The Econometrics Journal

Econometrics Journal (2010), volume **13**, pp. S80–S98.
doi: 10.1111/j.1368-423X.2010.00311.x

Non-parametric estimation of exact consumer surplus with endogeneity in price

ANNE VANHEMS[†,‡]

[†]*Toulouse Business School, 20, Boulevard Lascrosses, BP 7010, 31068 Toulouse Cedex 7, France.*

[‡]*Toulouse School of Economics, 21 Allee de Brienne, 31000 Toulouse, France.*
E-mail: a.vanhems@esc-toulouse.fr

First version received: March 2009; final version accepted: January 2010

Summary This paper analyses a structural microeconomic relation describing the exact consumer surplus in a non-parametric setting with endogenous prices. The exact consumer surplus can be characterized as the solution of a differential equation involving the observed demand function. The strategy put forward in this paper involves two steps: first, estimate the demand function with endogeneity using non-parametric IV, second, plug this estimator into the differential equation to estimate the exact consumer surplus. The rate of convergence for this estimator is derived and is shown to be faster than the rate for the underlying non-parametric IV regression estimator. Solving the differential equation smooths the demand estimator and leads to a faster rate of convergence. The implementation of the methodology is illustrated through a simulation study.

Keywords: *Exact consumer surplus*, *Inverse problem*, *Non-parametric instrumental regression*.

1. INTRODUCTION

This paper addresses the issue of evaluating exact consumer surplus in a non-parametric setting. Consumer surplus is a widely used tool in microeconomics and can be interpreted as a monetary measure of the impact on consumer welfare of a change in the price of a good. It defines what income would be necessary for the consumer to maintain his utility level constant for this price change (Varian, 1992). This quantity was introduced by John Hicks (see Hicks, 1956) and depends on the Hicksian unobserved demand function. Although it could be roughly approximated by integrating the Marshallian observed demand function (Willig, 1976), Hausman (1981) shows that we can derive a measure of the exact consumer's surplus from the observed demand curve without involving any approximation.

Consider one consumer, define by 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 exact consumer surplus associated with an income level y and denoted by S_y is characterized by the following relation:

$$\begin{cases} S_y'(p) = -q(p, y - S_y(p)), \\ S_y(p^1) = 0. \end{cases} \tag{1.1}$$

The link between S_y and q is given by this non-linear ordinary differential equation of order 1. The initial condition $S_y(p^1) = 0$ means that with no variation of price the exact consumer surplus is equal to 0. The approach classically used to estimate and analyse the function S_y involves two steps: first, the estimation of the demand function q; second, the resolution of the differential equation.

To be more precise, consider (Q, P, Y) a continuously distributed random vector defining demand, price and income, and a sample $(Q_i, Y_i, P_i)_{i=1,\ldots,n}$ of observations. The demand function q can be approximated by the function g defined as the regression of Q given P and Y:

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

Hausman and Newey (1995) propose a semi-parametric estimation of the demand function with a non-parametric estimation of g, and an additive parametric part including several exogenous variables such as the year of survey and the city/state of the household. They assume that the identification assumption $E(U|P, Y) = 0$ is satisfied. In a second step, they plug this demand estimator into the differential equation and solve it numerically. Finally, they analyse its statistical properties (see also Vanhems, 2006, for the asymptotic properties).

Our work extends this setting by relaxing the exogeneity assumption $E(U|P, Y) = 0$ and considering the case where price can be an endogenous variable. Endogeneity issues occur frequently in economics, for example if an additional variable causes both independent and dependent variables and is not included in the regression model. Consider the example of hourly individual wages explained by the level of education (this example is quoted from Angrist and Krueger, 2001, Hall and Horowitz, 2005). The error term U may include personal unobserved characteristics such as individual ability, that would influence both level of education and wage. Another classical example is given by the Engel curve relationship that describes the expansion path for commodity demands with respect to household budget. In this setting, the total budget variable is a choice variable in the consumer's allocation of income and acts as an endogenous regressor (see e.g. Blundell et al., 2007).

The price endogeneity issue is also raised in several research articles. Brown and Walker (1989) argue that the hypothesis of random utility maximization implies that the additive error U can depend on P (see also Lewbel, 2001, and Matzkin, 2007) and the error term is interpreted as consumer preference heterogeneity.[1] In an industrial organization framework, Berry et al. (1995) analyse demand and supply in differentiated product markets and highlight the problem involved by correlation between prices and product characteristics, some of which are observed by the consumer but not by the econometrician. They use the instrumental variables approach to estimate the demand system, and apply their techniques to the analysis of equilibrium in the US automobile industry. Yatchew and No (2001), proposing an analysis of household gasoline demand in Canada, also raise the problem of price endogeneity. In fact, they observe significant

[1] Note however that this literature mainly focuses on heteroscedasticity of U.

variations in prices within a given urban area, with a 5% higher coefficient of variation, which lead them to conclude that this heterogeneity in price variation depends on location and may affect consumers' choices. The authors suggest that one instruments the observed price variable with the average price over a relatively small geographical area, such as the average inter-city price. In general, in any equilibrium determination of market outcomes, prices and demands will be determined simultaneously.

