Non Parametric Estimation of Simultaneous Equations

Senay SOKULLU [*]

April 6, 2009

Abstract

In this paper we develop a nonparametric estimation technique for the simultaneous equations systems with nonadditive random terms. 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 nonparametric estimation literature is combining the estimation of simultaneous equations with the estimation of nonparametric functions with nonadditive random terms. 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 better results compared to GMM , but also that the choice of regularization parameter together with the bandwidth is very important to get good estimates. Furthermore, as an application example, we apply the network diffusion models to the case of two-sided markets.

Keywords: Nonadditive random terms, Nonlinear inverse problems, Tikhonov Regularization

JEL Classification: C13, C14, C30

---

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

1 Introduction

Nonparametric econometrics is becoming more and more important. Roehrig (1988) explaines the usefullness 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, the assumption of additively separable random terms stays very questionable. As mentioned by Matzkin (2003), random terms instead should be genereated from within the model and this corresponds to the case of nonadditive random terms. The examples she has given are very illustrative: a nonadditive random term may represent a heterogeneity parameter in a utility function or some productivity shock in a production function. In most of the economic models, the dependent variable is a function of both the explanatory variables and the random term, and the random term is not additive.

Our contribution to the nonparametric estimation literature is combining the estimation of simultaneous equations with the estimation of nonparametric functions with nonadditive random terms. Although there are papers about these subjects separately, none of them considers the case of the two situations together. For example, Matzkin (2003) presents estimators for nonparametric functions with additively nonseparable random terms where she also assumes unknown distribution function for random term. She exploits some properties that are implied by economic theory and kernel representation of distribution functions to estimate the functions. On the other hand Brown and Matzkin (1998) give the identification conditions to estimate nonparametric simultaneous equations systems with additive random terms. To estimate these functions, they adopt Manski's Closest Empirical Distribution method. In this paper, we present the identification conditions and the estimators of a semiparametric simultaneous equations model with nonadditive random terms. For the estimation process we adopted the nonparametric instrumental variable estimator of Darolles, Florens, and Renault (2009) which is based on writing the conditional moment conditions by their kernel counterparts. To come over the problem of simultaneity, we use limited information method where we estimate our equations one by one.

One thing should be pointed out here. The estimation equation gives us an inverse problem which needs to be regularized to be solved. 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.

After developing our method, we made simulations to see how it works. To do this, we generated a sample of size 250 and replicated them 100 times. We made both parametric and nonparametric estimations. For parametric estimation we used GMM. We saw that, when the regularization parameter is chosen optimally, our estimated curve fits very well to the actual one in our nonparametric estimation while the GMM estimates are highly biased and have high mean squared errors. 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.

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 apply its use to the case of two-sided markets where we wrote the demand equations of both sides using the network diffusion model.

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 endogeneous variables, Y, Z and two exogeneous variables X, W. These variables generate the random vector Ω which has a cumulative distribution function F. Then for each F, we can define the subspaces of our variables as $L^2_F(Y)$, $L^2_F(Z)$, $L^2_F(W)$ and $L^2_F(X)$ which belong to commmon Hilbert space denoted by L^2_F. The relationship between the variables are given by the following equations system:

$$Y = S_1(\varphi(Z) + \beta X + U) \tag{1}$$

$$Z = S_2(\psi(Y) + \gamma W + V) \tag{2}$$

where U and V are the error terms.

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). So let's start with equation (1).Our model is identified[1] under the assumptions[2] $U \perp (X, W)$, $V \perp (X, W)$, (Y, Z, X, W) are identically and independently distributed with function F, and $\beta = \gamma = 1$[3].

First of all, we take the inverse of S_1 to linearize our problem. It does not affect anything

---

[1] Proof of identification and the identification conditions can be found in the appendix.

[2] If we had assume additively separable error term identification would be reached by theorem 5.1 in Roehrig (1988)

[3] this is just a simplifying assumption

in the problem, but we will recover $\hat{S}_1$ from $\hat{H}_1$[4].

$$S_1^{-1}(Y) = H_1(Y)$$

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

Our conditional independence condition can be written:

$$E(U|X, W) = 0$$

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

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

This relationship can be written using the operator $T : L^2_Y \times L^2_Z \to L^2_{X,W}$:

$$T(H_1(Y), \varphi(Z)) \to E[H_1(Y) - \varphi(Z)|X, W]$$

$$T(H_1, \varphi) = X \tag{3}$$

Equation (3) gives us an ill-posed inverse problem as H_1 and φ are both unknown. We need to regularize our problem to get a solution. For this we use Tikhonov Regularization. So the inverse of T is given by $(\alpha I + T^*T)^{-1}T^*$ where α is the regularization parameter to be selected. Then solution is given by:

$$(H_1, \varphi) = (\alpha I + T^*T)^{-1}T^*X$$

where T^* is the adjoint operator of T. To be able to solve our problem, we first need to

---

[4] By taking the inverse of S_1, we are easing the estimation. Without such a transformation we need to estimate 3 dimensional kernels which will cause problems in the simulation.

calculate T^*.

$$\langle T(H_1, \varphi), \phi(X, W)\rangle = \langle (H_1, \varphi), T^* \phi(X, W)\rangle$$

We start by writing the left hand side, then we will get the right hand side from which we can recover the T^*.

