1}(p) &= - (I + T_{y}^{+} \Delta)^{-1}T_{y}^{+} \langle \Delta g(., y); m_y(p, .) \rangle \\
A_{n3y2}(p) &= - (I + T_{y}^{+} \Delta)^{-1}T_{y}^{+} \Delta \left( A_{n1y}(p) - \langle g(., y); m_y(p, .) \rangle \right) \\
A_{n3y}(p) &= A_{n3y1}(p) + A_{n3y2}(p)
\end{align*}
\]

Following Hall and Horowitz (2005) argument and using Cauchy-Schwartz inequality, it can be shown that:

\[
E \| A_{n3y2} \|^2 \leq \left( E \| (I + T_{y}^{+} \Delta)^{-1}T_{y}^{+} \Delta \|^4 E \| A_{n1y}(p) - \langle g(., y); m_y(p, .) \rangle \|^4 \right)^{1/2} = O \left( n^{-2(\beta + \nu - 1)/2} \right)
\]

The second term is again decomposed in several sub-terms, each of them being controlled in the same vein as for \( A_{n1y}(p) \). Tedious moment calculus show that

\[
E \| A_{n3y1} \|^2 = \left( n^{-2(\beta + \nu - 1)/2} \right).
\]

The last result (A.2) with \( j = 4 \) follows with the rates of \( A_{n2y} \) and \( A_{n3y} \).

\[
\square
\]

References