The purpose of this work is to provide a theoretical analysis of the non-parametric exact consumer surplus estimator under the assumption of price endogeneity using an instrumental variable approach to identify the structural demand relationship. The instrumental variable approach has also been investigated in many recent econometric studies such as Darolles et al. (2002), Ai and Chen (2003), Newey and Powell (2003), Hall and Horowitz (2005), Blundell et al. (2007), Chen (2007) and Gagliardini and Scaillet (2007), to name but a few. In this paper, we apply the purely non-parametric kernel regression model used in Darolles et al. (2002) or Hall and Horowitz (2005). The regression estimators proposed in the two papers are similar and we finally adopted the methodology developed in Hall and Horowitz (2005) in order to stick to the consumer surplus illustration with one common variable Y in the regressors and in the instruments.[2]

To implement the instrumental variable approach we introduce some continuously distributed random variable W, called an instrument, such that $E(U|Y, W) = 0$. The underlying function g is then defined through a second equation:

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

As pointed out by recent econometric analysis of non-parametric instrumental regression, the study of g defined by (1.2) is a difficult ill-posed inverse problem that cannot be solved using standard tools, and equation (1.2) needs to be stabilized before estimation (see Engl et al., 2000, for a general overview of ill-posed inverse problems and regularization methods). Both steps of stabilization and estimation are discussed in detail in the body of the paper. A major property we find is that the rate of convergence of the estimated exact consumer surplus is improved, compared to the rate of estimated demand function. Solving the differential equation smooths the demand estimator and leads to a faster rate of convergence. This smoothing effect is consistent with the results obtained in the exogenous case (see Vanhems, 2006, for more details) and is completely driven by the resolution of the differential equation.

The paper proceeds in the following way. In the next section, we set out the notations, give the main equations to be solved and establish the link with inverse problems theory. We then present our non-parametric estimator and recall the theoretical properties of each inverse problem (equations (1.1) and (1.2)). In Section 4, we study the asymptotic behaviour of our estimator and conclude the analysis with some simulations.

2. MODEL SPECIFICATION

In this section, we set out the notation and link our model with inverse problem theory.

[2] Note that other identification methods could have been used such as control function approach (see e.g. Newey et al., 1999, Blundell and Powell, 2003, or Newey and Imbens, 2009, for a non-parametric setting).

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 modelling of Hall and Horowitz (2005). Consider (Q, P, Y, W, U) a continuously distributed random vector with all scalar random variables (to simplify the notations and fit with the microeconomic illustration). P and Y are endogenous and exogenous explanatory variables, respectively, and W is the instrument. 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 $1 \leq i \leq n$, be observed data independently and identically distributed as (Q, P, Y, W, U).

Let f_{PYW} denote the joint density 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:

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

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 non-linear equation model*

Consider a price value $p^1 \in]0, 1[$.[4] Our functional parameter of interest S_y is the solution of the differential equation (1.1) depending on the demand function q. When q is replaced by the approximation function g, the differential equation to solve is 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 definition of S_y involves the function g which depends on the distribution of (Q, P, Y, W). Under standard regularity assumptions on the function g, there exists a unique local solution to (2.2). The analysis of these two problems (2.1) and (2.2) is closely linked to inverse problem theory and we recall below the characteristics of each of them.

[3] This assumption is directly taken from Hall and Horowitz (2005) and is not a restrictive one as they argue in their article, p. 2908. Moreover, in our case, we are interested in solutions of differential equations which are by construction uniquely defined in a neighbourhood of the initial condition $S_y(p^1) = 0$, which will restrict the support of the functions and random variables.

[4] We fix a price value p^1 in the interior of]0, 1[so that a neighbourhood of p^1, on which S_y is defined, can also be included in [0, 1].

2.3. Link with inverse problems theory

The methodology used to study S_y is in two steps by solving successively the two equations (2.1) and (2.2). As we will see below, they have different regularity properties that impact the way to solve them and the properties of their solutions. Consider first the relation (2.2). The function S_y is defined implicitly as the solution of this non-linear differential equation, which can be considered as an inverse problem to solve. The standard issue is to check whether or not the inverse problem is well-posed, that is if there exists a unique stable solution to (2.2) (see Tikhonov and Arsenin, 1977, Kress, 1999, or Engl et al., 2000, for a general definition). This relation is characterized by the differential operator A_y defined by $A_y(g, S_y) = S_y' + g(\cdot, y - S_y)$ and solving (2.2) is equivalent to inverting this operator under the initial condition $S_y(p^1) = 0$. Although the operator is non-linear, it can be proved (see Vanhems, 2006, for details) that this inverse problem is in fact well-posed and defines a unique local stable solution. Under regularity assumptions on g (recalled in the next section), there exists a unique solution: $S_y(p) = \Phi_y[g](p)$, where Φ_y is continuous with respect to g.

Consider now the first relation (2.1). This second inverse problem, which defines implicitly the parameter of interest g, requires to invert the linear integral operator T_y. As recalled in the introduction, this model is the foundation of many studies, and it was proved (see e.g. Tikhonov and Arsenin, 1977, or Kress, 1999) that 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. In particular, the solution is not stable and a regularization step is required to solve the problem. In our case, as in the problems studied by Darolles et al. (2002), Hall and Horowitz (2005), Carrasco et al. (2007) or Johannes et al. (2010), f_{PYW} is unknown and has to be estimated from a sample of (P, Y, W). The way to proceed is the following: first, equation (2.1) is stabilized using standard regularization method (recalled in the next section); second, the operator T_y is replaced by an estimator and the estimated stabilized equation is solved. Under regularity assumptions on the function g and the operator T_y, there exists a unique regularized solution g.

The purpose of the next section is to recall separately the estimation procedure for the two equations (2.1) and (2.2) as well as the theoretical properties of their estimated solutions. Both inverses will then be mixed in Section 4.

REMARK 2.1. A potentially better way to proceed would have been to directly study the parameter of interest S_y in one step and invert one operator instead of two. However, this one step approach raises several issues. First, contrary to the operator A_y, T_y depends on the law of the data set and has to be estimated (which we do in a first step). Second, as we will see in the next section, it is possible to write an explicit solution to the linear inverse problem, whereas it turns out to be impossible for the non-linear one. Only a numerical approximation is available. Due to these two reasons, we decided it preferable not to treat our model as a single inverse problem.

