

Non Parametric Estimation of Semiparametric Transformation Models*

Jean-Pierre FLORENS[†] & Senay SOKULLU[‡]

March 16, 2010

Abstract

In this paper we develop a nonparametric estimation technique for the simultaneous transformation equations systems. Identification and asymptotic properties of our model are also analysed. Our estimation method is based on writing the conditional moment conditions by their kernel counterparts so it is very easy to apply and works very well even in very nonlinear environments. Our contribution to the literature is introducing a nonparametric component to the right hand side of the transformation model and estimating the equation nonparametrically. While doing this, we solve the ill-posed inverse problem we run across, by using the Tikhonov Regularization. In the simulations we made, we saw not only that our nonparametric estimation gives very good fits, but also that the choice of regularization parameter together with the bandwidth is very important.

Keywords: Transformation Models, Nonparametric Estimation, Inverse problems, Tikhonov Regularization

JEL Classification: C13, C14, C30

*Preliminary and incomplete

[†]Toulouse School of Economics, E-mail: florens@cict.fr

[‡]Toulouse School of Economics, E-mail: Senay.Sokullu@tse-fr.eu

1 Introduction

Nonparametric econometrics is becoming more and more important. Roehrig (1988) explains the usefulness of nonparametric techniques by pointing out the needlessness of approximations about the parameters. In turns, the results that are independent of these approximations become more robust. In addition to this, transformation models that are highly used in applied econometrics and statistics has the form:

$$H(Y) = X'\beta + U \tag{1}$$

where H is an unknown function to be estimated. The examples of models like in equation 1 are given in Horowitz (1996). It includes, parametric and semiparametric proportional hazard models, log-linear regression and accelerated failure time models, the Box Cox model, etc. These models are generally parametric, and include only a parametric specification on the right hand side. However, in some cases, it may be not be possible to specify effects of some explanatory variables with a parametric form, and the economist may need to use a semiparametric model in the form:

$$Y = X'\beta + \varphi(Z) + U \tag{2}$$

When the literature is reviewed, it will not be hard to find the papers with this kind of semi parametric specification. Florens, Johannes, and Van Bellegem (2009) gives its examples where they study the estimation β in equation 2.

In this paper we introduce a system of equations where each of the equation is a semi-parametric transformation model, and estimate our equations of interest and the parameter β with the non parametric instrumental variable estimation defined in Darolles, Florens, and Renault (2009). We start with a basic system where we normalize the parameters of interest

to 1:

$$H_1(Y) = \varphi(Z) + X + U$$

$$H_2(Z) = \psi(Y) + W + V$$

Then we generalize it as:

$$H_1(Y) = \varphi(Z) + X'\beta + U$$

$$H_2(Z) = \psi(Y) + W'\gamma + V$$

We study the asymptotics of the estimators under both settings though, in the second setting the importance was given to show the \sqrt{N} convergence rate of β . The estimation of the systems like in equations above is done with limited information method, in other words, we estimate them equation by equation and leave the study of estimation with full information method for the future work. Using the assumption of conditional mean independence, we can write for the first equation of the first system:

$$\mathbb{E}[U|X, W] = 0$$

$$\mathbb{E}[H_1(Y) - \varphi(Z)|X, W] = X$$

However the solution of the above equation will give us an ill-posed inverse problem which needs to be regularized. In that step, we use Tikhonov regularization scheme as it is easier to implement than the other schemes. So, we do not only introduce a very useful nonparametric estimation method but also show that we can still get consistent results in case of solution of inverse problems with regularization.

The paper differs from the existing literature in many ways. First of all, it covers the most general case, as it considers a semiparametric transformation model. In most examples of transformation models, parametric models are used and the estimations are also done parametrically. Horowitz (1996) presents semi parametric estimation of equation 1, though

it makes a parametric specification at the right hand side. On the other hand, Feve and Florens (2009) estimate the same model with nonparametric instrumental regression. Florens, Johannes, and Van Bellegem (2009) assumes the partially linear, semiparametric model as in equation 2. So, when these papers are examined, it is seen that we are covering the most general case, as well as we are introducing a system of equations. Moreover, we are not making strong assumptions about the form of our functions, they can all be satisfied by the economic theory. For this reason, our model is very important for the estimation of structural equations. When the economic theory is examined it can be seen that the form of equations we propose to estimate is very common. In other words, simultaneous equations exist in many fields of economic theory from best-response equations of Cournot game to the one of the most recent topics, two-sided markets. In the paper, after developing the estimation theory, we give an application example where we adapt the network diffusion models to the two sided markets, derive the structural equations and explain how to estimate them with the technique we introduced. In fact, this makes another distinction of our paper from the others. We introduce the theory and we give a very detailed application example. The practical implementation of the model, as well as the choice of regularization parameter, is very well defined.

The econometric model and its estimation is supported by simulations, as well. We generated samples of different sizes and perform our estimation. We saw that, when the regularization parameter is chosen optimally, our estimated curves fits very well to the actual ones. However, in cases where we chose the regularization parameter arbitrarily, we may have very oscillating or very flat curves as the theory suggests. This result also proves the importance of selection of regularization parameter in inverse problems which can be ran across very often in nonparametric estimation.

In the following part of the paper, we will give our simultaneous equations model, derive its asymptotic properties and present our simulation results. Later, we will give an example of application of the model we have. In *Section 2* we will introduce our model and the

estimation technique while the asymptotic properties will be developed in *Section 3*. In *Section 4*, we will present our simulation method and simulation results and in *Section 5* we will give an application example from two-sided markets. Finally, in *Section 6*, we will conclude.

2 A Simultaneous Equations Model and Its Nonparametric Estimation

We have a semiparametric simultaneous equations model in which we have two endogenous variables, Y, Z and two exogenous variables X, W . These variables generate the random vector Ξ which has a cumulative distribution function F . Then for each F , we can define the subspaces of our variables as $L_F^2(Y)$, $L_F^2(Z)$, $L_F^2(W)$ and $L_F^2(X)$ which belong to common Hilbert space denoted by L_F^2 .

In the simplest case, the relationship between the variables are given by the following equations system:

$$H_1(Y) = \varphi(Z) + X + U \tag{3}$$

$$H_2(Z) = \psi(Y) + W + V \tag{4}$$

where U and V are the error terms and $H_{(i)}$ for $i = 1, 2$ is a one-to-one monotonic function.

We use the limited information method for simultaneous equations, i.e, we do an equation by equation estimation, and for the estimation we adopted the nonparametric instrumental regression of Darolles, Florens, and Renault (2009).

2.1 Identification

Our conditional independence condition can be written:

$$\mathbb{E}(U|X, W) = 0$$

$$\mathbb{E}[H_1(Y) - \varphi(Z) - X|X, W] = 0$$

Assumption 1 $\mathbb{E}[U|X, W] = 0$

Assumption 2 (Y, Z) are strongly identified by (X, W) , i.e.:

$$\forall m(Y, Z) \in L^2, \quad \mathbb{E}[m(Y, Z)|X, W] = 0 \quad \Rightarrow \quad m(Y, Z) = 0 \quad a.s$$

Assumption 3 Y and Z are measurably separable:

$$m(Y) = l(Z) \Rightarrow m(\cdot) = l(\cdot) = \text{constant}$$

Assumption 4 Normalization :

$$l(\cdot) = \text{constant} \Rightarrow \text{constant} = 0$$

For simplicity, we will assume that $\varphi(\cdot)$ is normalized by the condition $\mathbb{E}(\varphi(Z)) = 0$.

Under this assumption, we consider as parametric space, the space:

$$\mathcal{E}_0 = (H, \varphi) \in L_y^2 \times L_z^2 \quad \text{such that} \quad \mathbb{E}[\varphi] = 0 \quad (5)$$

Theorem 1 Under the assumptions 1-4, the functions $H(Y)$ and $\varphi(Z)$ are identified.

2.2 Estimation

Let us define the operator

$$T : \mathcal{E}_0 = \left\{ L_Y^2 \times \tilde{L}_Z^2 \right\} \rightarrow L_{X,W}^2: \quad T(H, \varphi) = \mathbb{E}(H(Y) - \varphi(Z)|X, W)$$

where $\tilde{L}_Z^2 = \{\varphi \in L_Z^2 / \mathbb{E}(\varphi) = 0\}$ and the inner product is defined as:

$$\langle (H_1, \varphi_1), (H_2, \varphi_2) \rangle = \langle H_1, H_2 \rangle + \langle \varphi_1, \varphi_2 \rangle$$

Then our estimation problem is given by:

$$T(H, \varphi) = r \tag{6}$$

where $r = \mathbb{E}[X|X, W]$. Equation 6 gives us an ill-posed inverse problem as the operator T has infinite dimensional range and in general it is a compact operator. So, we need to regularize our problem to get a consistent solution. For this we have chosen the Tikhonov Regularization as it is easier to work with. Basically, we will control the norm of the solution by a penalty term, α , which we will call it as *regularization parameter*. The choice of α is very important since it characterizes the balance between the fitting and the smoothing, though in the following sections we will introduce a data based selection rule for it. So solution is given by:

$$(H(Y), \varphi(Z))' = (\alpha_N I + T^*T)^{-1} T^* X \tag{7}$$

where I is the identity operator in $L_Y^2 \times L_Z^2$.