$$LHS = \int E[(H_1 - \varphi)|X, W)]\phi(X, W) f(x, w) dx dw$$

$$= \int [H_1(Y) - \varphi(Z)]\phi(X, W) f(x, w, y, z) dx dw dy dz$$

$$= \int H_1(Y)\phi(X, W) f(x, w, y) dx dw dy - \int \varphi(Z)\phi(X, W) f(x, w, z) dx dw dz$$

$$= \int H_1(Y) E[\phi(X, W)|Y] f y dy - \int \varphi(Z) E[\phi(X, W)|Z] f z dz = RHS$$

So, our adjoint operator of T,T^*:

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

To solve our regularized inverse problem we need the following assumption:

Assumption 1 *The operator T is one-to-one, and operators T, T^*,TT^* and T^*T are compact. Moreover, TT^* and T^*T have no zero eigenvalues.*

The above assumption guarantees the uniqueness of the solution even if the instruments and the explanatory variables have no elements in common. In addition, the following assumption is made to be able to cover the more general cases of regularization parameter α converging itself to zero.

Assumption 2 *(α_N, $N > 0$) is a sequence of nonnegative real numbers, converging to zero, and such that:*

$$\lim_{N\to\infty} \frac{1}{\alpha_N} \left\| \hat{T}^*\hat{T} - T^*T \right\| = 0$$

Then, we can solve our equation as follows:

$$(\alpha_N I + T^*T)(H_1, \varphi) = T^*X$$

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

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

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 increse 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:

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

$$= \frac{\sum_i x_i K_y\left(\frac{y - y_i}{h_y}\right)}{\sum_i K_y\left(\frac{y - y_i}{h_y}\right)}$$

For the second line:

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

$$= \frac{\sum_i x_i K_z\left(\frac{z - z_i}{h_z}\right)}{\sum_i K_z\left(\frac{z - z_i}{h_z}\right)}$$

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:

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

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

Our estimated functions are the solutions of the following system:

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

$$\begin{pmatrix} \hat{H}_1 \\ \hat{\varphi} \end{pmatrix} = \begin{pmatrix} \alpha_N I + A_y A_{xw} & -A_y A_{xw} \\ -A_z A_{xw} & \alpha_N I + A_z A_{xw} \end{pmatrix}^{-1} \begin{pmatrix} A_y X \\ -A_z X \end{pmatrix} \tag{4}$$

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:

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

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 assumptions[5].

Assumption 3 *$\alpha_N \in \Psi^F_\beta$, $\beta > 0$, where Ψ^F_β is defined in Darolles, Florens, and Renault (2009)*

Assumption 4 *There exists $s \geq 2$ such that:*

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

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

Assumption 5

$$\left\| \hat{T}^*X - \hat{T}^*\hat{T}\Phi \right\|^2 = O\left(\frac{1}{N} \right)$$

where $\Phi = (H_1(Y), \varphi(z))$.

Assumption 6

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

$$\lim_{N\to\infty} \alpha_N^2 N h_N^p = \infty$$

---

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

Theorem 1 *Let us define* $(H_1(Y), \varphi(z))$ *as* Φ*. Under assumptions* 1 *to* 6*:*

- $\left\| \hat{\Phi}_N^{\alpha_N} - \Phi \right\|^2 = O\left(\left(\frac{1}{\alpha_N N} \right) + \frac{1}{\alpha_N^2} \left(\frac{1}{Nh^2} + h^{2s} \right) \alpha_N^\beta + \alpha_N^\beta \right)$
- *and* $\left\| \hat{\Phi}_N^{\alpha_N} - \Phi \right\| \to 0$ *in probability.*

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.

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

$$\underbrace{(\alpha_N I + \hat{T}^* \hat{T})^{-1} \hat{T}^* \hat{T} \Phi - (\alpha_N I + T^* T)^{-1} T^* T \Phi}_{II}$$

$$\underbrace{(\alpha_N I + T^* T)^{-1} T^* T \Phi - \Phi}_{III}$$

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 $O\left(\frac{1}{\alpha_N^2}\right)$ by Groetsch (1984) and the second term is $O\left(\frac{1}{N}\right)$ by assumption 5.

Now, let us look at the second term II:

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

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

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

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

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

The first term in (II) is $O(\frac{1}{\alpha_N^2})$ by Groetsch (1984) while the second one is of order $O(\frac{1}{Nh^2} + h^{2s})$ by assumption 4 and the third was shown to be equal to $O(\alpha_N^\beta)$ in Darolles, Florens, and Renault (2009).

The third term can be examined more straightforwardly:

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

$$= \Phi_N^\alpha - \Phi$$

and $\|III\|^2 = \|\Phi_N^\alpha - \Phi\|^2$ is shown to be $O(\alpha_N^\beta)$ in Darolles, Florens, and Renault (2009). Finally if we combine all what we have:

$$\left\|\hat{\Phi}_N^{\alpha_N} - \Phi\right\|^2 = O\left(\left(\frac{1}{\alpha_N N}\right) + \frac{1}{\alpha_N^2}\left(\frac{1}{Nh^2} + h^{2s}\right)\alpha_N^\beta + \alpha_N^\beta\right)$$

The second term can be eliminated by using *Assumption 4*, so if we equalize the speeds of convergence of the first and the last term, we can get the speed of convergence of our estimators as:

$$\left\|\hat{\Phi}_N^{\alpha_N} - \Phi\right\|^2 = O(N^{\frac{-\beta}{\beta+2}})$$

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:

$$\left\|\hat{\Phi}_N^\alpha - \Phi\right\| \leq \underbrace{\left\|\hat{\Phi}_N^\alpha - \Phi_N^\alpha\right\|}_{A} + \underbrace{\|\Phi_N^\alpha - \Phi\|}_{B}$$

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

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

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

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

If we look at A term by term:

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

the first term is of order $O(\frac{1}{\alpha_N})$ by Groetsch (1984) and the second term is of order $O(\frac{1}{\sqrt{N}})$ by assumption 5. So the the first term (I) converges to $\left(\frac{1}{\alpha_N\sqrt{N}}\right)$.