3. ESTIMATION AND IDENTIFICATION

In this section, we present the non-parametric 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. Estimation of consumer demand

We first consider the non-parametric instrumental regression defined in equation (2.1) and present the methodology developed in Hall and Horowitz (2005). As recalled in the previous section, solving the relation (2.1) generates a linear ill-posed inverse problem which implies that a consistent estimator of g is not found by a simple inversion of the estimated operator $\widehat{T}_y$. For the purpose of estimation, we need to replace the inverse of T_y by a regularized version. In what follows, we use the well-known Tikhonov regularization and replace $\widehat{T}_y^{-1}$ by $(\widehat{T}_y + aI)^{-1} = \widehat{T}_y^{+}$, where I is the identity operator and $a > 0$ (see Engl et al., 2000, for an overview of the main regularization methods).

3.1.1. Estimation. The function g is estimated using kernel estimation. Consider K a kernel function of one dimension, centred 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 $K_h(\cdot, \cdot)$ such that if t is not close to either 0 or 1 then $K_h(u, t) = K_h(u)$. In what follows, in order to simplify the formulas and notations, we simply denote it by $K_h(u)$.

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

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

The non-parametric estimator of $g(p, y)$ is then defined by:

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

3.1.2. Theoretical properties. In order to derive rates of convergence for $\widehat{g}(p, y)$ it is necessary to impose regularity conditions on the operator T_y. By construction T_y is linear and we assume that for each $y \in [0, 1]$, T_y is a compact operator. Compactness is a standard and often used regularity assumption for integral operators that allow in particular to define a discrete spectrum. We denote by $\{\phi_{y1}, \phi_{y2}, \ldots\}$ the orthonormalized sequence of eigenvectors and $\lambda_{y1} \geq \lambda_{y2} \geq \cdots > 0$ 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 decay of the coefficients b_{yj}, λ_{yj} and d_{yjk} depending on constants α and β, it is proved in Hall and Horowitz (2005) 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, $|b_{yj}| \leq C j^{-\beta}$, $j^{-\alpha} \leq C \lambda_{yj}$ and $\sum_{k \geq 1} |d_{yjk}| \leq C j^{-\alpha/2}$, $C > 0$, uniformly in $y \in [0, 1]$.

3.2. *Estimation of exact consumer surplus*

Consider now the second non-linear inverse problem defined by equation (2.2).

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

$$\begin{cases} \widehat{S}'_y(p) = -\widehat{g}(p, y - \widehat{S}_y(p)), \\ \widehat{S}_y(p^1) = 0. \end{cases} \tag{3.3}$$

3.2.1. Estimation. The Cauchy–Lipschitz theorem states that under some regularity assumptions on g, for each $y \in]0, 1[$, there exists a unique solution S_y defined in a neighbourhood of the initial condition $(p^1, 0)$.[5] Again, under regularity conditions on $\widehat{g}$, following the Cauchy–Lipschitz theorem, there exists a unique solution $\widehat{S}_y$ defined on a neighbourhood of the initial condition $(p^1, 0)$.[6]

The estimated solution $\widehat{S}_y$ can be approximated using numerical implementation. Various classical algorithms can be used such as the Euler–Cauchy algorithm, Heun's method or the Runge–Kutta method (see Ascher and Petzold, 1998, or Collatz, 1960, for a general overview of these numerical methods). As an illustration, Hausman and Newey (1995) use a Buerlisch–Stoer algorithm from Numerical recipes and Vartia (1983) details the polygon method. Let us briefly recall the general methodology. Consider a grid of equidistant points $p_1, \ldots, p_n$, where $p_{i+1} = p_i + h$ and $p_1 = p^1$. The differential equation (2.2) is transformed into a discretized version where $\widehat{g}_h$ is an approximation of $\widehat{g}$:

$$\begin{cases} \widehat{S}_{y(i+1)} = \widehat{S}_{yi} - h \widehat{g}_h(p_i, y - \widehat{S}_{yi}), \\ \widehat{S}_{y0} = 0. \end{cases} \tag{3.4}$$

[5] We fix y in the interior of $[0, 1]$ for convenience, to make sure that $y - S_y(p)$ still belongs to $[0, 1]$.

[6] See e.g. Coddington and Levinson (1955) for a general presentation of the Cauchy–Lipschitz theorem and Vanhems (2006) for an application in econometrics.

In the particular case of the Euler algorithm, $\widehat{g}_h = \widehat{g}$. These numerical algorithms converge faster than the non-parametric estimators and hence the numerical approximation of $\widehat{S}_y$ does not affect the theoretical properties of the estimator (as detailed in Vanhems, 2006).

3.2.2. Theoretical properties. Existence and uniqueness of both solutions S_y and $\widehat{S}_y$ is proved under Cauchy–Lipschitz assumptions imposed on both functions g and $\widehat{g}$. Consider a fixed income value $y \in]0, 1[$. Denote $I = [p^1 - \epsilon_1, p^1 + \epsilon_1]$, for $\epsilon_1 > 0$ a closed neighbourhood of p^1, $J = [y - \epsilon_2, y + \epsilon_2]$ with $\epsilon_2 > 0$ and $D_y = I \times J$. The regularity conditions required to prove existence and uniqueness for S_y are the following:

ASSUMPTION 3.1. $\max_{(p,\widetilde{y}) \in D_y} |g(p, \widetilde{y})| < \epsilon_2/\epsilon_1$.

ASSUMPTION 3.2. $|g(p, y_2) - g(p, y_1)| \leq k|y_2 - y_1|, \forall (p, y_i) \in D_y$ *such that* $c = k\epsilon_1 < 1$.

Note that the important condition to prove existence and uniqueness of a solution for the differential equation (2.2) is the second one. Indeed, Assumption 3.1 is just imposed by the local definition of our solution on I and the Cauchy–Lipschitz theorem proves existence and uniqueness of a solution defined on I. In particular, if the function g is assumed to be continuous, this assumption is very easily checked.[7] Assumption 3.2 imposes 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 $\widehat{S}_y$, we impose this last stronger condition.