Remember that the adjoint operator of T satisfies:

$$\langle T(H, \varphi), \psi \rangle = \langle (H, \varphi), T^* \psi \rangle$$

for any $(H, \varphi) \in \mathcal{E}$ and $\psi \in L_{X,W}^2$. From this equality it follows immediately that

$$T^* \psi = (\mathbb{E}[\psi|Y], \mathbb{E}[\psi|Z])$$

However, our parametric space is \mathcal{E}_0 defined in equation 5. Let us denote the restriction of T to \mathcal{E}_0 by T_0 ($T_0 = T|_{\mathcal{E}_0}$) and the projection of \mathcal{E} under \mathcal{E}_0 by P ($P(H, \varphi) = (H, \varphi - \mathbb{E}(\varphi))$). Then we have the following lemma to characterize the adjoint operator T^* of T :

Lemma 2 *Let us define the operator $T : \mathcal{E} \rightarrow \mathcal{F}$ with the dual $T^* : \mathcal{F} \rightarrow \mathcal{E}$. Moreover, let us define $T_0 = T|_{\mathcal{E}_0}$, where $\mathcal{E}_0 \in \mathcal{E}$. Then, $T_0^* = PT^*$ where P is the projection operator on \mathcal{E}_0*

Proof. Note that, we can write:

$$x \in \mathcal{E}_0, \quad \langle Tx, y \rangle = \langle T_0x, y \rangle$$

moreover, by using the fact that, for any z_1, z_2 , $\langle z_1, z_2 \rangle = \langle Pz_1, z_2 \rangle$:

$$\langle x, T^*y \rangle = \langle x, PT^*y \rangle$$

where $PT^* \in \mathcal{E}_0$ so $PT^* = T_0^*$. ■

Then we can write the adjoint operator of T as:

$$T^* = \begin{pmatrix} E(\phi|Y) \\ PE(\phi|Z) \end{pmatrix}$$

where P is the projection of L_Z^2 on L_Z^{2*} ($P\varphi = \varphi - \mathbb{E}(\varphi)$). Although in equation 7 we have used the same α s for the regularization of different equations, they do not necessarily be the same. ¹ Then, we can solve our equation as follows:

$$(\alpha_N I + T^*T)(H_1, \varphi)' = T^*X$$

$$\begin{pmatrix} \alpha_N H_1 + E[E(H_1|X, W) - E(\varphi|X, W)|Y] \\ -\alpha_N \varphi + PE[E(H_1|X, W) - E(\varphi|X, W)|Z] \end{pmatrix} = \begin{pmatrix} E(X|Y) \\ PE(X|Z) \end{pmatrix}$$

¹Infact, in the simulations we made, we have seen that, for equal values of α s, we can not get good fits.

$$\begin{pmatrix} \alpha_N H_1 + E[E(H_1|X, W)|Y] - E[E(\varphi|X, W)|Y] \\ -\alpha_N \varphi + PE[E(H_1|X, W)|Z] - PE[E(\varphi|X, W)|Z] \end{pmatrix} = \begin{pmatrix} E(X|Y) \\ PE(X|Z) \end{pmatrix}$$

As we do not know the true distribution of our variables, we need to estimate them first. This, in turns, brings about the second source of distortion in our problem. The first one was due to the regularization parameter, α_N and the second one is coming from the bandwidths of the kernels. One thing should be noted here, as was shown in Darolles, Florens, and Renault (2009), the dimension of instruments does not have a negative effect on the speed of convergence, on the contrary, the speed of convergence increases with the dimension of instruments. So, if we have large number of instruments this will not cause a problem about the curse of dimensionality, instead it will increase the speed of convergence of our estimator. Now, let us write the above system of equations in terms of kernels to get the following system:

$$\begin{aligned} \alpha_N H_1(y) + \frac{\sum_i \frac{\sum_j H_1(y_j) K_x\left(\frac{x_i - x_j}{h_x}\right) K_w\left(\frac{w - w_j}{h_w}\right)}{\sum_j K_x\left(\frac{x_i - x_j}{h_x}\right) K_w\left(\frac{w - w_j}{h_w}\right)} K_y\left(\frac{y - y_i}{h_y}\right)}{\sum_i K_y\left(\frac{y - y_i}{h_y}\right)} - \frac{\sum_i \frac{\sum_j \varphi(z_j) K_x\left(\frac{x_i - x_j}{h_x}\right) K_w\left(\frac{w - w_j}{h_w}\right)}{\sum_j K_x\left(\frac{x_i - x_j}{h_x}\right) K_w\left(\frac{w - w_j}{h_w}\right)} K_y\left(\frac{y - y_i}{h_y}\right)}{\sum_i K_y\left(\frac{y - y_i}{h_y}\right)} \\ = \frac{\sum_i x_i K_y\left(\frac{y - y_i}{h_y}\right)}{\sum_i K_y\left(\frac{y - y_i}{h_y}\right)} \end{aligned}$$

For the second line:

$$\begin{aligned} -\alpha_N \varphi(z) - \frac{\sum_i \frac{\sum_j \varphi(z_j) K_x\left(\frac{x_i - x_j}{h_x}\right) K_w\left(\frac{w - w_j}{h_w}\right)}{\sum_j K_x\left(\frac{x_i - x_j}{h_x}\right) K_w\left(\frac{w - w_j}{h_w}\right)} K_z\left(\frac{z - z_i}{h_z}\right)}{\sum_i K_z\left(\frac{z - z_i}{h_z}\right)} + \frac{\sum_i \frac{\sum_j H_1(y_j) K_x\left(\frac{x_i - x_j}{h_x}\right) K_w\left(\frac{w - w_j}{h_w}\right)}{\sum_j K_x\left(\frac{x_i - x_j}{h_x}\right) K_w\left(\frac{w - w_j}{h_w}\right)} K_z\left(\frac{z - z_i}{h_z}\right)}{\sum_i K_z\left(\frac{z - z_i}{h_z}\right)} \\ = \frac{\sum_i x_i K_z\left(\frac{z - z_i}{h_z}\right)}{\sum_i K_z\left(\frac{z - z_i}{h_z}\right)} \end{aligned}$$

for some bandwidth parameters h_y , h_z , h_w and h_x . Note that, \hat{T}^* is not the adjoint of \hat{T} . One thing should be noted about the above equation. We wrote it as an empirical counterpart

of $-\alpha_N\varphi + E[E(H_1|X, W)|Z] - E[E(\varphi|X, W)|Z] = E(X|Z)$, which means that we did not use the projector. To recover this, after the estimation process $\hat{\varphi}$ should be transformed to its final version by the application of the estimator of the projector, $\hat{P} = \hat{\varphi} - \frac{1}{N} \sum_{i=1}^N \hat{\varphi}_i$.

Let $A_{xw}(w)$ be the matrix whose (i,j)th element is:

$$A_{xw}(w)(i, j) = \frac{K_x\left(\frac{x_i - x_j}{h_x}\right) K_w\left(\frac{w - w_j}{h_w}\right)}{\sum_j K_x\left(\frac{x_i - x_j}{h_x}\right) K_w\left(\frac{w - w_j}{h_w}\right)}$$

Moreover, A_y and A_z are the matrices with the (i,j)th elements:

$$A_y(i, j) = \frac{K_y\left(\frac{y_i - y_j}{h_y}\right)}{\sum_j K_y\left(\frac{y_i - y_j}{h_y}\right)}$$

$$A_z(i, j) = \frac{K_z\left(\frac{z_i - z_j}{h_z}\right)}{\sum_j K_z\left(\frac{z_i - z_j}{h_z}\right)}$$

Our estimated functions are the solutions of the following system:

$$\begin{pmatrix} \alpha_N \hat{H}_1 + A_y A_{xw} \hat{H}_1 - A_y A_{xw} \hat{\varphi} \\ -\alpha_N \hat{\varphi} + A_z A_{xw} \hat{H}_1 - A_z A_{xw} \hat{\varphi} \end{pmatrix} = \begin{pmatrix} A_y X \\ A_z X \end{pmatrix}$$

$$\begin{pmatrix} \hat{H}_1 \\ \hat{\varphi} \end{pmatrix} = \begin{pmatrix} \alpha_N I + A_y A_{xw} & -A_y A_{xw} \\ +A_z A_{xw} & -(\alpha_N I + A_z A_{xw}) \end{pmatrix}^{-1} \begin{pmatrix} A_y X \\ A_z X \end{pmatrix} \quad (8)$$

Equation (4) is a system of $2n$ equations in $2n$ unknowns which means that we can recover \hat{H}_1 and $\hat{\varphi}$, hence \hat{S}_1 . For the estimation of the second equation in our system the procedure is exactly the same and \hat{H}_2 and $\hat{\psi}$ are given by:

$$\begin{pmatrix} \hat{H}_2 \\ \hat{\psi} \end{pmatrix} = \begin{pmatrix} \alpha_N I + A_z A_{xw} & -A_z A_{xw} \\ A_y A_{xw} & -(\alpha_N I + A_y A_{xw}) \end{pmatrix}^{-1} \begin{pmatrix} A_z W \\ A_y W \end{pmatrix}$$

2.3 Consistency and Rate of Convergence

In our estimation process, we are estimating our functions through the estimation of the operators. For this reason, to be able to talk about the consistent estimation of the functions of interest, first we have to estimate the operators T^*T and T^*X consistently. To show this, we are going to make a set of assumptions².