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

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

The first part is $\|\Phi - \Phi_N^\alpha\|$ and has zero limit, the second term is of order $O(\frac{1}{\sqrt{Nh}} + h^s)$ by aasumption and the last term is smaller than $O(\frac{1}{\alpha_N})$. So, the second term (II) of A converges to $O\left(\frac{1}{\alpha_N}\left(\frac{1}{\sqrt{Nh}} + h^s\right)\right)$. Then, we can conclude that $\left\|\hat{\Phi}_N^\alpha - \Phi\right\|$ converges to zero in probability if $\alpha_N \to 0$, $\frac{h^{2s}}{\alpha_N^2} \to 0$ and $\frac{1}{Nh\alpha_N^2} \sim O(1)$. ■

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*.

4 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.

4.1 Generation of Data

We generated our data according to the models in equations (1) and (2). For $S_i(.)$ we have chosen the following form:

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

which we will recover it by the estimation of $H_i(.)$. Moreover, $\varphi(.)$ and $\psi(.)$ were chosen to be:

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

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

Then our simultaneous equations system became:

$$Y = \frac{1}{1+ke^{AZ^a+\beta X+U}} \tag{5}$$

$$Z = \frac{1}{1+ke^{BY^b+\gamma W+V}} \tag{6}$$

Whose parameters were associated with the following values: $k = 0.1$, $A = 1$, $a = 2$, $\beta = 1$, $B = 1$, $b = 0.5$ and $\gamma = 1$. (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 uniform distribution with mean 0. 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 process[6]. We generated 100 samples of 250.

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 three different estimations, the difference coming from the moment conditions[7]. 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. In the last estimation we adapted the efficient moment conditions of Florens and Carrasco (2004) which has the form $e^{\tau' x}$. With this formulation one can get a continuum of moment conditions and the estimators obtained are asymptotically efficient.

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

4.2 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

---

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

[7] we give the set of moment conditions in the appendix

the *3rd* 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 the first equation the most efficient results are obtained with the *2nd* set and the least efficient results are obtained with the *3rd*. For the second equation, *1st* set of moment equations gives the best results while the *3rd* one gives the worst. Moreover, very high MSE and variance values for $\hat{a}$ and $\hat{A}$ in the estimation of first equation with third set of moments worth noticing. 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 satisfactory[8]. 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 $\hat{H}_1(.)$ is very also sensitive to the regularization parameter we used for $\hat{\varphi}(.)$.

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. However, for $\hat{H}_1$ the best results were obtained in the range $[10^{-8}, 10^{-7}]$, though this weak regularization was

---

[8] Figures are presented in appendix

balanced with the strong regularization for $\hat{\varphi}$. The range of α we used for $\hat{\varphi}$ is $[0.09, 0.095]$.

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 simulataneous choice.

5 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 future 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 adds to a channel which is watched by a lot of viewer however viewers would like to watch a channel with fewer adds. 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 situtations : (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 $S(.)$ 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 simulataneous 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:

$$I_1 = S_1(\varphi(I_2) + \beta' X_1 + u) \qquad for \quad side \quad one \tag{7}$$

$$I_2 = S_2(\psi(I_1) + \gamma' X_2 + v) \qquad for \quad side \quad two \tag{8}$$

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.

6 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 peresented 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. Although data based selection methods for regularization parameter exist, none of them considers our case, i.e., none of them considers the case of using two regularization parameter and optimizing them simultaneuosly. So, this is one the cases that we leave for future work. 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 estimation of equations simultaneously 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 Proof of Identification with function S(.)

First of all we need to make following assumptions:

Assumption 7 $Y, Z, W_t \sim iid$ *with function* F *where* $W_t = (X, W)$

Assumption 8 $U \perp W_t$

Assumption 9 $S_0' \neq 0$

Assumption 10 $g_0(u|w_t) \neq 0$

Assumption 11 S_U *is given*

Assumption 12 $\beta_{01} = 1$

Assumption 13 *normalization* $\tilde{\varphi} = cte \Rightarrow \tilde{\varphi} = 0$

