Nonparametric Analysis of Two-Sided Markets

Senay Sokullu

Discussion Paper No. 12/628

October 3 2012

Department of Economics
University of Bristol
8 Woodland Road
Bristol BS8 1TN

Nonparametric Analysis of Two-Sided Markets*

Senay SOKULLU †
University of Bristol

October 3, 2012

Abstract

This paper considers an empirical semiparametric model for two-sided markets. Contrary to existing empirical literature on two-sided markets, we do not rely on linear network effects. Instead, network effects and probability distribution functions of net benefits of two sides are specified nonparametrically. The demand functions and the network effect functions of readers and advertisers are estimated by nonparametric IV estimation using a data set from German magazine industry. The ill-posed inverse problem faced during the estimation is solved by Tikhonov Regularization. We show that semiparametric specification is supported by the data and the network effects on readers' side are neither linear nor monotonic. With a numerical illustration we demonstrate that the mark-up of the magazine on readers' side is 27% higher with the nonlinearly specified network effects than in the case with linear network effects.

Keywords: Two-sided markets, Network externality, Nonparametric IV, Ill-posed inverse problems, Tikhonov Regularization

JEL Classification: C14, C30, L14

---

*I am indebted to my advisor Jean-Pierre Florens for his generous advice, support and encouragement. I thank Alexei Alexandrov, Richard Blundell, Lapo Filistrucchi, Marc Ivaldi, Gregory Jolivet, Toru Kitagawa, Costas Meghir, David Pacini, Jean-Marc Robin, Christophe Rothe, Anna Simoni, Tuba Toru, Frank Windmeijer and the seminar participants at University of Bristol and at CEMMAP for numerous helpful comments. All errors are mine.

†E-mail: *Senay.Sokullu@bristol.ac.uk*

1 Introduction

In the past 10 years a lot of work has been done both in terms of theory and in terms of empirics of two-sided markets; see, for instance, Rochet and Tirole (2003); Armstrong (2006); Rysman (2004); Kaiser and Wright (2006). However, most empirical papers use traditional parametric tools and specify the network effect with a constant parameter, see Argentesi and Filistrucchi (2007); Borsenberger et al. (2010). More precisely, the externality exerted on side one by side two is assumed to be linear in the number of agents on side two and vice versa. In this paper we develop a structural model for two sided markets where we do not specify any functional form for the network effects and we estimate them with nonparametric instrumental variables estimation. We solve the ill-posed inverse problem we come across during nonparametric IV estimation by regularizing it with Tikhonov Regularization scheme. The nonparametric specification of network effects allows us to capture nonlinearities and non-monotonicities in the network effect function with the increasing number of agents on the other side. The results of both nonparametric estimation and parametric estimation with a nonlinear specification show that network effects are not linear. The implications of this result for misspecified linear models are demonstrated with a numerical illustration.

The main feature of a two-sided market is the existence of externalities between the two sides of the market. More precisely, the benefit of agents on one side of the platform depends on the number of agents on the other side. However, as is pointed out in Rochet and Tirole (2003), a market with network externalities is a two-sided market if platforms can effectively cross-subsidize between different categories of end users. So, it is not only the interdependence of the sides to enter the platform but also the pricing structure of the platform which defines a market as two (or multi) sided. So far various industries have been examined under this setting: media, academic journals, dating agencies, credit cards, shopping malls, etc. For example, in the magazine industry, the decision of advertisers to advertise in a particular magazine depends on the circulation rate of that magazine. On the other side of the market, the readers may care about the advertising content of the magazine they buy. The platform, namely the magazine, can use this interdependence between the two sides when deciding on its pricing scheme. In this case, it is going to be more aggressive with the advertisers if they benefit more from contacting the readers on the platform.

While in some industries, the network effect is continuously increasing, in some other industries it is nonlinear and non-monotone. For example, in the credit card industry, where the credit card is the platform and buyers and sellers are the two sides, increasing the number of sellers who accept the credit cards would unambiguously increase the benefit and thus the number of buyers who hold a credit card. However, when we think of the magazine market, although it is found that readers get utility from seeing adverts in a magazine (see Borsenberger et al., 2010; Kaiser and Wright, 2006; Kaiser and Song, 2009), increasing the share of ad pages relative to content pages may start to give disutility to readers and thus make them leave the platform. In this case, the positive network externality for readers may become negative after some threshold level of ads. These network externalities play a crucial role in the platform's decision of pricing scheme since a price change for one side does not only affect the agents on that side, but it also affects the agents on the other side through the network effects. More concretely, the pricing equations of the platform can be explained by the usual *Lerner index* plus an extra term coming from the relationship with the other side of the market.[1] An empirical study where the network effects are specified linearly with a parameter may give misleading results if in fact we have nonlinear and non-monotonic network effects. In the case of nonlinear and monotonic network effects where the specification is linear, the misspecification will lead to quantitative errors, such as under or over estimation of markup which in turn may result in erroneous conclusion about the market power of the platform. The case of non-monotonicity is even more crucial for two-sided markets, since it is the nonmonotonic network effect functions that lead to the emergence of many platforms. For example, suppose that there is a publisher with 100 potential pages of advertisement and the readers start to get disutility after the $50th$ page. In such a situation, the optimal strategy for the publisher is to publish two similar magazines with 50 pages of advertisement in each so that it can extract the profits from both sides. In an empirical study with linear specification of network effects we may not uncover the true structure of the market and we may therefore make erroneous conclusions in our analysis.