Assumption 5 *Source Condition* There exists $\nu > 0$ such that:

$$\sum_{j=1}^{\infty} \frac{\langle \Phi, \phi_j \rangle^2}{\lambda_j^{2\nu}} < \infty$$

where $\Phi = (H, \varphi)$, $\{\phi_j\}$ is the system of functions of \mathcal{E} and λ_j are strictly positive singular values of T ³

This assumption is to define a regularity space for our functions. In other words, we can say that our unknown value of $\Phi_0 \in \Psi_\nu$ where

$$\Psi_\nu = \left\{ \Phi \in \mathcal{E} \text{ such that } \sum_{j=1}^{\infty} \frac{\langle \Phi, \phi_j \rangle^2}{\lambda_j^{2\nu}} < \infty \right\}$$

In fact, assuming that $\Phi_0 \in \Psi_\nu$ just adds a smoothness condition to our functional parameter of interest. As was pointed out by Carrasco, Florens, and Renault (2007), this regularity assumption will give us an advantage in calculation the rate of convergence of the regularization bias.

Assumption 6 *There exists $s \geq 2$ such that:*

- $\left\| \hat{T} - T \right\|^2 = O\left(\frac{1}{Nh_N^{p+q+1}} + h_N^{2s}\right)$
- $\left\| \hat{T}^* - T^* \right\|^2 = O\left(\frac{1}{Nh_N^{p+q+1}} + h_N^{2s}\right)$

²In this part we will present our results based on the first equation of our system, everything holds for the second equation, too

³Moreover, we can write our source condition in a more explicit way as $\sum_{j=1}^{\infty} \frac{[\langle H, \phi_{j1} \rangle + \langle \varphi, \phi_{j2} \rangle]^2}{\lambda_j^{2\nu}} < \infty$.

where s is the minimum between the order of the kernel and the order of the differentiability of f .

Assumption 7

$$\left\| X - \hat{T}\Phi \right\|^2 = O\left(\frac{1}{Nh_N^{p+q+1}} + h_N^{2s}\right)$$

Assumption 8

$$\lim_{N \rightarrow \infty} \alpha_N = 0$$

$$\lim_{N \rightarrow \infty} \frac{h_N^{2s}}{\alpha_N} = 0$$

$$\lim_{N \rightarrow \infty} \alpha_N N h_N^{p+q+1} \rightarrow \infty$$

Theorem 3 *Let us define $(H_1(Y), \varphi(z))$ as Φ . Under assumptions 5 to 8:*

- $\left\| \hat{\Phi}_N^{\alpha_N} - \Phi \right\|^2 = O\left(\frac{1}{\alpha} \left(\frac{1}{Nh_N^{p+q+1}} + h_N^{2s}\right) + \frac{1}{\alpha} \left(\frac{1}{Nh_N^{p+q+1}} + h_N^{2s}\right) (\alpha^{\min(\nu+1, 2)}) + \alpha^\beta\right)$
- $\left\| \hat{\Phi}_N^{\alpha_N} - \Phi \right\| \rightarrow 0$ in probability.

Optimal speed of convergence is obtained by the calculation of optimal α . To do this we equalize the first and the third term of the rate of convergence above, as the second term is negligible. Then the optimal bandwidth is given by:

$$h = N^{-\frac{1}{p+q+1+2s}}$$

and the speed of convergence is given by:

$$\left\| \hat{\Phi}^\alpha - \Phi \right\|^2 \sim O\left(N^{-\left(\frac{2s}{2s+p+q+1}\right)\left(\frac{\nu}{\nu+1}\right)}\right)$$

The asymptotic normality is also attained with the addition of some other assumptions. For a detailed discussion of this see Darolles, Florens, and Renault (2009) *Section 4.3* and Florens, Johannes, and Van Bellegem (2009) *Section 3.2*.

3 Semiparametric Transformation Model: The General Case

In this section, we will generalize the basic model above, and we will show that, in the semiparametric transformation models with many explanatory variables, we can get \sqrt{N} -consistency for the estimated parameters. In this general case, the new system of equations is given as:

$$H_1(Y) = \varphi(Z) + X_0 + X'\beta + U \quad (9)$$

$$H_2(Z) = \psi(Y) + W_0 + W'\gamma + V \quad (10)$$

Identification of this general model is not very different from the previous one's, nonetheless we need some additional assumptions.

Assumption 9 (Y, Z, X) are strongly identified by (X, W) .

$$\mathbb{E}[g(Y, Z, X)|X, W] = 0 \quad g(Y, Z, X) = 0 \quad a.s.$$

Remark: Assumption 9 may be weakened by considering only the function $g(Y, Z, X)$ satisfying:

$$g(Y, Z, X) = g_1(Y) + g_2(Z) + g_3(X)$$

Assumption 10 (Y, Z) and X are measurably separable:

$$m(Y, Z) = l(X) \Rightarrow m(\cdot) = l(\cdot) = constant$$

Theorem 4 Under the assumptions 1-4, 9 and 10 the functions $H(Y)$ and $\varphi(Z)$ and the parameter β are identified.

Now, we can proceed with the estimation. Let us keep the same our operator T as in the previous section, and introduce an additional operator $T_X : \mathbb{R}^k \rightarrow L_{X,W}^2 : \beta \mapsto X'\beta$.

Equivalently its adjoint is defined $T_X^* : L_{X,W}^2 \rightarrow \mathbb{R}^k : g \mapsto \mathbb{E}(Xg(X, W))$. Then we can write the following:

$$T(H, \varphi) - T_X \beta = X_0 \quad (11)$$

The normal equations are:

$$T^*T(H, \varphi) - T^*T_X \beta = T^*X_0 \quad (12)$$

$$T_X^*T(H, \varphi) - T_X^*T_X \beta = T_X^*X_0 \quad (13)$$

From equation 12, we get⁴ $(H, \varphi) = (\alpha_n I + T^*T)^{-1}(T^*T_X \beta + T^*X_0)$ and if we substitute this into equation 13, we obtain an expression for the beta:

$$\beta = (T_X^*(P^\alpha - I)T_X)^{-1}T_X^*(I - P^\alpha)X_0$$

where $P^\alpha = T(\alpha I + T^*T)^{-1}T^*$. Then the estimator is given by:

$$\hat{\beta} = \left(\widehat{T_X^*T}(\alpha I + \hat{T}^*\hat{T})^{-1}\widehat{T^*T_X} - \widehat{T_X^*T} \right)^{-1} \left(T_X - \widehat{T_X^*T}(\alpha I + \hat{T}^*\hat{T})^{-1}\widehat{T^*T_X} \right) X_0$$

We can continue with the asymptotic properties of the $\hat{\beta}$. Let us introduce some additional assumptions to get the \sqrt{N} -consistency of $\hat{\beta}$.

Assumption 11 (*Source Condition*) *There exists $\eta > 0$ such that:*

$$\max_{i=1, \dots, k} \sum_{j=1}^{\infty} \frac{\langle \tilde{\phi}_i, \phi_j \rangle^2}{\lambda_j^{2\eta}} < \infty$$

where $\{\tilde{\phi}_i\}$ is system of functions in $L_{X,W}^2$, $\{\phi_j\}$ is the system of functions of \mathcal{E} and λ_j are the strictly positive singular values of T .

This source condition explains the collinearity between (Y, Z) and (X) . In other words, as

⁴ α in this generalized version and α in the simple version need not necessarily be the same. We are using the letter α just to refer regularization parameter.

was explained in Florens, Johannes, and Van Bellegem (2009), it means that T_X is "adapted" to the operator T .

Assumption 12 *There exists $s \geq 2$ such that:*

$$\left\| \widehat{T^*T_X} - T^*T_X \right\|^2 = O\left(\frac{1}{Nh_N} + h_N^{2s}\right)$$

where s is the minimum between the order of the kernel and the order of the differentiability of f .

Assumption 13 • $\left\| \widehat{T_X^*T_X} - T_X^*T_X \right\|^2 = O\left(\frac{1}{N}\right)$

• $\left\| \widehat{T_X^*T} - T_X^*T \right\|^2 = O\left(\frac{1}{N}\right)$

Assumption 14

$$\lim_{N \rightarrow \infty} \alpha_N^{\eta \wedge 2} h^{2s} = 0$$

$$\lim_{N \rightarrow \infty} \frac{\alpha_N^{\eta \wedge 2}}{Nh^{p+q+1}} \rightarrow 0$$

$$\lim_{N \rightarrow \infty} \frac{\alpha_N^{\eta \wedge 2}}{N} \rightarrow 0$$

$$\lim_{N \rightarrow \infty} \frac{h^{2s}}{N} \rightarrow 0$$

Now, we can state the our theorem about the \sqrt{N} -consistency of $\hat{\beta}$.