Assumption 14 $E[\mu(z, u, x)|u, w_t] = 0 \Rightarrow \mu = 0$

Now, we can start by writing our independency condition:

$$U \perp W \Rightarrow Pr(U \geq u|W = w) = Pr(U \geq u)$$

$$\int Pr(U \geq u, Z = z|W = w)dz = Pr(U \geq u)$$

$$\int f(z|w)Pr(U \geq u|z, w)dz = S_U(u)$$

$$\int f(z|w)Pr(\varphi(Z) + \beta' X + U \geq \varphi(Z) + \beta' X + u)dz = S_U(u)$$

$$\int f(z|w)Pr(S(\varphi(Z) + \beta' X + U) \leq S(\varphi(Z) + \beta' X + u))dz = S_U(u)$$

$$\int f(z|w)Pr(Y \leq S(\varphi(Z) + \beta' X + u)|z,w)dz = S_U(u)$$

So, our main equation will be:

$$\int f(z|w)F(S(\varphi(Z) + \beta' X + u)|z,w)dz = S_U(u) \tag{9}$$

As we know the distribution of F and f, *equation 9* will give our identification conditions. So, the identification[9] will be obtained if:

$$(f_0, F_0, \lambda_0), \quad (f_0, F_0, \lambda_1) \Rightarrow \lambda_0 = \lambda_1$$

Let us define $K_0(\lambda) = 0$ be the linear operator in the neighborhood of the true values, i.e it is the sum of the derivatives of equation 1 with respect to λ.

$$K_0(\lambda) = \int f_0 \tilde{S} + \int f_0 S_0' \tilde{\varphi} + [\int f_0 S_0' X^T]\beta' - \tilde{S}_U$$

We need to check if $K_0(\lambda) = 0$ has a unique solution or not. With the assumptions we made in the beginning our operator K_0 simplifies to:

$$K_0(\lambda) = \int f_0 \tilde{S} + \int f_0 S_0 \tilde{\varphi} = 0$$

More precisely:

$$\int f_0(S_0(\varphi_0(z) + x + u), z|w) S_0' [\tilde{S}(\varphi_0(z) + x + u) + S_0'(\varphi_0(z) + x + u)\tilde{\varphi}(z)]/S_0' dz = 0 \tag{10}$$

Note that,

$$Pr(U \leq u, z|w) = F_0(S_0(\varphi_0(z) + x + u), z|w)$$

---

[9] We started our analysis of identification by considering *local identification*, because if we can not identify our parameter locally, there is no way to identify them globally.

So the distribution of u is given by $\frac{\partial F_0}{\partial u} = S_0' f_0$. Hence we can write *equation 10* as:

$$\int g(u|w) g_0(z|u,w)[\tilde{S}(\varphi_0(z)+x+u) + S_0'(\varphi_0(z)+x+u)\tilde{\varphi}(z)]/S_0' dz = 0 \tag{11}$$

which means

$$E[\mu(z,u,x)|u,w] = 0$$

With the assumptions (7) and (8):

$$\tilde{S}(\xi) + S_0'(\xi)\tilde{\varphi}(z) = 0$$

$$\tilde{\varphi}'(z) = 0 \Rightarrow \tilde{\varphi}'(z) \quad is \quad constant$$

Then, $\tilde{\varphi} = 0$ and $\tilde{S} = 0$.

B Proof of identification with function S^{-1}

After linearizing our problem by taking the inverse of function $S(.)$, we can identify our equations with weaker assumptions.

Assumption 15

$$E(U|X,W) = 0$$

Assumption 16 *(Y,Z) is strongly identified by (X,W)*

$$E[\mu(Y,Z)|X,W] = 0 \Rightarrow \mu = 0$$

Assumption 17 *Separability condition:*

$$a(Y) = b(Z) \Rightarrow a(.) = b(.) = constant$$

Let us write our equation using the function $H(.)$ one more time.

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

Assume that we have $H_1(.)$ and $\varphi_1(.)$ and $H_2(.)$ and $\varphi_2(.)$ such that

$$H_1(Y) = \varphi_1(Z) + X + U$$

$$H_2(Y) = \varphi_2(Z) + X + U$$

If we take expectations and substract the first one from the second:

$$E\left[(H_2(Y) - H_1(Y)) - (\varphi_2(Z) - \varphi_1(Z))|X, W\right] = 0$$

By *Assumption 9*:

$$(H_2 - H_1)(Y) = (\varphi_2 - \varphi_1)(Z)$$

By *Assumption 10*:

$$H_2 = H_1 + c \quad and \quad \varphi_2 = \varphi_1 + c$$

And finally by *Assumption 7* we get the identification as $c = 0$.

C Moment conditions for the GMM

Our first and second set of moment conditions are coming from the independency condition between the U and (X, W). Only difference is that we add some more conditions to the secon set.

1st set:

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

2nd set

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

For the third set of conditions, we used the efficient moment conditions of Florens and Carrasco (2004) which are in the form of $e^{\tau' x}$ which becomes e^{tx+ws} in our problem. Although one can get infinite number of moment conditions with this setting, we have used $t \in \{0, 0.25, 0.50, 0.75\}$ and $s \in \{0, 0.25, 0.50, 0.75\}$ and obtained 16 moment conditions. *3rd set*:

$$M3 = \left\{e^{tx+sw} : t \in \{0, 0.25, 0.50, 0.75\}\,; s \in \{0, 0.25, 0.50, 0.75\}\right\}$$

D Simulation Results

D.1 GMM estimation

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

Table 3: Simulation results for the $1st$ equation with the $3rd$ set of moments

	$\hat{k}$	$\hat{A}$	$\hat{a}$
Bias	0.0401	-2.2380	5.8534
Variance	0.0600	48.35	237.56
MSE	0.0600	48.37	238.61