Let us turn now to the existence and uniqueness of a solution $\widehat{S}_y$. Indeed, we study the exact consumer surplus in a two-step procedure and we also have to take into account the estimated differential equation (3.3). We use again the Cauchy–Lipschitz theorem and introduce the parameters ϵ_{1n} and ϵ_{2n}, define the neighbourhoods I_n and D_{yn} such that $\widehat{g}$ satisfies the two following assumptions:

ASSUMPTION 3.1′. $\max_{(p,\widetilde{y}) \in D_{yn}} |\widehat{g}(p, \widetilde{y})| < \epsilon_{2n}/\epsilon_{1n}$.

ASSUMPTION 3.2′. $|\widehat{g}(p, y_2) - \widehat{g}(p, y_1) \leq k_n|y_2 - y_1|, \forall (p, y_i) \in D_{yn}$ *such that* $|c_n = k_n\epsilon_{1n} < 1$.

Again, in order to derive rates of convergence in the next section, we will transform these conditions into regularity conditions on the kernel function used to construct $\widehat{g}$. At last, in order to define both solutions S_y and $\widehat{S}_y$ on the same neighbourhood D_y, we need an additional assumption of convergence of the Lipschitz factor k_n to k. In other words, under the condition that $\frac{\partial}{\partial e_2}\widehat{g}$ (i.e. the derivative of $\widehat{g}$ with respect to the second variable) converges uniformly to $\frac{\partial}{\partial e_2}g$, 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 main issue of this differential inverse problem is its non-linearity and the next step to derive rates of convergence is to linearize the relation between S_y and g. The methodology used to transform the non-linear equation into a linear problem is closely related to the functional delta method and is similar to Hausman and Newey (1995) and Vanhems (2006). Then, under the assumptions of existence uniqueness and stability for $\widehat{S}_y$ and S_y, it can be proved that:

[7] From a practical point of view, it could be interesting to check if the solution can be extended to a larger interval to take into account larger price variations. Under the same assumptions, it can be proved that a unique maximal solution exists, which can be constructed by piecing together local solutions if the intersection of their definition intervals is not empty.

$$\forall p \in I, \quad \widehat{S}_y(p) - S_y(p) = I(p, y) + R_n(p, y), \tag{3.5}$$

where $R_n(p, y) = o_P(\|\widehat{g} - g\|)$ is the residual term and the counterpart in the Taylor expansion. Under assumptions on the estimated function $\widehat{g}$, this term converges to zero in probability and will be neglected in the asymptotics. The first term $I(p, y)$ is linear in $\widehat{g} - g$ and has an explicit form that 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 $\|\cdot\|$. In particular, $\|\widehat{g} - g\|^2 = \int_{D_y} (\widehat{g} - g)^2(a, b)\, da db$. If other norms are used it will be clearly specified.

4. ASYMPTOTIC BEHAVIOUR OF THE ESTIMATED SOLUTION

The objective of this section is to combine both inverse problems and derive the asymptotic behaviour of the solution of the differential equation obtained after estimating the regression function observed in an endogenous setting. We use the delta method to transform the non-linear differential equation into a linear relation, up to the residual term. We show that, under assumptions detailed below, we are able to control the residual term and derive the rate of convergence for the leading linear term.

4.1. Assumptions

In order to prove theoretical properties on the estimated exact consumer surplus $\widehat{S}_y$, we need to impose a set of regularity conditions. These assumptions are derived from the analysis of each inverse problem (estimation of consumer demand and estimation of exact consumer surplus) and are adapted from Hall and Horowitz (2005) and Vanhems (2006). The regularity conditions on g and $\widehat{g}$ discussed in Section 3.2.2 are given in Assumptions 4.1, 4.5 and 4.7. Assumption 4.1 is equivalent to equation (1.2). Assumptions 4.2, 4.3 and 4.6 imply that T_y is a compact operator; Assumption 4.4 describes the sizes of the tuning parameters. Moreover, we also introduce the generalized Fourier decomposition for the following function:

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

with specific assumptions on the rate of decay of the coefficients c_{yjk} given in Assumption 4.3.[8] All the required assumptions are summarized below:

ASSUMPTION 4.1. *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* $[0, 1]$ *and* $E(Q - g(P, Y)|W, Y) = 0$.

ASSUMPTION 4.2. *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*

[8] 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 $m(p, t, y)$.

$E(Q^2|Y=y, W=w)$ 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 1 on $[0,1]^2$.

ASSUMPTION 4.3. *The constants α, β, ν satisfy $\beta > 1/2$, $\nu > 1/2$, $\alpha > 1$ and $\max(\beta + \nu - 1/2; 2\nu - 1) < \alpha < \min(2\nu; 2\beta; \beta + \nu)$. Moreover, $|b_{yj}| \leq Cj^{-\beta}$, $j^{-\alpha} \leq C\lambda_{yj}$, $\sum_{k\geq 1} |d_{yjk}| \leq Cj^{-\alpha/2}$ and $\sum_{k\geq 1} |c_{yjk}| \leq Cj^{-2\nu}$ uniformly in y, for all $j \geq 1$.*

ASSUMPTION 4.4. *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 $\tau = 2r/(2r+1)$.*

ASSUMPTION 4.5. *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 $L_2([0,1])$.*

ASSUMPTION 4.6. *For each $y \in [0,1]$, the function ϕ_{yj} form an orthonormal basis for $L_2[0,1]$ and $\sup_p \sup_y \max_j |\phi_{yj}(p)| < \infty$.*

ASSUMPTION 4.7. $\forall y \in [0,1]$, $\sup_{D_y} |\frac{\partial}{\partial e_2}\widehat{g}(p,\tilde{y}) - \frac{\partial}{\partial e_2} g(p,\tilde{y})|$ *converges in probability to* 0.