Theorem 5 *Under the assumptions 5, 6, 7, 8, 11, 12, 13, 14:*

$$\sqrt{N} \left\| \hat{\beta} - \beta \right\| = O_p(1)$$

4 Data Based Selection of α

As we have mentioned before, selection of regularization parameter is crucial. It is of great importance because it characterizes the balance between the fitting and the smoothing.

In Heinz W. Engl and Neubauer (1996), a heuristic selection rule is proposed based on a comparison between the residual and the assumed bound for the noise level. Moreover, it has been proved that the regularization method where α is defined via the aforementioned rule, namely *the discrepancy principle* is convergent and of optimal order.

Florens and Lestringant (2007) have proposed two adapted methods for the selection of optimal regularization parameter: *Method of residuals* and the *L-curve method*. In this paper, we introduce the adaptation of method of residuals to our model and we use this one in our simulations. Nonetheless, as was noted in Florens and Lestringant (2007) as well, the aim is more to locate α where we could have the optimal solution. The α that is given by the adaptive selection rule may not be the final value, it can give a clue to work on the optimal solution.

The method of residuals, where the squared norm of residuals are used, is defined for a given bandwidth. Since the statistic $\|\hat{\epsilon}^\alpha\|$ can be proved to reach it minimum at $\alpha = 0$ and it is an increasing function of α , Florens and Lestringant (2007) has introduced two modifications: the first one is to calculate the residuals of an estimation obtained by an iterated Tikhonov regularization of order two. The second modifications is the division of this norm by α^2 .

Now, let us adopt this rule to our basic model:

$$\hat{\epsilon}_{(2)}^\alpha = \begin{pmatrix} A_y X \\ A_z X \end{pmatrix} - \begin{pmatrix} A_y A_{xw} & -A_y A_{xw} \\ A_z A_{xw} & -A_z A_{xw} \end{pmatrix} \begin{pmatrix} \hat{H}_{1,(2)} \\ \hat{\varphi}_{(2)} \end{pmatrix}$$

where $\begin{pmatrix} \hat{H}_{1,(2)} \\ \hat{\varphi}_{(2)} \end{pmatrix}$ are the estimators obtained by using an iterated Tikhonov regularization of order two, i.e.:

$$(\hat{H}_{1,(2)} \hat{\varphi}_{(2)})' = (\alpha_N I_{2N} + \hat{T}^* \hat{T})^{-1} (X + \alpha ((\hat{H}_{1,(1)} \hat{\varphi}_{(1)})')$$

And the optimal α is given by:

$$\alpha^* = \underset{\alpha}{\operatorname{argmin}} \frac{1}{\alpha^2} \|\hat{\epsilon}_{(2)}^\alpha\|^2 \quad (14)$$

5 A Simulation Analysis

We made a simulation analysis to see if our method is working well. After simulating the data we estimated it both parametrically (by GMM) and nonparametrically (by our method). In the end we showed that, high nonlinearity of the model gives very high mean squared errors in the parametric estimation, on the other hand, the choice of regularization parameter α is crucial in the nonparametric estimation.

5.1 Generation of Data

We generated our data according to the models in equations 3 and 4. For $H_i(\cdot)^{-1} = S_i(\cdot)$ we have chosen the following form:

$$S_i(x) = \frac{1}{1 + ke^x} \quad i = 1, 2$$

Moreover, $\varphi(\cdot)$ and $\psi(\cdot)$ were chosen to be:

$$\varphi(x) = Ax^a$$

$$\psi(x) = Bx^b$$

Then our simultaneous equations system became:

$$Y = \frac{1}{1 + ke^{AZ^a + X + U}} \quad (15)$$

$$Z = \frac{1}{1 + ke^{BY^b + W + V}} \quad (16)$$

Whose parameters were associated with the following values: $k = 0.1$, $A = 1$, $a = 2$, $B = 1$ and $b = 0.5$. (X, W) were drawn from a joint normal distribution with mean $[2, 3]$ and covariance

$$\Sigma = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$$

Moreover, u and v were drawn independently from a standard normal distribution. For these given structure, we solved our two nonlinear equations in two unknowns to get the values of Y and Z to use in the estimation process⁵. We generated samples of different sizes also to control for the effect of sample size on the estimation results.

We first estimated the model by a parametric method to be able to compare it with the results of nonparametric estimation. For the parametric estimation we used 2-stage GMM. Moreover we did two different estimations, the difference coming from the moment conditions⁶. In the first estimation, the set of moment conditions was composed of the explanatory variables and their powers while in the second one, we kept the same form of the moments but increased their number.

As was mentioned before, for nonparametric estimation we used the method we have presented in this paper.

5.2 Results

The results of GMM estimation are not very satisfactory, since the MSE's are very large. Moreover, the efficiency of estimations are not consistent for the two equations. According to the theory we expect an increase in the efficiency when we increase the number of moment conditions. So, the most efficient estimation in our case should be the one obtained with the *2nd* set of moment conditions while the least should be the one obtained with the *1st* set. However, it is not the result we get for each equation. While for the first equation the most efficient results are obtained with the *2nd* set, for the second equation, *1st* set of moment

⁵Given the sample we have, we used the MATLAB to solve for equilibrium values of Y and Z .

⁶we give the set of moment conditions in the appendix

equations gives the best results. We display our result tables in the appendix.

The inconsistent and bad results for GMM estimations can be explained by the high non-linearity of our problem. In addition to this we can not find any other explanation. Nonetheless, this suggest us that there is a need for other techniques.

On the contrary, the results of nonparametric estimation are very satisfactory⁷. We can get very close to the true values of our estimated functions in the estimation of both equations. There are a couple of issues to worth noting. Firstly, the estimated functions are very sensitive both to the regularization parameter and the bandwidth of the kernel estimator. For example, for a relatively high bandwidth, which gives us very smooth curves, when the regularization parameter decreases, the consistency of the estimated curves deteriorates. In other words, the estimated curve from one simulation to the other differs a lot. Moreover, in this case, it is very hard to optimize in terms of regularization parameter and nearly impossible to reach the true curve. Secondly, as we are estimating two functions simultaneously, we need two different regularization parameters, which makes the estimation even harder. In the simulations we made, we saw that $\hat{H}_1(\cdot)$ is very also sensitive to the regularization parameter we used for $\hat{\varphi}(\cdot)$. For this reason, we introduced the data based selection of α to our simulations and to be able to get two different optimal regularization parameters for the two equations, we used a constant ratio between α_H and α_φ . Another point worth noticing is the change of optimal α parameter with the sample size. In the simulations we made we saw that, an α^* which gives very oscillating estimated curves with a sample size of 200, can give very smooth curve with a sample size of 500. This also supports the theory that the optimality of the regularization parameter is related with the sample size.

In addition to those, we know that, for the operators whose smallest eigenvalue is close to 0, we need a strong regularization, i.e., a large regularization parameter. In our case, not only the smallest eigen value was very close to zero but also the 3rd largest one. In our simulations, the optimal α for \hat{H}_1 lies between the ranges $[10^{-2}, 10^0]$ while the for $\hat{\varphi}$, it was

⁷Figures are presented in appendix

in $[10^2, 10^4]$.

By looking at the results of our simulation, we can conclude that, for very nonlinear models, like the one we have, parametric methods may stay weak. On the other hand, in nonparametric estimation the choice of bandwidth and regularization parameter is very important. So, we still need to work on the theory of choice of these parameters, especially on the simultaneous choice.

6 Application: A Network Diffusion Model for Two Sided Markets

The topic of two-sided markets is not very old in economics. Nonetheless, lots of works have been done in terms of theory. In contrast, there are very few papers looking at the topic from an empirical point of view. In this section we adopt network diffusion models to the case of two-sided markets and we suggest using the nonparametric method we have presented to estimate the model.

The main feature of a two-sided market is the existence of externalities between the two sides of the market. In other words, decision of one side to enter the market or not depends on the decision of the other side. So, the platforms- we can think of them as the suppliers- have to “get both sides on the board”. There are many examples of such platforms: Magazines and newspapers, academic journals, television channels, dating agencies, credit cards, shopping malls, etc. In the case of television channels and magazines for example, firms would like to give ads to a channel which is watched by a lot of viewer however viewers would like to watch a channel with fewer ads. In the case of credit cards, a consumer wants to hold a card which is most widely accepted by the retailers and retailers would like to have the machine of a card which is most widely used by consumers.

We observe different structures in two-sided markets: The externality to each group might be of different size relatively, which in turn brings about different pricing structure.

The pricing can be like fixed fees, per-transaction charges or two-part tariffs. Moreover, different sides can do single-homing or multi-homing which affects pricing a lot. If an agent chooses to use only one platform then it is said that it "*single-homes*", such as a reader buying only one newspaper every day. If an agent uses many platforms then we say that it "*multi-homes*" like companies giving adds to many newspapers. In this case we could have three different situations : (i) both sides can do single-homing, (ii) one side can do single-homing and the other can do multi-homing and (iii) both sides do multihoming. The interesting case to be examined is the case (ii), which is called '*competitive bottlenecks*'. In the case of competition, the platforms will have monopoly power over providing access to their single-homing consumers for the multi-homing sides and thus in the multi-homing side the prices are going to be higher and there will be too few agents of that side on the platform.