Second equation

Table 4: 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 5: 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

Table 6: Simulation results for the $2nd$ equation with the $3rd$ set of moments

	$\hat{k}$	$\hat{B}$	$\hat{b}$
Bias	-0.0068	0.1118	0.0941
Variance	0.0162	3.5116	0.1062
MSE	0.0161	3.4890	0.1140

D.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

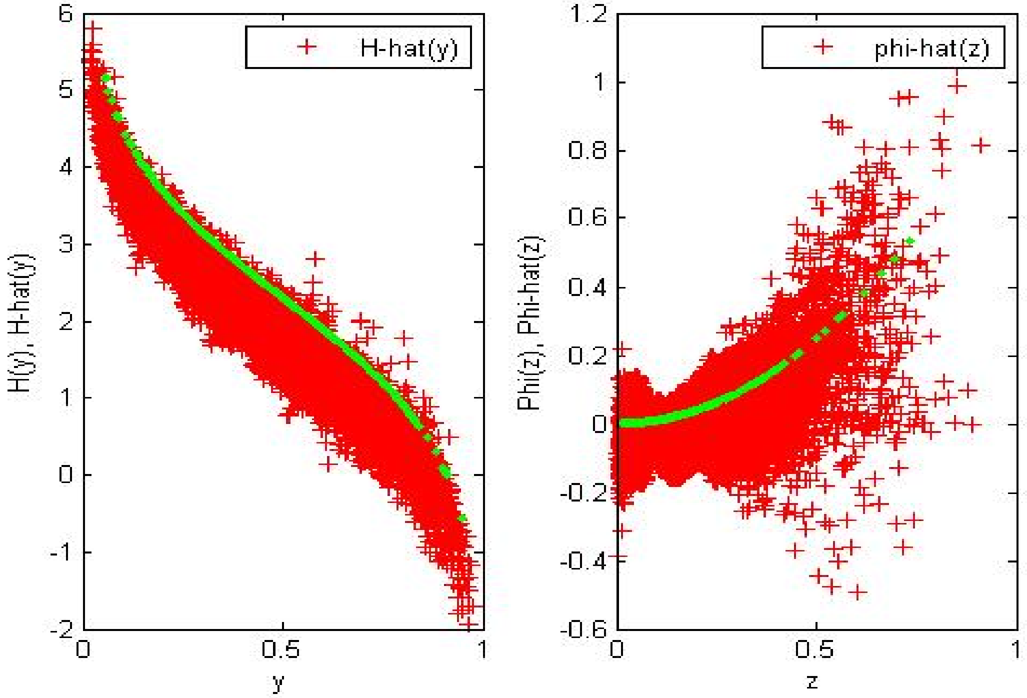

Figure 1: *Estimated functions for* $\alpha_H = 10^{-8}$ *and* $\alpha_\varphi = 0.09$

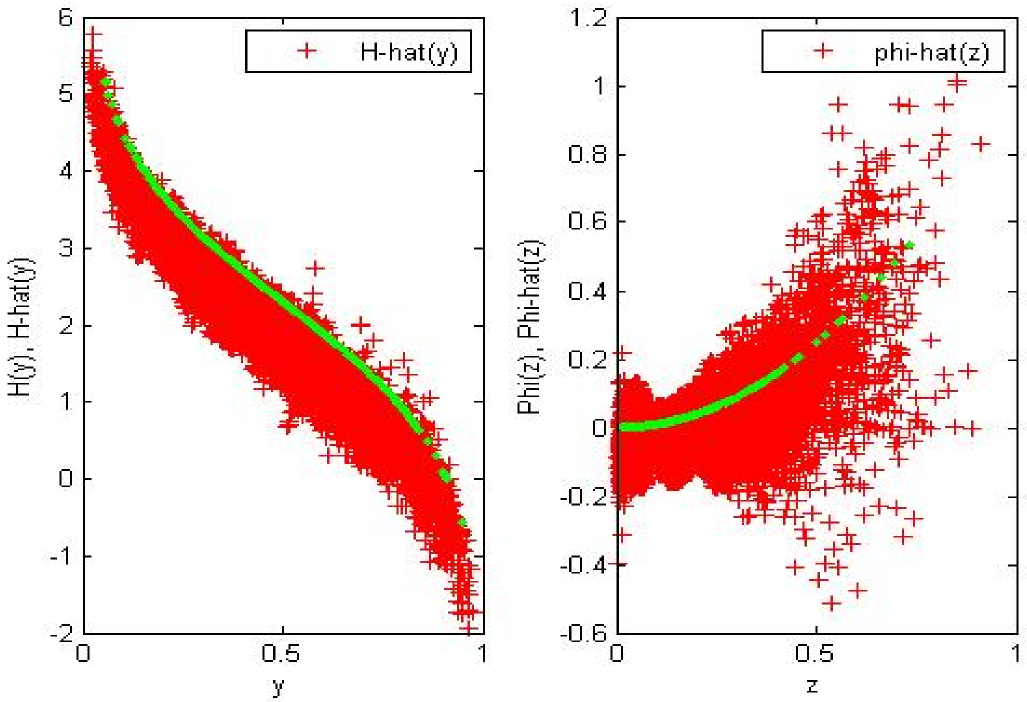

Figure 2: *Estimated functions for* $\alpha_H = 7 * 10^{-8}$ *and* $\alpha_\varphi = 0.091$