REMARK 4.1. (i) In order to estimate the demand function g, a standard kernel function K has been introduced in Assumption 4.5. As recalled in Section 3.1.1 (see also Hall and Horowitz, 2005), in order to prevent from edge effects, a generalized kernel function or 'boundary kernel' has to be used. It corrects in particular for the bad behaviour of the non-parametric estimator around 0 or 1. However, to simplify the expansions in the proofs, we simply use the notation K. (ii) Assumption 4.3 specifies a polynomial rate of decay for the coefficients b_{yj}, c_{yjk}, d_{yjk} and λ_{yj}. However, other rates of decay could be used, such as exponential rate, which would lead to different rates of convergence for the non-parametric estimator (see Johannes et al., 2010, for a general overview).

4.2. *Theoretical properties*

Consider Assumptions 4.1 to 4.7. Then we can prove the following results.

THEOREM 4.1. *For each $y \in]0,1[$, there exist unique solutions S_y and $\widehat{S}_y$ defined on a common neighbourhood I of p^1.*

This first result proves that both solutions S_y and $\widehat{S}_y$ exist and are defined in the same neighbourhood I. It implies that the estimated solution $\widehat{S}_y$ is stable and will converge to S_y as soon as $\widehat{g}$ converges to g. In order to derive rates of convergence, we now need to linearize the differential equation.

THEOREM 4.2. *(i) Linear decomposition. Consider $y \in]0,1[$. For any $p \in I$,*

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

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

with $R_n(p, y)$ the residual term introduced in equation (3.5), which converges to zero. (ii) Convergence in mean square. Under the additional property:

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

we can prove that:

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

We give below some comments on this rate of convergence and the condition (4.3) required to derive it.

REMARK 4.1. Note first that the rate of convergence obtained here is faster than the rate given in Hall and Horowitz (2005). This finding is consistent with the conclusions in Vanhems (2006): 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 faster. Moreover, compared to the Hall and Horowitz (2005) result, an additional parameter ν appears in the rate of convergence. In fact, the linear term $I(p, y)$ can be rewritten using the scalar product in $L_2[0, 1]$: $I(p, y) = \langle(\widehat{g}-g)(\cdot, y); m_y(p, \cdot)\rangle$ and our objective is then to analyse the scalar product of the estimator $\widehat{g}$ with a smooth function (instead of the function $\widehat{g}$ itself, as in Hall and Horowitz, 2005). Our rate of convergence will depend on the smoothness of the function $m_y(p, \cdot)$ characterized by the parameter ν. This parameter captures the regularity induced by solving the differential equation. That explains why the rate of convergence of $\widehat{S}_y$ is faster than $n^{-\tau\frac{2\beta-1}{2\beta+\alpha}}$ the rate of convergence for $\widehat{g}$ obtained by Hall and Horowitz (2005).

REMARK 4.2. In order to derive the rate of convergence in Theorem 4.2, we need an additional condition, given by the inequality (4.3). This condition is not restrictive as the income value y is initially fixed in $]0, 1[$. Since S_y takes values in a neighbourhood of 0 and $\widehat{g}-g$ 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). From an economic point of view, it acts as if the compensated income were finally neglected in the surplus equation, as it is in the definition of the observed consumer surplus, when the demand function is integrated over price with fixed income.

5. SOME SIMULATIONS AND CONCLUDING REMARKS

We present a small Monte Carlo study in order to demonstrate the practical implementation of the proposed method. The function g is defined as follows: $g(p, y) = \frac{0.2y}{(p+0.1)}$. This form fits with the classical demand function derived from the Cobb–Douglas utility (up to an additive term 0.1 to ensure the function is well-defined on $[0, 1]$). For fixed values y and p^1, the differential equation can be explicitly solved and S_y is defined by:

$$S_y(p) = y\left(1-\left(\frac{p+0.1}{p^1+0.1}\right)^{0.2}\right).$$

We consider the trigonometric basis in $L_2[0,1]$, that is $\phi_1 = 1, \phi_{2j}(\cdot) = \sqrt{2}\cos(2\pi j.)$, $\phi_{2j+1}(.) = \sqrt{2}\sin(2\pi j.)$. The variables P, Y and W are uniformly distributed on $[0, 1]$ and the joint density of (P, Y, W) is defined by $f_{PYW}(p, y, w) = \sum_{j=1}^{\infty} \lambda_j \phi_j(p)\phi_j(y)\phi_j(w)$ with singular values satisfying $\lambda_1 = 1$ and $\lambda_j = j^{-1}(2\sum_{j=1}^{\infty} l^{-1})^{-1}$, $j \geq 2$. For computational purposes, the infinite series were truncated at $j = 100$. We then generate $Q = E[g(P, Y)|W] + V$, where V is distributed as Normal(0, 0.1).

To compute the exact consumer surplus, the income value is fixed and equal to 0.5 and the price reference p^1 is equal to 1. The estimated solution of the differential equation is calculated using the Euler algorithm (see Section 3.2.1).

We generate samples of size $n = 200$, and perform 500 Monte Carlo replications. The experiments are carried out in R. The kernel function is the Gaussian kernel and the values of the smoothing parameters are fixed and equal to $h = 0.5$ and $a = 0.05$.

Results are illustrated graphically in Figures 1 and 2. The figures show $g(p, 0.5)$ and $S_{0.5}(p)$ in the solid line, and Monte Carlo approximation to $E(\widehat{g}(p, 0.5))$ and $E(\widehat{S_{0.5}}(p))$ in the dotted line. Performances of both estimators are compared using the average of Monte Carlo approximations to mean squared error (MSE). The results are the following: $MSE(g) = 0.01687601$ and $MSE(S_y) = 0.0003646748$. This illustrates clearly the fact that solving the differential equation smooths the demand and improve its properties (see Section 4.2), although the smoothing parameters h and a are not chosen optimally.