On the other hand, the diffusion model of physics, has been widely used in economics. Except its heavy use in finance, it was first adopted by Bass (1969) where he developed a growth model for new consumer durables. Then it was also used in models of network economics, like in Larribeau (1993) where she derive the demand for telephone in Spain or in Feve, Florens, Rodriguez, and Soteri (2008), where they derived the demand for mail, by looking at the growth in internet advertising. So, the fact that network externalities exist between the two sides of the market in a two-sided market framework makes it seems reasonable to use a network diffusion model for this topic, too.

Let us first introduce our set-up. We are in a two-sided market with sides $i = 1, 2$. Each side decides whether to enter the platform or not by looking at the benefits and costs of entering that platform. Benefit of entering a platform for side i , is a function of the platform characteristics X_i , the share of agents of side j on that platform, I_j , and the cost of entering the platform c_i . For the moment we assume that the cost of entering the platform is exogenous. So, an agent n on side i will enter the platform if her net benefits is higher than its cost or more precisely:

$$b_{n,i} \geq A_i(X_i, I_j, c_i)$$

where $b_{n,i}$ is the net benefit of entering the platform of agent n . Thus the probability of entering the platform and hence the share of buyers who enter the platform at time t is given by:

$$I_{i,t} = S_i(A_i(X_{i,t}, I_{j,t}, c_{i,t}, u_t))$$

where $S(\cdot)$ is the survivor function defined as $Pr(\theta \geq t) = S(t)$ and u_t is the error term. It should be noted that, contrary to many nonparametric studies, we are not using the additively separable error term. From an intuitive point of view this will mean that the random term is coming from within our model.

For the sake of simplicity, we are taking prices as exogenous and estimate only the demand equations as a benchmark model. So our simultaneous equations to be estimated are:

$$I_{1,t} = S_1(A_1(X_{1,t}, I_{2,t}, c_{1,t}, u_t))$$

$$I_{2,t} = S_2(A_2(X_{2,t}, I_{1,t}, c_{2,t}, v_t))$$

Moreover, we assume that the function A has the form

$$A_1(X_{1,t}, I_{2,t}, c_{1,t}, u_t) = \varphi(I_{2,t}) + \beta X_{1,t} + u_t$$

where in this new form X_t captures both the platform characteristics and the cost. So we have:

$$I_1 = S_1(\varphi(I_2) + \beta' X_1 + u) \quad \text{for side one} \quad (17)$$

$$I_2 = S_2(\psi(I_1) + \gamma' X_2 + v) \quad \text{for side two} \quad (18)$$

We have decided to use long-run equilibrium equation of network diffusion models. For this, we have two motivations: First, we believe that many of the two sided industries are now mature industries so it is hard to observe a convergence process. In other words, we believe that many of the two sided industries have already reached their steady states. Nonetheless,

in estimation with real data this can be checked by looking at both of the models. Second, in the adaption of dynamic model we would have lagged variable of the share of people using the same platform on the other side and we would not have simultaneous equations. For econometric theory considerations we would like to develop a nonparametric model that addresses the estimation of simultaneous equations. For the moment, we are leaving the estimation of dynamic equation to the future work.

7 Conclusion

In this paper, we developed a nonparametric estimation technique for simultaneous equations with nonseparable error terms and showed that it works well with the simulations we made. Moreover, we presented the adoption of network diffusion models to the case of two sided markets and suggested to estimate it with our new method.

The method itself is very easy to apply as it is based on kernels though simulations showed that it is very sensitive to the choice of bandwidth and regularization parameters. Moreover, the model we presented, requires the estimation of two functions simultaneously, thus leads to the problem of choosing two optimal regularization parameters simultaneously. We solved this problem by introducing two different regularization parameters and preserving a constant ratio between the two. Moreover, we think that in the nonparametric estimation part, methods other than kernels, like splines, polynomials, etc. can be used and an efficiency comparison can be made as an extension.

A more important extension, can be the full information method rather than following a limited information approach. In fact, this is a topic that we are still working on and for the moment we can conclude that this can be done in a very similar way to ours in this paper. Finally, an estimation which is derived from a structural economic model can be done with real data and the performance of a parametric and our nonparametric estimation can be compared.

Appendices

A Proofs of Theorems

A.1 Theorem 1

Proof. by assumption1

$$\mathbb{E}[H_1(Y) - \varphi(Z) - X|X, W] = 0$$

Let us recall two more functions $H_1^*(Y)$ and $\varphi^*(Z)$. By assumption1 again, we can write:

$$\mathbb{E}[H_1(Y) - \varphi(Z) - X|X, W] = 0 \quad \mathbb{E}[H_1^*(Y) - \varphi^*(Z) - X|X, W] = 0$$

If we take the difference of the two expectations:

$$\mathbb{E}[(H_1(Y) - H_1^*(Y)) - (\varphi(Z) - \varphi^*(Z)) + (X - X)|X, W] = 0$$

then by assumption2:

$$(H_1(Y) - H_1^*(Y)) - (\varphi(Z) - \varphi^*(Z)) = 0$$

finally by assumption3

$$(H_1(Y) - H_1^*(Y)) = (\varphi(Z) - \varphi^*(Z)) = c$$

and by assumption4:

$$c = 0$$

then:

$$H_1(Y) = H_1^*(Y) \quad \text{and} \quad \varphi(Z) = \varphi^*(Z)$$

■

A.2 Theorem 3

Proof. Remember that the solution of our problem was given by

$$\Phi = (\alpha_N I + T^* T)^{-1} T^* X$$

For the proof of the first part, we will decompose our equation into three parts as was done in Darolles, Florens, and Renault (2009) and look at the rates of convergence term by term.

$$\begin{aligned} \hat{\Phi}_N^\alpha - \Phi &= \underbrace{(\alpha_N I + \hat{T}^* \hat{T})^{-1} \hat{T}^* X - (\alpha_N I + \hat{T}^* \hat{T})^{-1} \hat{T}^* \hat{T} \Phi}_I \\ &\quad \underbrace{(\alpha_N I + \hat{T}^* \hat{T})^{-1} \hat{T}^* \hat{T} \Phi - (\alpha_N I + T^* T)^{-1} T^* T \Phi}_{II} \\ &\quad \underbrace{(\alpha_N I + T^* T)^{-1} T^* T \Phi - \Phi}_{III} \end{aligned}$$

The first term (*I*) is the estimation error about the right hand side (X) of the equation, the second term (*II*) is the estimation error coming from the kernels and the third term (*III*) is the regularization bias coming from regularization parameter α .

Now, let's first examine the first term:

$$I = (\alpha_N I + \hat{T}^* \hat{T})^{-1} \hat{T}^* X - (\alpha_N I + \hat{T}^* \hat{T})^{-1} \hat{T}^* \hat{T} \Phi$$

$$I = (\alpha_N I + \hat{T}^* \hat{T})^{-1} \hat{T}^* (X - \hat{T} \Phi)$$

$$\|I\|^2 = \left\| (\alpha_N I + \hat{T}^* \hat{T})^{-1} \right\|^2 \left\| \hat{T}^* X - \hat{T}^* \hat{T} \Phi \right\|^2$$

where the first term is $O\left(\frac{1}{\alpha_N}\right)$ by Fève and Florens (2009) and the second term is $O\left(\frac{1}{N h_N^{p+q+1}} + h_N^{2s}\right)$ by assumption 7.

Now, let us look at the second term II :

$$\begin{aligned}
II &= (\alpha_N I + \hat{T}^* \hat{T})^{-1} \hat{T}^* \hat{T} \Phi - (\alpha_N I + T^* T)^{-1} T^* T \Phi \\
&= \left[\left[I - (\alpha_N I + \hat{T}^* \hat{T})^{-1} \hat{T}^* \hat{T} \right] - \left[I - (\alpha_N I + T^* T)^{-1} T^* T \right] \right] \Phi \\
&= \left[\alpha_N (\alpha_N I + \hat{T}^* \hat{T})^{-1} - \alpha_N (\alpha_N I + T^* T)^{-1} \right] \Phi \\
&= (\alpha_N I + \hat{T}^* \hat{T})^{-1} (\hat{T}^* \hat{T} - T^* T) \alpha_N (\alpha_N I + T^* T)^{-1} \Phi \\
\|II\|^2 &= \left\| (\alpha_N I + \hat{T}^* \hat{T})^{-1} \right\|^2 \left\| (\hat{T}^* \hat{T} - T^* T) \right\|^2 \left\| \alpha_N (\alpha_N I + T^* T)^{-1} \Phi \right\|^2
\end{aligned}$$

The first term in (II) is $O(\frac{1}{\alpha_N})$ by Fève and Florens (2009) while the second one is of order $O\left(\frac{1}{N h_N^{p+q+1}} + h_N^{2s}\right)$ as a result of relation $\left\| \hat{T}^* \hat{T} - T^* T \right\| = O\left(\max\left\|\hat{T} - T\right\|, \left\|\hat{T}^* - T^*\right\|\right)$ by assumption 6 and by Florens, Johannes, and Van Belleghem (2009). Finally, the third is equal to $O(\alpha_N^{\min(\nu+1, 2)})$ by Darolles, Florens, and Renault (2009).