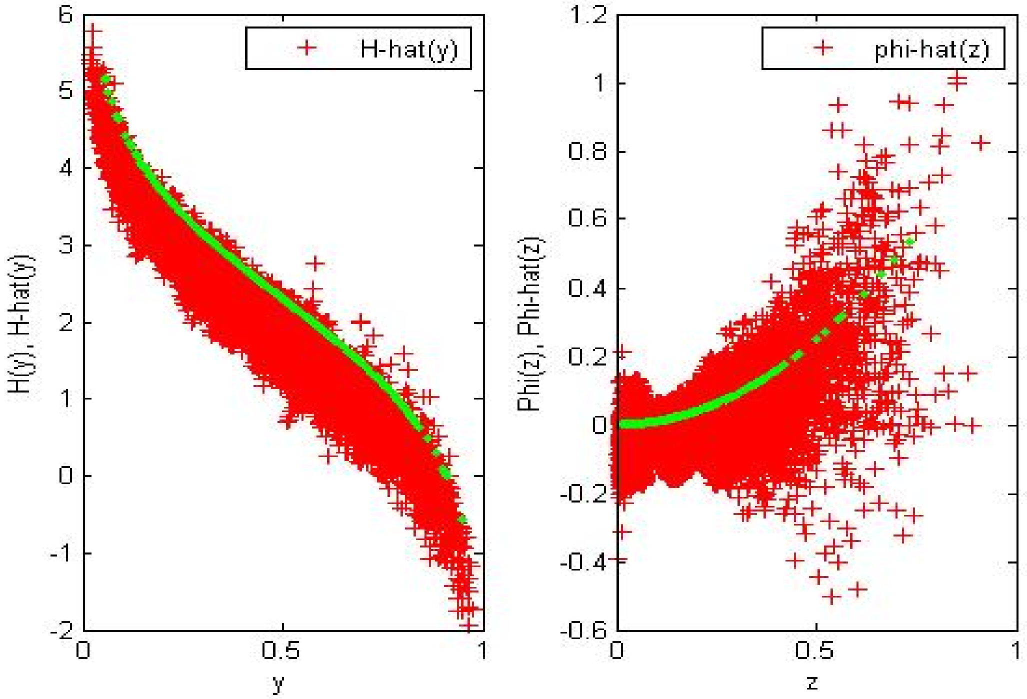

Figure 3: *Estimated functions for* $\alpha_H = 5 * 10^{-8}$ *and* $\alpha_\varphi = 0.092$

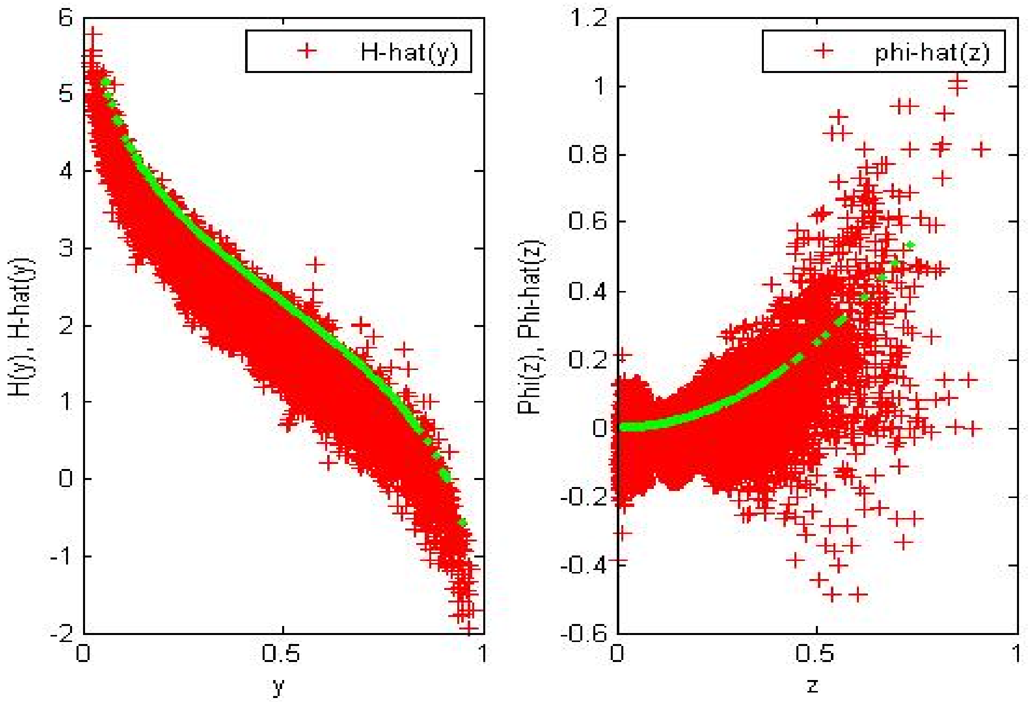

Figure 4: *Estimated functions for* $\alpha_H = 3 * 10^{-8}$ *and* $\alpha_\varphi = 0.093$