In order to address this important issue, we set up a semiparametric model, in which we include the network effects nonparametrically in the demand functions. This way of specification lets us see if the increasing number of agents on one side may effect the participation decision of the agents on the other side negatively or positively. In addition to this, we do not specify any probability distribution function for the net benefits of agents, which leads to a nonparametrically specified demand functions. So, we estimate the network effect functions and the demand functions of the two sides nonparametically. To the best of our knowledge, neither the functional specification of the network effects nor the nonparametric approach has been used in the empirical two-sided market literature before. More broadly, nonparametric approaches have not yet been used in the empirical analysis of network industries.

Nonparametric estimation has gained a lot of attention as it has many advantages. First of all, the model is not approximated by a set of parameters and hence not affected by

[1] *The Lerner Index*, in other words, *markup*, shows the ratio between the profit margin and the profit maximizing price and is inversely proportional to the demand elasticity, i.e. $\frac{p-c}{p} = \frac{1}{\epsilon}$. (See Tirole, 1988)

any specification error. Secondly, the estimation results are independent of parametric restrictions and show us if the model is really supported by the data or not. We estimate the functions of interest by nonparametric IV estimation. However, it is well known that nonparametric IV estimation causes an ill-posed inverse problem which needs to be regularized. There are many papers in the literature that cope with this problem with different regularization schemes. Following the approach of Darolles et al. (2011), we regularize our inverse problem with *Tikhonov Regularization* and estimate the unknown density functions of the variables with kernels. Depending on the regularity of the function, this may give an optimal convergence rate or slower, but at the same time prevents the possible specification errors coming from the misspecification of the parametric form.

Furthermore we perform parametric estimation whose functional specifications are based on the nonparametric estimation results. More precisely, using the same structural equations, we approximate the unknown functions by nonlinear parametric forms, and estimate demand equations simultaneously. The results are consistent with what we have obtained in our nonparametric analysis.

Two main groups of literature are related to this paper. The first one is the two-sided markets literature. The theoretical literature on two-sided markets has focused on credit card markets, buyer-to-buyer platforms, academic journals and media. The model we use in this paper is related to that of Armstrong (2006), as it is more suitable to the magazine industry. Empirical studies have concentrated mostly on media. Kaiser and Wright (2006), Kaiser (2007), Kaiser and Song (2009) and Borsenberger et al. (2010) analyze the magazine industry, whereas Argentesi and Filistrucchi (2007) and Behringer and Filistrucchi (2010) are examples of the newspaper industry. In a very recent paper VanDalen (2010) analyzes the radio broadcasting industry in the Netherlands. In contrast to our paper, all these papers are using parametric tools and specify network effects linearly. Rysman (2004) uses a nonlinear specification for network effects in his paper where he examines the market for *Yellow Pages* directories. Kaiser (2007), Kaiser and Song (2009) and Borsenberger et al. (2010) consider a two sided market model for the magazine industry and analyze the effect of advertisements on readers. Kaiser (2007) finds that readers that are over 50 of age, are less educated and/or have low income like advertisements more compared to others. Using the same data set, Kaiser and Song (2009) conclude that readers of *Women's*, *Business and Politics* and *Car* magazines value ads, as ads in these magazines are informative. Borsenberger et al. (2010) arrive at a similar conclusion based on data from France. They show that readers of entertainment magazines value ads while the readers of information magazines do not. However, what we are interested in this paper is the level of advertisements from which readers start to get disutility or utility. With our nonparametric specification of network effects, we do not make *a priori* assumptions on the externality exerted on readers by advertisers, and we search for a threshold level from where this externality changes its sign. Our results show that the readers like advertisements. Moreover, we find that their benefits decrease when there are too few advertisements.