The third term can be examined more straightforwardly:

$$\begin{aligned}
III &= (\alpha_N I + T^* T)^{-1} T^* T \Phi - \Phi \\
&= \Phi_N^\alpha - \Phi
\end{aligned}$$

and $\|III\|^2 = \|\Phi_N^\alpha - \Phi\|^2$ is $O(\alpha_N^\nu)$ by assumption 5. Finally if we combine all what we have:

$$\left\| \hat{\Phi}_N^{\alpha_N} - \Phi \right\|^2 = O\left(\frac{1}{\alpha} \left(\frac{1}{N h_N^{p+q+1}} + h_N^{2s}\right) + \frac{1}{\alpha} \left(\frac{1}{N h_N^{p+q+1}} + h_N^{2s}\right) (\alpha^{\min(\nu+1, 2)}) + \alpha^{nu}\right)$$

Now we can continue with the proof of the second part of our theorem. To do this, we will decompose the estimation error into two as estimation bias and regularization bias:

$$\left\| \hat{\Phi}_N^\alpha - \Phi \right\| \leq \underbrace{\left\| \hat{\Phi}_N^\alpha - \Phi_N^\alpha \right\|}_A + \underbrace{\left\| \Phi_N^\alpha - \Phi \right\|}_B$$

We know that the regularization bias goes to 0 as $\alpha_N \rightarrow 0$, so let us examine A :

$$\begin{aligned} \hat{\Phi}_N^\alpha - \Phi_N^\alpha &= (\alpha_N I + \hat{T}^* \hat{T})^{-1} \hat{T}^* X - (\alpha_N I + T^* T)^{-1} T^* T \Phi \\ &= \underbrace{(\alpha_N I + \hat{T}^* \hat{T})^{-1} (\hat{T}^* X - \hat{T}^* \hat{T} \Phi)}_I \\ &= \alpha_N \underbrace{\left[(\alpha_N I + \hat{T}^* \hat{T})^{-1} - (\alpha_N I + T^* T)^{-1} \right]}_{II} \Phi \end{aligned}$$

If we look at A term by term:

$$I = \left\| (\alpha_N I + \hat{T}^* \hat{T})^{-1} \hat{T}^* \right\| \left\| (X - \hat{T} \Phi) \right\|$$

the first term is of order $O(\frac{1}{\sqrt{\alpha_N}})$ by Feve and Florens (2009) and the second term is of order $O(\frac{1}{\sqrt{N h_N^{p+q+1}}} + h_N^s)$ by assumption 7. So the the first term (I) converges to $\left(\frac{1}{\sqrt{\alpha_N}} \left(\frac{1}{\sqrt{N h_N^{p+q+1}}} + h_N^s \right) \right)$.

$$II = \alpha_N \left[(\alpha_N I \hat{T}^* \hat{T})^{-1} - (\alpha_N I + T^* T)^{-1} \right] \Phi$$

$$\|II\| = \left\| \alpha_N (\alpha_N I + T^* T)^{-1} \Phi \right\| \left\| \hat{T}^* \hat{T} - T^* T \right\| \left\| (\alpha_N I + T^* T)^{-1} \right\|$$

The first part is $\|\Phi - \Phi_N^\alpha\|$ and has zero limit, the second term is of order $O(\frac{1}{\sqrt{N h_N^{p+q+1}}} + h_N^s)$ by assumption 6 and by Florens, Johannes, and Van Belleghem (2009) and the last term is smaller than $O(\frac{1}{\sqrt{\alpha_N}})$. So, the second term (II) of A also converges to $\left(\frac{1}{\sqrt{\alpha_N}} \left(\frac{1}{\sqrt{N h_N^{p+q+1}}} + h_N^s \right) \right)$. Then, we can conclude that $\left\| \hat{\Phi}_N^\alpha - \Phi \right\|$ converges to zero in probability if $\alpha_N \rightarrow 0$, $\frac{h_N^{2s}}{\alpha_N} \rightarrow 0$ and $\frac{1}{N h_N^{p+q+1} \alpha_N} \sim O(1)$.

■

A.3 Theorem 4

Proof.

$$H_1(Y) - \varphi(Z) - X_0 - X'\beta = U$$

$$\mathbb{E}[H_1(Y) - \varphi(Z) - X_0 - X'\beta|X, W] = 0 \quad \text{by } \textit{assumption1}$$

Let us recall two more functions $H_1^*(Y)$, $\varphi^*(Z)$ and β^* such that:

$$H_1^*(Y) - \varphi^*(Z) - X_0 - X'\beta^* = U$$

Then, again by *assumption1*:

$$\mathbb{E}[H_1^*(Y) - \varphi^*(Z) - X_0 - X'\beta^*|X, W] = 0 \quad \text{by } \textit{assumption1}$$

If we take the difference of the two expectations:

$$\mathbb{E}[(H_1(Y) - H_1^*(Y)) - (\varphi(Z) - \varphi^*(Z)) - (X'\beta - X'\beta^*)|X, W] = 0$$

Then, by *assumption9*:

$$(H_1(Y) - H_1^*(Y)) - (\varphi(Z) - \varphi^*(Z)) - (X'\beta - X'\beta^*) = 0$$

By *assumption10*:

$$(H_1(Y) - H_1^*(Y)) - (\varphi(Z) - \varphi^*(Z)) = (X'\beta - X'\beta^*)$$

Finally by *assumptions 3 and 4* we get the identification:

$$H_1(Y) = H_1^*(Y) \quad \varphi(Z) = \varphi^*(Z) \quad X'\beta = X'\beta^*$$

■

A.4 Theorem 5

Proof.

$$\hat{\beta} - \beta = \underbrace{\hat{M}_\alpha^{-1}}_I \left\{ \underbrace{[\hat{T}_X^* - \widehat{T}_X^* T(\alpha I + \hat{T}^* \hat{T})^{-1} \hat{T}^*]}_{II} (X_0 - \hat{T}(H, \varphi) + T_X \beta) \underbrace{[\hat{T}_X^* - \widehat{T}_X^* T(\alpha I + \hat{T}^* \hat{T})^{-1} \hat{T}^*]}_{III} T(H, \varphi) \right\} \quad (19)$$

where $\hat{M}^\alpha = \widehat{T}_X^* T(\alpha I + \hat{T}^* \hat{T})^{-1} \widehat{T}^* T_X - \widehat{T}_X^* T_X$. To prove our final result, we will show the following asymptotic convergences:

$$\begin{aligned} I &= \left\| \hat{M}_\alpha^{-1} - M_\alpha^{-1} \right\| = O_p \left(\alpha^{\frac{\eta \wedge 2}{2}} \left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{N h^{p+q+1}}} + h^s + \frac{1}{N h} + h^s \right) + \frac{1}{\sqrt{N}} \right) \\ II &= O_p \left(\left[\frac{1}{\sqrt{N}} \left(1 + \frac{1}{\sqrt{\alpha}} \right) \left(\frac{1}{\sqrt{N h^{p+q+1}}} + h^s \right) \right] \right. \\ &\quad \left. + \frac{1}{\sqrt{\alpha}} \left(\frac{1}{N h^{p+q+1}} + h^s \right) + \alpha^{\frac{\eta \wedge 2}{2}} \left(\frac{1}{N h^{p+q+1}} + h^s \right) + \alpha^{\frac{\eta \wedge 2}{2}} \left(\frac{1}{\sqrt{N h^{p+q+1}}} + h^s \right) \right) \\ III &= O_p \left(\frac{1}{\sqrt{N}} + \sqrt{\alpha} \left(\frac{1}{\sqrt{N h^{p+q+1}}} + h^s \right)^{\frac{\nu \wedge 2}{2}} + \alpha^{\frac{1 \wedge (1+\nu)}{2}} \right) \end{aligned}$$

The assumptions we made to state *Theorem 5* ensure that I has the rate $o_p(1)$ while II and III have the rate $O_p(N^{-1/2})$. Let us begin with I .