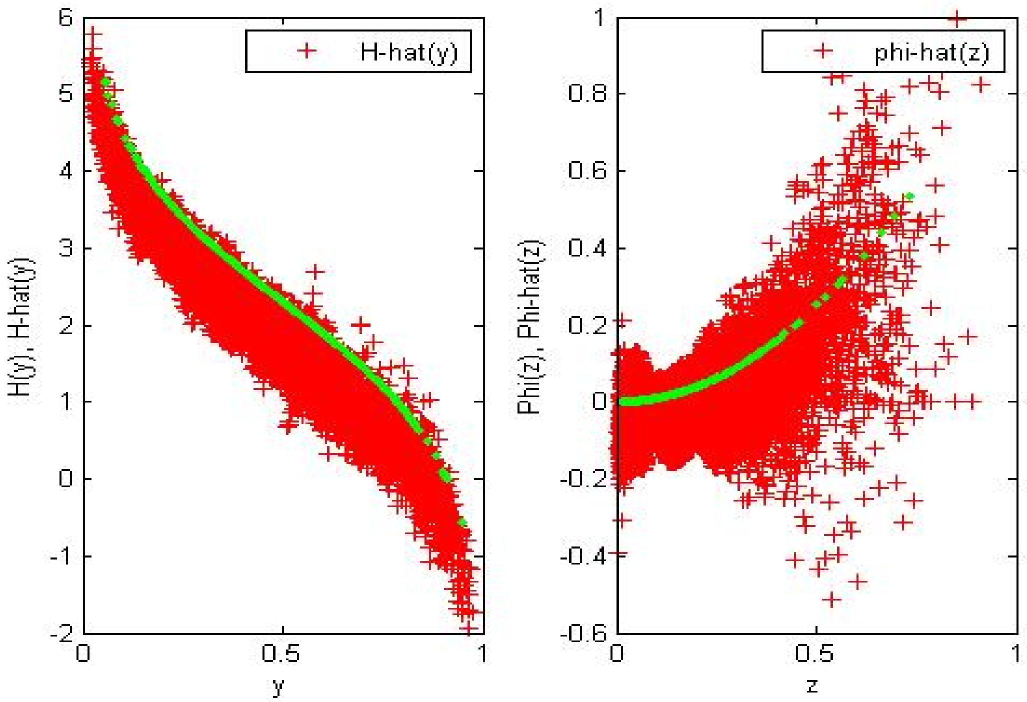

Figure 5: *Estimated functions for* $\alpha_H = 10^{-7}$ *and* $\alpha_\varphi = 0.095$

When we use a larger bandwidth, with the same values of alpha as in *figure 1* we got the results in *figures 6*.

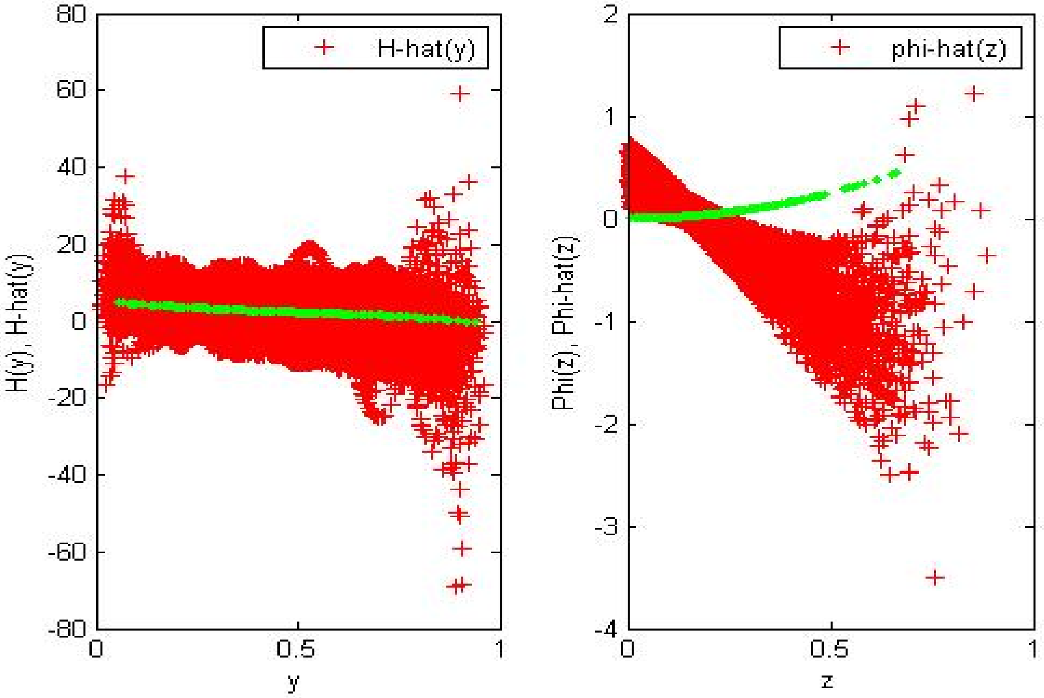

Figure 6: *Estimated functions for* $\alpha_H = 10^{-8}$ *and* $\alpha_\varphi = 0.09$ *with a larger bandwidth*