To conclude, this article develops a non-parametric estimator of exact consumer surplus where price is specified to be endogenous. We combine the methodology of the non-parametric instrumental variable of Hall and Horowitz (2005) with the estimation of solution of differential equations by Vanhems (2006) in a two-step procedure: first non-parametric estimation of demand; second, non-parametric estimation of exact consumer surplus. We analyse the asymptotic property of our estimator and show that the rate derived for the estimated exact consumer surplus is faster than the rate obtained for the estimated demand (due to the resolution of the differential equation linking both functions). This result is illustrated via a small Monte Carlo simulation.

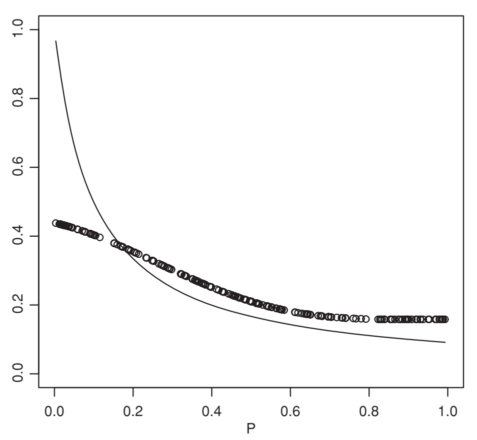

Figure 1. Graph of functions g (solid line) and $E(\widehat{g})$ (dotted line).

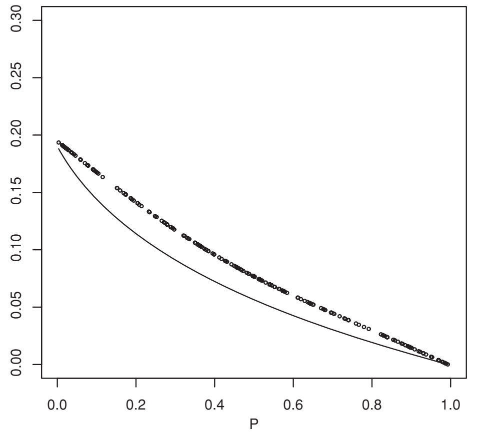

Figure 2. Graph of functions S_y (solid line) and $E(\widehat{S_y})$ (dotted line).

ACKNOWLEDGMENTS

I am grateful to Richard Blundell for stimulating conversations, suggestions and advice. I also thank Jean-Pierre Florens, Jean-Marc Robin, Jan Johannes, Stefan Hoderlein, Sebastien Van Bellegem and two anonymous referees for most helpful comments.

REFERENCES

Ai, C. and X. Chen (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica 71*, 1795–843.

Angrist, J. and A. Krueger (2001). Instrumental variables and the search for identification: from supply and demand to natural experiments. *Journal of Economic Perspectives 15*, 69–85.

Ascher, U. and L. Petzold (1998). *Computer Methods for Ordinary Differential Equations and Differential Algebraic Equations*. Philadelphia: Society for Industrial and Applied Mathematics (SIAM).

Berry, S., J. Levinsohn and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*, 841–90.

Blundell, R., X. Chen and D. Kristensen (2007). Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica 75*, 1613–69.

Blundell, R. and J. Powell (2003). Endogeneity in semiparametric and nonparametric regression models. In H. L. Dewatripont, M. and S. Turnovsky (Eds.), *Advances in Economics and Econometrics: Theory and Applications*, *Volume 43*, 111–21. Cambridge: Cambridge University Press.

Brown, B. and M. Walker (1989). The random utility hypothesis and inference in demand systems. *Econometrica 57*, 815–29.

Carrasco, M., J.-P. Florens and E. Renault (2007). Linear inverse problems in structural econometrics: estimation based on spectral decomposition and regularization. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, *Volume 6B*, 5633–751. Amsterdam: Elsevier.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, *Volume 6B*, 5548–632. Amsterdam: Elsevier.

Coddington, E. and N. Levinson (1955). *Theory of Ordinary Differential Equations*. New York: McGraw-Hill.

Collatz, L. (1960). *The Numerical Treatment of Differential Equations*. New York: Springer.

Darolles, S., J.-P. Florens and E. Renault (2002). Nonparametric instrumental regression. IDEI Working Paper No. 228, Université de Toulouse I.

Engl, H. W., M. Hanke and A. Neubauer (2000). *Regularization of Inverse Problems*. Dordrecht: Kluwer Academic.

Gagliardini, P. and O. Scaillet (2007). A specification test for nonparametric instrumental variable regression. Research Paper No. 07-13, Swiss Finance Institute.

Hall, P. and J. L. Horowitz (2005). Nonparametric methods for inference in the presence of instrumental variables. *Annals of Statistics 33*, 2904–29.

Hausman, J. (1981). Exact consumer's surplus and deadweight loss. *American Economic Review 71*, 662–76.

Hausman, J. and W. Newey (1995). Nonparametric estimation of exact consumer surplus and deadweight loss. *Econometrica 63*, 1445–76.

Hicks, J. R. (1956). *A Revision of Demand Theory*. Oxford: Clarendon Press.

Johannes, J., S. Van Bellegem and A. Vanhems (2010). Convergence rates for ill-posed inverse problems with an unknown operator. Forthcoming in *Econometric Theory*.

Kress, R. (1999). *Linear Integral Equations*, *Applied Mathematical Sciences Volume 82*, New York: Springer.

Lewbel, A. (2001). Demand systems with and without errors. *American Economic Review 91*, 611–18.