Proof of I:

$$\left\| \hat{M}_\alpha^{-1} - M_\alpha^{-1} \right\| \leq \left\| M_\alpha^{-1} \right\| \left\| \hat{M}_\alpha^{-1} \right\| \left\| \hat{M}_\alpha - M_\alpha \right\|$$

The first term above is bounded and the second term above is bounded in probability so we need to look at the convergence of the third term.

$$\begin{aligned} \left\| \hat{M}_\alpha - M_\alpha \right\| &= \left\| \left[\widehat{T}_X^* T(\alpha I + \hat{T}^* \hat{T})^{-1} \widehat{T}^* T_X - \widehat{T}_X^* T_X \right] - \left[T_X^* T(\alpha I + T^* T)^{-1} T^* T_X - T_X^* T_X \right] \right\| \\ &\leq \left\| \widehat{T}_X^* T(\alpha I + \hat{T}^* \hat{T})^{-1} \widehat{T}^* T_X - T_X^* T(\alpha I + T^* T)^{-1} T^* T_X \right\| \end{aligned}$$

$$\begin{aligned}
& + \left\| T_X^* T_X - \widehat{T_X^* T_X} \right\| \\
& \leq \underbrace{\left\| \widehat{T_X^* T} - T_X^* T \right\| \left\| (\alpha I + T^* T)^{-1} T^* T_X \right\|}_A \\
& + \underbrace{\left\| T_X^* T [(\alpha I + \widehat{T}^* \widehat{T})^{-1} - (\alpha I + T^* T)^{-1}] T^* T_X \right\|}_B \\
& + \underbrace{\left\| T_X^* T (\alpha I + T^* T)^{-1} \right\| \left\| \widehat{T^* T_X} - T^* T_X \right\|}_C + \underbrace{\left\| T_X^* T_X - \widehat{T_X^* T_X} \right\|}_D
\end{aligned}$$

- The first term in A is of order $O(1/\sqrt{N})$ by assumption 13 and the second term is $O(\alpha^{\frac{\nu \wedge 2}{2}})$ by Florens, Johannes, and Van Bellegem (2009)
- B can be decomposed as the following:

$$\|B\| \leq \left\| T_X^* T_X (\alpha I + \widehat{T}^* \widehat{T})^{-1} \right\| \left\| T^* T - \widehat{T}^* \widehat{T} \right\| \left\| (\alpha I + T^* T)^{-1} T^* T_X \right\|$$

The first term is bounded. The second term is of order $O(\frac{1}{\sqrt{N}h^{p+q+1}} + h^s)$ by assumption 6 and the third term is of order $O(\alpha^{\frac{(\nu \wedge 2)}{2}})$ again by Florens, Johannes, and Van Bellegem (2009).

- The first of C is $O(\alpha^{\frac{(\nu \wedge 2)}{2}})$ and the second term is $O\left(\frac{1}{\sqrt{N}h} + h^s\right)$ by assumption 12.
- Finally D is of order $O\left(\frac{1}{\sqrt{N}}\right)$ by assumption 13.

Proof of II:

We can denote $\hat{e} = X_0 - \widehat{T}(H, \varphi) + T_X \beta$. Then:

$$\|\hat{e}\| \leq \left\| \widehat{T} - T \right\|$$

which is of order $O(\frac{1}{\sqrt{N}h^{p+q+1}} + h^{2s})$ by assumption 6. Then we can write II as:

$$[T_X^* - \widehat{T_X^* T} (\alpha I + \widehat{T}^* \widehat{T})^{-1} \widehat{T}^*] \hat{e}$$

$$\begin{aligned}
&= \left\{ (\hat{T}_X^* - \widehat{T}_X^* T (\alpha I + \hat{T}^* \hat{T})^{-1} \hat{T}^*) - (T_X^* - T_X^* T (\alpha I + T^* T)^{-1} T^*) \right\} \hat{e} \\
&\quad + (T_X^* - T_X^* T (\alpha I + T^* T)^{-1} T^*) \hat{e}
\end{aligned}$$

The first part is of order $O_p \left(\left[\frac{1}{\sqrt{N}} \left(1 + \frac{1}{\sqrt{\alpha}} \right) \left(\frac{1}{\sqrt{N h^{p+q+1}}} + h^s \right) \right] + \frac{1}{\sqrt{\alpha}} \left(\frac{1}{N h^{p+q+1}} + h^s \right) + \alpha^{\frac{\eta \wedge 2}{2}} \left(\frac{1}{N h^{p+q+1}} + h^s \right) \right)$ and the second part is of order $O(\alpha^{\frac{(\nu \wedge 2)}{2}} (\frac{1}{\sqrt{N h^{p+q+1}}} + h^s))$.

Proof of III:

By assumption 5 and following Florens, Johannes, and Van Bellegem (2009), we can write:

$$\begin{aligned}
&\left\| [T_X^* - \widehat{T}_X^* T (\alpha I + \hat{T}^* \hat{T})^{-1} \hat{T}^*] \hat{T}(H, \varphi) \right\| \leq \left\| \widehat{T}_X^* T \right\| (H, \varphi) \\
&\quad + \left\| \widehat{T}_X^* T (\alpha I + \hat{T}^* \hat{T})^{-1} \hat{T}^* \hat{T} \right\| \left\| (T^* T)^{\nu/2} - (\hat{T}^* \hat{T})^{\nu/2} \right\| \|g\| \\
&\quad + \left\| \widehat{T}_X^* T (\alpha I + \hat{T}^* \hat{T})^{-1} \hat{T}^* \hat{T} (\hat{T}^* \hat{T})^{\nu/2} \right\| \|g\|
\end{aligned}$$

The first term is $O(\frac{1}{\sqrt{N}})$, the second is of order $O(\alpha^{1/2})$. Moreover by Heinz W. Engl and Neubauer (1996) $\left\| (T^* T)^{\nu/2} - (\hat{T}^* \hat{T})^{\nu/2} \right\| \leq \left\| T^* T - \hat{T}^* \hat{T} \right\|^{(\nu \wedge 2)/2}$. The rate of the last part is also given by Heinz W. Engl and Neubauer (1996) and is equal to $O(\alpha^{\frac{1 \wedge (1+\nu)}{2}})$.

■

B Simulation Results

B.1 GMM estimation

The two instrument set that we used are the following:

1st set:

$$M1 = \{1, x, w, x^2, w^2, xw\}$$

2nd set

$$M2 = \{1, x, w, x^2, w^2, x^3, w^3, xw, x^2w, xw^2\}$$

First equation

Table 1: Simulation results for the 1st equation with the 1st set of moments

	\hat{k}	\hat{A}	\hat{a}
Bias	-0.0391	0.4914	0.7188
Variance	0.0001	7.4886	4.9070
MSE	0.0016	7.6552	5.3746

Table 2: Simulation results for the 1st equation with the 2nd set of moments

	\hat{k}	\hat{A}	\hat{a}
Bias	-0.0376	0.0989	0.4702
Variance	0.0002	0.2745	0.2618
MSE	0.0016	0.2816	0.4803

Second equation

Table 3: Simulation results for the 2nd equation with the 1st set of moments

	\hat{k}	\hat{B}	\hat{b}
Bias	-0.0290	-0.0819	0.1449
Variance	0.0007	0.2245	0.0242
MSE	0.0015	0.2289	0.0450

Table 4: Simulation results for the 2nd equation with the 2nd set of moments

	\hat{k}	\hat{B}	\hat{b}
Bias	-0.0208	-0.2035	0.1454
Variance	0.0086	2.3614	0.0263
MSE	0.0089	2.3792	0.0471

B.2 Nonparametric estimation

Below, we present the nonparametric estimation results of the first equation of our system.

Figures contain both the \hat{H}_1 and $\hat{\varphi}$ for different values α_H and α_φ