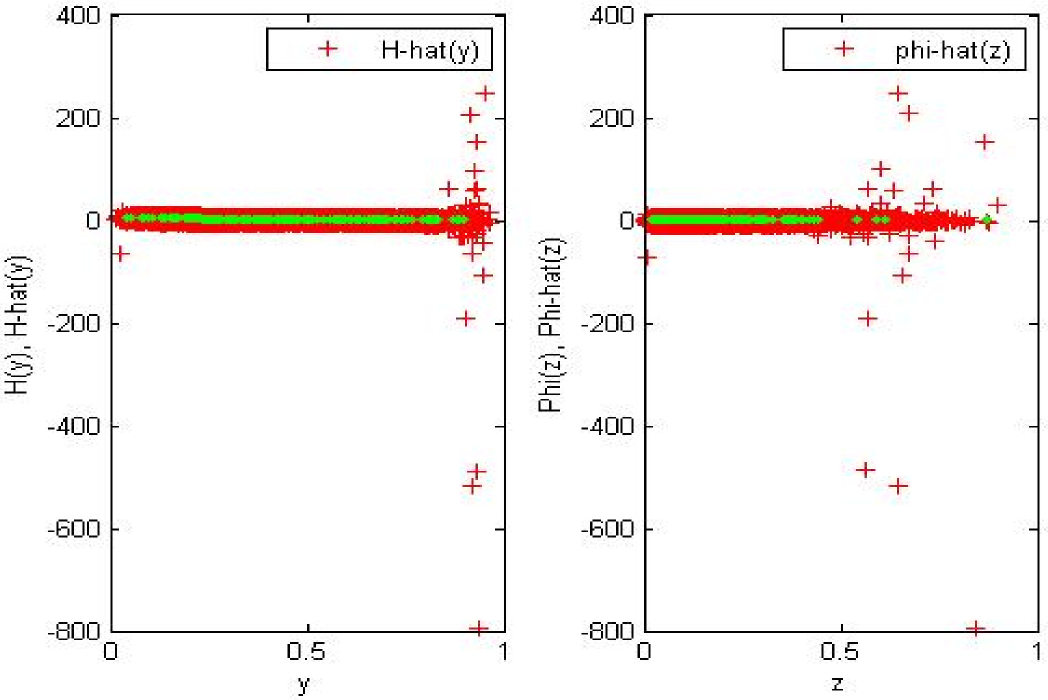

Figure 7: *Estimated functions for* $\alpha_H = 10^{-8}$ *and* $\alpha_\varphi = 10^{-8}$

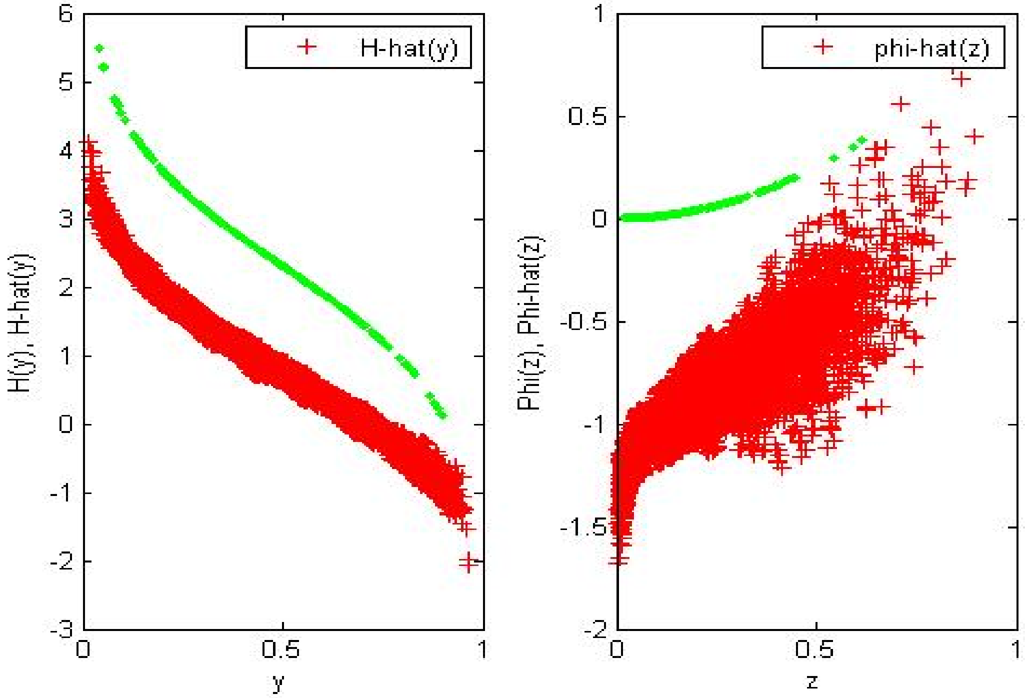

Figure 8: *Estimated functions for* $\alpha_H = 0.091$ *and* $\alpha_\varphi = 0.091$

References

Bass, F. M. (1969): "A New Product Growth for Model Consumer Durables," *Management Science*, 15, 215–227.

Brown, D. J., and R. L. Matzkin (1998): "Estimation of Nonparametric Functions in Simultaneous Equations Models with an Application to consumer Demand," Discussion paper, Northwestern University.

Darolles, S., J.-P. Florens, and E. Renault (2009): "Nonparametric Instrumental Regression," 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., and M. Carrasco (2004): "On the Asymptotic Efficiency of GMM," .

Groetsch, C. W. (1984): *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. Pitman, London.

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

Matzkin, R. L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339–1375.

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