The second group of related literature is on nonparametric estimation methods with endogenous variables. Ai and Chen (2003), Newey and Powell (2003), Carrasco et al. (2007), Darolles et al. (2011), Feve and Florens (2009), Florens and Sokullu (2012) all deal with the problem of nonparametric IV estimation in the presence of endogenous variables. Although Ai and Chen (2003) and Newey and Powell (2003) use sieve methods and get over the problem of ill-posedness by putting bounds on integrals of higher order derivatives, all the other papers use kernel estimation and regularize the ill-posed inverse problem by *Tikhonov Regularization*. Florens and Sokullu (2012) is different from the other ones, in the sense that they develop their estimation technique for semiparametric transformation models. This paper also studies the nonparametric estimation of semiparametric transformation models, however, different from Florens and Sokullu (2012), all explanatory variables in our system are endogenous. Another related work to ours is that of Blundell et al. (2007). They present a nonparametric estimation technique and estimate shape invariant *Engel curves* using the U.K. Family Expenditure Survey.

The paper proceeds as follows. In *Section 2*, we introduce our model, derive the structural demand equations for readers and advertisers and pricing equations of the platform. We perform the nonparametric empirical analysis of the magazine industry in *Section 3*, where we describe the data, specify the model and prove identification, and we also define our estimation method and show its asymptotic properties. *Section 4*, presents the results of the empirical analysis with parametric tools while in *Section 5*, we present a numerical illustration to show the importance of misspecification of two-sided network effects. Finally, in *Section 6* concludes. All the proofs are presented in the appendices.

2 The Model

In this section we introduce our model taking into account the fact that an agent on one side will consider the number of agents on the other side when she makes her decision whether to enter the platform or not. In this paper we use a data from the magazine industry to perform our empirical analyses. The two-sided market model for the magazine industry is defined as follows: Each magazine produces content pages and ad pages for its readers and provides advertising outlet for firms who want to reach the readers of the magazine. It maximizes its profits by setting a cover price for the readers and an advertising rate for firms given the prices of its competitors. On the other hand, readers decide to buy the magazine or not and the firms decide to buy ad space from the magazine or not by looking at their net benefits. As it is a two-sided industry, the benefits of readers depend on the number of ad pages in the magazines and the benefits of advertisers depend on the number of readers of the magazine, as well as some other magazine characteristics.

We obtain the demand functions of readers and advertisers following the approach used by Larribeau (1993) and Feve et al. (2008). Let us begin with the reader side. We have $i = 1, ..., I$ readers. We assume that the readers are heterogeneous in their net benefit b_i^r of buying the magazine and these benefits are drawn from a continuous distribution. So, the reader i decides to buy magazine m if its net benefit is higher than a threshold (say, its net costs) $\underline{b}^r$:

$$b_i^r \geq \underline{b}^r(N^a, X, U)$$

where $\underline{b}^r(.)$ is the threshold benefit level for readers which is a function of the share of advertisers on the same platform N^a, the observable magazine characteristics X and the unobservable magazine characteristics U. All the readers whose net benefits are higher than this threshold will buy the magazine. Thus, the probability of buying the magazine and hence the market share of readers is given by:

$$N^r = P(b_i^r \geq \underline{b}^r(N^a, X, U)) = 1 - F^r(\underline{b}^r(N^a, X, U)) \tag{1}$$

where $F^r(.)$ is the cdf of net benefits of readers. We can rewrite the equation 1 as:

$$N^r = S^r(\underline{b}^r(N^a, X, U)) \tag{2}$$

where $S^r(.) = 1 - F^r(.)$ is the survival function. Equation (2) gives us the demand of readers for the magazine. We assume that it is strictly decreasing in the threshold benefit level. Given the distribution of benefits of readers, the higher the threshold is, the less readers buy the magazine. Furthermore, if the observable magazine characteristics are price and number of content pages, we expect the threshold benefit level to be increasing in cover price and decreasing in the number of content pages thus the demand is decreasing in cover price and increasing in number of content pages. The effect of share of ad pages is ambiguous, since the readers may like the ads or not, depending on the type of the magazine or depending on the readers' tastes. Note that, here we include the market share of advertisers as an explanatory variable in the reader demand. What one would expect is that readers instead care about the total number of ad pages and/or the ratio of ad pages to content pages in a magazine. The reason we use the market share in the modeling is to have a fully simultaneous system, i.e, $N^r = S^r(N^a, ., .)$ and $N^a = S^a(N^r, ., .)$. Moreover, in our application we show that our results on readers' side are robust to different definition of advertising amount in the magazine.

Now, let us consider the advertisers. We have $j = 1, ..., J$ advertisers. Each of them has a net benefit b_j^a from advertising in a magazine and these net benefits are drawn from a continuous distribution $F^a(.)$. They advertise in the magazine if their net benefit is higher than $\underline{b}^a$, i.e:

$$b_j^a \geq \underline{b}^a(N^r, W, V)$$

where $\underline{b}^a(.)$ is the threshold benefit level for advertisers and it is a function of the share of readers of the magazine N^r, observable magazine characteristics for advertisers W and unobservable magazine characteristics V. Like in the case of readers, the probability of advertising in the magazine and thus the share of advertisers who join the magazine is given by:

$$N^a = P(b_j^a \geq \underline{b}^a(N^r, W, V)) = 1 - F^a(\underline{b}^a(N^r, W, V)) \tag{3}$$

We can rewrite the equation 3 as:

$$N^a = S^a(\underline{b}^a(N^r, W, V)) \tag{4}$$

where $S^a(.) = 1 - F^a(.)$. Equation (4) is the demand equation of firms for advertising in the magazine. It is strictly decreasing in the threshold benefit function. The threshold benefit function is expected to be increasing in the ad rate and decreasing in the share of readers. So, more firms would like to advertise in a magazine with a higher readership and a lower ad rate.