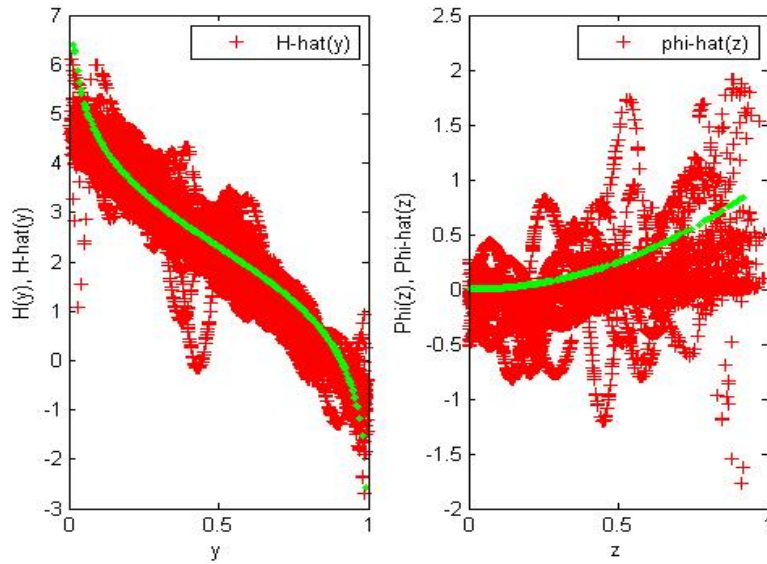


Figure 1: *Estimated functions for $\alpha = 7.8 \times 10^{-4}$ and $c = 100$*

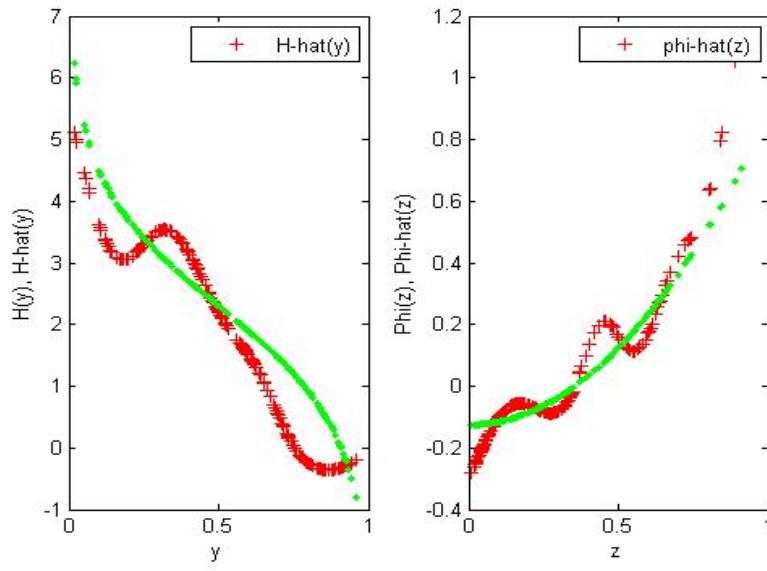


Figure 2: *Estimated functions for a sample of 200*

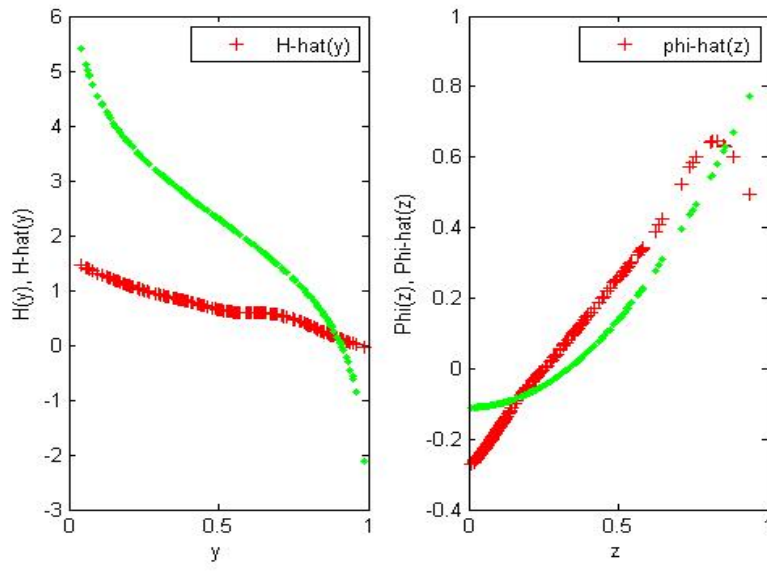


Figure 3: *Estimated functions for $\alpha_H = 1$ and $c = 1$*

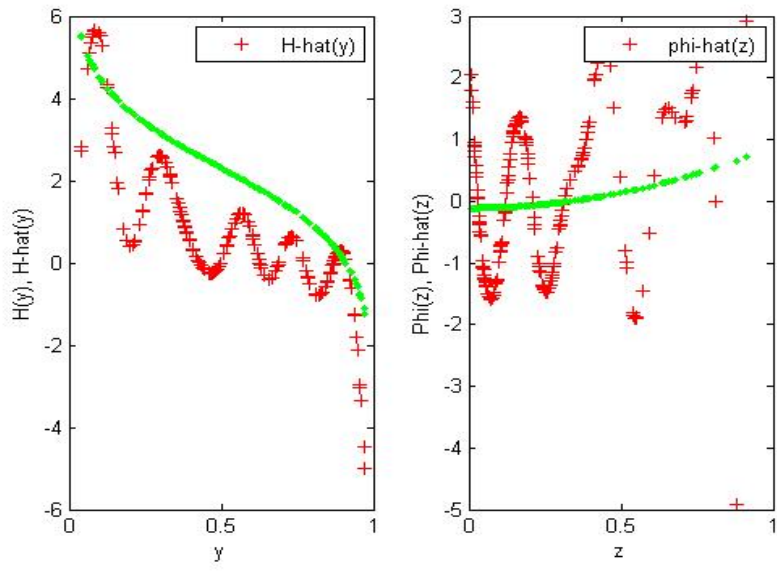


Figure 4: *Estimated functions for $\alpha_H = 10^{-5}$ and $c = 1$*

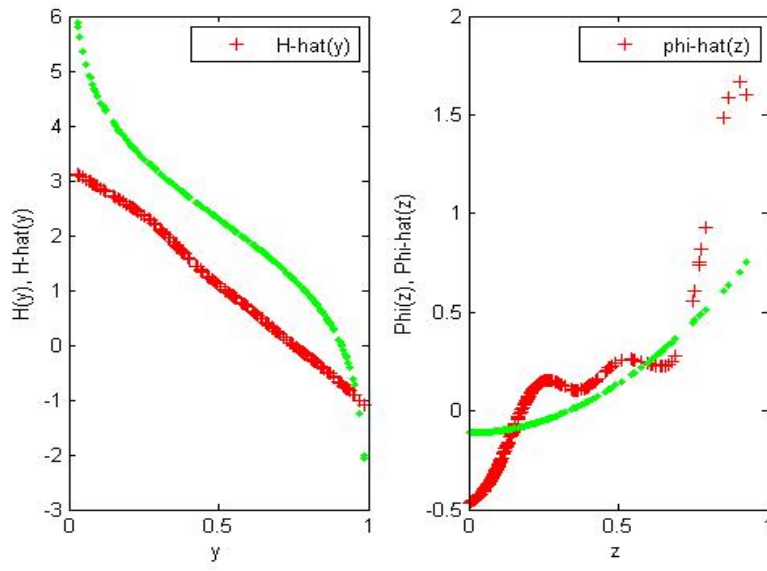


Figure 5: *Estimated functions for $\alpha_H = 10^{-1}$ and $c = 1$*

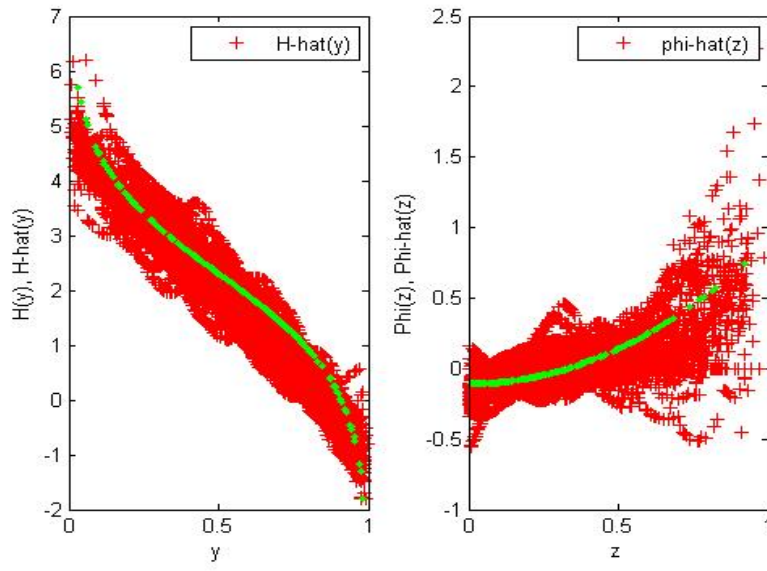


Figure 6: *Monte Carlo simulation*

References

- BASS, F. M. (1969): “A New Product Growth for Model Consumer Durables,” *Management Science*, 15, 215–227.
- CARRASCO, M., J.-P. FLORENS, AND E. RENAULT (2007): “Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization,” in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6 of *Handbook of Econometrics*, chap. 77. Elsevier.
- DAROLLES, S., J.-P. FLORENS, AND E. RENAULT (2009): “Nonparametric Instrumental Regression,” Discussion paper, Toulouse School of Economics.
- FEVE, F., AND J.-P. FLORENS (2009): “The Practice of Non Parametric Estimation by Solving Inverse Problems: The Example of Transformation Models,” Discussion paper, Toulouse School of Economics.
- FEVE, F., J.-P. FLORENS, F. RODRIGUEZ, AND S. SOTERI (2008): “Internet Advertising and Direct Mail: Trends and Analysis for the UK,” Discussion paper, Toulouse School of Economics, Toulouse.
- FLORENS, J.-P., J. JOHANNES, AND S. VAN BELLEGEM (2009): “Instrumental Regression in Partially Linear Models,” Discussion paper, Toulouse School of Economics.
- FLORENS, J.-P., AND R. LESTRINGANT (2007): “The Practice of Non-Parametric Instrumental Variables Estimation,” Discussion paper, Toulouse School of Economics, Toulouse.
- HEINZ W. ENGL, M. H., AND A. NEUBAUER (1996): *Regularization of Inverse Problems*. Kluwer Academic Publications, Dordrecht.
- HOROWITZ, J. L. (1996): “Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable,” *Econometrica*, 64, 103–137.

LARRIBEAU, S. (1993): “Microeconometrie De La Demande de Telephone,” Ph.D. thesis, Universite des Sciences Sociales de Toulouse, Toulouse, France.

ROHRIG, C. S. (1988): “Conditions for Identification in Nonparametric and Parametric Models,” *Econometrica*, 56, 433–447.