Matzkin, R. (2007). Heterogeneous choice. In N. W. Blundell, R. and T. Persson (Eds.), *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometrics Society*, *Volume 43*, 111–21. Cambridge: Cambridge University Press.

Newey, W. and G. Imbens (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica 77*, 1481–512.

Newey, W. K. and J. L. Powell (2003). Instrumental variable estimation of nonparametric models. *Econometrica 71*, 1565–78.

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

Tikhonov, A. and V. Arsenin (1977). *Solutions of Ill-Posed Problems*. Washington, DC: Winston and Sons.

Vanhems, A. (2006). Nonparametric study of solutions of differential equations. *Econometric Theory 22*, 127–57.

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

Vartia, Y. (1983). Efficient methods of measuring welfare change and compensated income in terms of ordinary demand functions. *Econometrica 51*, 79–98.

Willig, R. D. (1976). Consumer's surplus without apology. *American Economic Review 66*, 589–97.

Yatchew, A. and J. No (2001). Household gasoline demand in Canada. *Econometrica 69*, 1697–709.

APPENDIX A: PROOFS OF RESULTS

Proof of Theorem 4.1: Existence and uniqueness of solutions S_y and $\widehat{S}_y$ is proved using the Cauchy–Lipschitz theorem, under the sufficient condition that both functions g and $\widehat{g}$ are continuously differentiable of order 1, which is assumed in Assumptions 4.2 and 4.5. Moreover, under Assumption 4.7 of uniform convergence, we can define a common Lipschitz factor k for both functions g and $\widehat{g}$ and common neighbourhoods I and D_y (see Vanhems, 2006, proof of Lemma 2.2, on p. 150, for details). $\square$

Proof of Theorem 4.2: Linear decomposition. This proof is directly adapted from Vanhems (2006) (proof of Proposition 4.1, p. 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 Φ_y that is the exact dependence between S_y and g. Consider the operator A_y defined as follows:

$$A_y : \begin{cases} C^1(D_y) \times C^1_{\epsilon_2,0}(I) \to C(I), \\ \quad (u, v) \mapsto A_y(u, v), \end{cases}$$

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 $C_{\epsilon_2,0}(I)$ the space of continuous functions defined on I and satisfying both Assumptions 3.1 and 3.2 of Section 3.2.1. The space $C^1_{\epsilon_2,0}(I)$ stands for continuously differentiable functions of order 1 belonging to $C_{\epsilon_2,0}(I)$.

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

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

on $C^1_{\epsilon_2,0}(I)$. We can easily see that $(C^1_{\epsilon_2,0}(I), \|\cdot\|^{'})$ is a Banach space. The use of such a norm allows us to have the continuity and linearity of the following function:

$$D : \begin{cases} \left(C^1_{\epsilon_2,0}(I), \|\cdot\|'\right) \to (C(I), \|\cdot\|), \\ \qquad f \longmapsto f'. \end{cases}$$

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 $C^1(D_y) \times C^1_{\epsilon_2,0}(I)$ 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:

$$\forall h \in C^1_{\epsilon_2,0}(I), \quad \forall p \in I, \quad d_2A_y(g, S_y)(h)(p) = h'(p) + \frac{\partial}{\partial e_2} g(p, y - S_y(p)) \cdot h(p).$$

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

$$\forall v \in C(I), \quad \exists ?h \in C^1_{\epsilon_2,0}(I); \quad \forall p \in I, \quad h'(p) + \frac{\partial}{\partial e_2} g(p, y - S_y(p)) \cdot h(x) = v(p).$$

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

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

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

$$\mathrm{Ker}(d_2A_y(g, y - S_y)) = \{0\}.$$

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:

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

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 which 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 linearly 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:

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

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: $\forall h \in C^1(D_y) \times C^1_{\epsilon_2,0}(I)$,

$$\begin{aligned} dA_y(g, S_y)(h)(p) &= d_1 A_y(g, S_y)\, dg(h)(p) + d_2 A_y(g, S_y)\, dS_y(h)(p) \\ &= dg(h)(p, y - S_y(p)) + (dS_y(h))'(p) + \frac{\partial}{\partial e_2} g(p, y - S_y(p))\, dS(h)(p). \end{aligned}$$

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:

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

Solving it leads us to:

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

So the statement is proved. The convergence of the residual term is proved in Hall and Horowitz (2005).

Convergence in mean square. We analyse the following term: $\int (\widehat{g} - g)(t, y) m_y(p, t)\, dt$. The objective is to prove that:

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

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

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

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

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

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

We will then carefully detail the proof for equation (A.1) and very briefly indicate the way to prove equation (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

$$
\begin{aligned}
I_1 &= -a \sum_k \sum_j b_{yj} c_{yjk} (\lambda_j + a)^{-1} \phi_{yk}(p),\\
I_2 &= O\left(h_y^r\right) \int \left[\iint \left(T_y^{+} f_{PYW}\right)(t,y,w)\, q \frac{\partial}{\partial y^r} f_{QWY}(q,w,y)\,dq\,dw \right] m_y(p,t)\,dt.
\end{aligned}
$$

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

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

Using Cauchy–Schwarz inequality, we get:

$$
\begin{aligned}
\|I_1\|^2 &\le C^2 a^2 \left(\sum_j j^{-2\nu} \right) \left(\sum_j |b_{yj}|^2 j^{-2\nu} (\lambda_j + a)^{-2} \right)\\
&\le \text{const. } a^2 \left(\sum_j |b_{yj}|^2 j^{-2\nu} (\lambda_j + a)^{-2} \right),
\end{aligned}
$$

where here and below "const." denote a positive constant. We then divide the series up to the sum over $j \leq J \asymp a^{-1/\alpha}$ and the complementary part. Following Hall and Horowitz (2005), we bound the right-hand side by $a^2 \sum_{j \leq J} (b_{yj} j^{-\nu}/\lambda_j)^2 + \sum_{j>J} (b_{yj} j^{-\nu})^2$. Under Assumptions 4.3 and 4.4, we prove that:

$$\|I_1\|^2 = O\left(n^{-\tau \frac{2(\beta+\nu)-1}{2\beta+\alpha}}\right). \tag{A.3}$$

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

$$\begin{aligned} I_2 &\leq \text{const. } h_y^r \langle E_{W|Y}\left[\left(T_y^+ f_{PYW}\right)(\cdot, y, W) | Y = y\right]; m_y(p, \cdot)\rangle \\ &\leq \text{const. } h_y^r \sum_{j,k,l} \frac{d_{yjk} c_{ylj}}{\lambda_{yj} + a} \phi_{yl}(p). \end{aligned}$$

Therefore, we get:

$$\begin{aligned} \|I_2\|^2 &\leq \text{const. } h_y^{2r} \sum_l \left(\sum_{k,j} \frac{d_{yjk} c_{ylj}}{\lambda_{yj} + a} \right)^2 \\ &\leq \text{const. } h_y^{2r} \left(\sum_j \frac{j^{-2\nu-\alpha/2}}{\lambda_{yj} + a} \right)^2. \end{aligned}$$

Again, we can use Cauchy–Schwarz inequality and divide the series up to the sum over J and the complementary part to get:

$$\begin{aligned} \|I_2\|^2 &\leq \text{const. } h_y^{2r} a^{\frac{2\nu-\alpha-1}{\alpha}} \\ &= O\left(n^{-\tau \frac{2(\beta+\nu)-1}{2\beta+\alpha}}\right) \end{aligned}$$

and

$$\|EA_{n1y}(p) - \int g(t, y) m_y(p, t)\, dt\|^2 = O\left(n^{-\tau \frac{2(\beta+\nu)-1}{2\beta+\alpha}}\right). \tag{A.4}$$

Consider now the variance term. Using Assumption 4.2, we deduce that

$$nh_y \text{var}\{A_{n1y}(p)\} \leq \text{const. } E_{W|Y}\left[\left(\int (T_y^+ f_{PYW})(t, y, W) m_y(p, t)\, dt\right)^2\right].$$

Then we prove, from an expansion of $T_y^+ f_{PYW}$ and $m_y(p, \cdot)$ in their generalized Fourier series, that

$$\begin{aligned} \int \text{var}\{A_{n1y}(p)\} dp &\leq \text{const. } \frac{1}{nh_y} \sum_{jkiql} \frac{d_{yjk} d_{yiq} c_{ylj} c_{yli}}{(\lambda_{yj} + a)(\lambda_{yl} + a)} \\ &\leq \text{const. } \frac{1}{nh_y} \sum_l \left(\sum_j \frac{\sqrt{\lambda_{yj}} c_{ylj}}{\lambda_{yj} + a} \right)^2 \\ &\leq \text{const. } \frac{1}{nh_y} \left(\sum_j \frac{\sqrt{\lambda_{yj}} j^{-2\nu}}{\lambda_{yj} + a} \right)^2. \end{aligned}$$

Using again Cauchy–Schwarz and the series decomposition as previously, we prove that:

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

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

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

$$\begin{aligned} D_{nyi}(p) &= \int \left\{ \int g(x, y) f_{PYW}(x, y, w) T_y^+ \left(\widehat{f}_{PYW}^{(-i)} - f_{PYW} \right)(t, y, w)\, dx dw \right\} m_y(p, t)\, dt, \\ A_{n2y1}(p) &= \frac{1}{n} \sum_{i=1}^{n} \int \left\{ T_y^+ \left(\widehat{f}_{PYW}^{(-i)} - f_{PYW} \right) \right\}(t, y, W_i) Q_i K_{h_y}(y - Y_i) m_y(p, t)\, dt - D_{nyi}(p), \\ A_{n2y2}(p) &= \frac{1}{n} \sum_{i=1}^{n} (D_{nyi}(p) - D_{ny}(p)), \\ A_{n2y}(p) &= A_{n2y1}(p) + A_{n2y2}(p). \end{aligned}$$

We then study each term $\|A_{n2y1}\|^2$ and $\|A_{n2y2}\|^2$. It may be shown by tedious calculations that $E\|A_{n2y1}\|^2 = O(n^{-\tau \frac{2(\beta+\nu)-1}{2\beta+\alpha}})$. 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 $E\|A_{n2y2}\|^2 = O(n^{-\tau \frac{2(\beta+\nu)-1}{2\beta+\alpha}})$.

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

$$\begin{aligned} A_{n3y1}(p) &= -\left(I + T_y^+ \Delta\right)^{-1} T_y^+ \langle \Delta g(\cdot, y); m_y(p, \cdot) \rangle \\ A_{n3y2}(p) &= -\left(I + T_y^+ \Delta\right)^{-1} T_y^+ \Delta \left(A_{n1y}(p) - \langle g(\cdot, y); m_y(p, \cdot) \rangle \right) \\ A_{n3y}(p) &= A_{n3y1}(p) + A_{n3y2}(p). \end{aligned}$$

Following the Hall and Horowitz (2005) argument and using Cauchy–Schwarz inequality, it can be shown that:

$$\begin{aligned} E\|A_{n3y2}\|^2 &\le \left(E\|(I + T_y^+ \Delta)^{-1} T_y^+ \Delta\|^4 E\|A_{n1y}(p) - \langle g(\cdot, y); m_y(p, \cdot) \rangle\|^4 \right)^{1/2} \\ &= O\left(n^{-\tau \frac{2(\beta+\nu)-1}{2\beta+\alpha}}\right). \end{aligned}$$

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 $E\|A_{n3y1}\|^2 = (n^{-\tau \frac{2(\beta+\nu)-1}{2\beta+\alpha}})$.

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