Now, we can write the demand system for the magazine:

$$N^r = S^r(\underline{b}^r(N^a, X, U))$$

$$N^a = S^a(\underline{b}^a(N^r, W, V))$$

Given the demand equations of both sides, the magazine chooses its cover price and advertising rate to maximize its profit:

$$\max_{p^r, p^a} \Pi = p^r N^r M^r + p^a N^a M^a - C(N^r M^r, N^a M^a, IP) - FC \tag{5}$$

where $p^{(i)}$ and $M^{(i)}$, $i = a, r$ are price and market size for both sides, C is the cost function which depends on the number of agents on each side as well as other cost variables such as input prices, IP, and FC is the fixed costs. The maximization problem in equation (5) gives the following pricing equations:

$$\frac{p^r - \frac{\partial C}{\partial N^r}}{p^r} = -\frac{1}{\epsilon_{N^r}} - \frac{\left(p^a - \frac{\partial C}{\partial N^a}\right)\frac{\partial N^a}{\partial p^r}}{\frac{\partial N^r/\partial p^r}{p^r}} \tag{6}$$

$$\frac{p^a - \frac{\partial C}{\partial N^a}}{p^a} = -\frac{1}{\epsilon_{N^a}} - \frac{\left(p^r - \frac{\partial C}{\partial N^r}\right)\frac{\partial N^r}{\partial p^a}}{\frac{\partial N^a/\partial p^a}{p^a}} \tag{7}$$

It should be noted that both equation (6) and equation (7) are modified versions of the *Lerner Index*, in the sense that they also include the network externalities coming from the two-sidedness of the industry. For the reader side, the mark-up (the term on the left hand side), which is the ability of the magazine to price over its cost, depends on the inverse price elasticity of the readers, $1/\epsilon_{N^r}$ as well as on a second term. This second term captures the externality that readers have over advertisers. If they exert a positive externality on advertisers, the magazine charges readers a lower price compared to a situation where this externality is ignored. The intuition is simple. By lowering the cover price, the magazine can attract more readers which in turn attract more advertisers through network externalities, thus increasing the profits of the magazine. For the advertisers, the same holds. The mark-up depends not only on the inverse price elasticity of the advertisers, $1/\epsilon_{N^a}$ but also on the externality they exert on readers. If the readers don't like ads, the magazine prices ads higher than otherwise, so that it can subsidize the readers for the disutility caused by ads. In return, the readers can be kept in the platform and the profits are made from the advertisers.

So, it is important to identify the sign and magnitude of the externalities as they play a crucial role in the pricing. For example, in the case where readers like ads as long as they do not pass a certain share of the magazine content, the magazines pricing scheme will not be the same below and above of that certain level of ads. It can extract surplus from both sides when they both exerts a positive externality on each other, thus increasing it profits. However, when advertisers exert a negative externality on readers, the profits will be extracted from advertisers and the readers will be subsidized by paying a price lower than in the "no externality case". In the next section, where we conduct an empirical analysis of the industry, we will specify the network effects as unknown functions to be able to see if the exerted externality changes its sign with the increasing number of ad pages or not.

3 Empirical Analysis of Magazine Industry : A Nonparametric Approach

In this section, we present a nonparametric empirical analysis of the magazine industry. We will specify network effects nonparametrically, and estimate the semiparametric demand equations using a nonparametric IV estimation.

3.1 Model Specification

In the studies of magazine markets, two approaches have been adopted for externality between advertisers and readers. In the first approach it is assumed that there is a positive externality of advertisement on readers whereas in the second one the externality is assumed to be negative. Kaiser (2007), uses data on the magazine industry in Germany and shows that readers are in fact ad-lovers. On the other hand, Borsenberger et al. (2010) conclude that the readers of entertainment magazines like advertisements while the readers of information magazines do not. Their result is quite intuitive. It is natural to expect that the readers of women's magazines value advertisements because ads in these types of entertainment magazines are in fact informative. In contrast, the information magazines are valued for their content and advertisements that are not related to their content can be perceived as nuisance. In this section our aim is to see if this network externality between the two sides changes sign as the share of agents on one side increases, rather than to see if it is negative or positive. So, instead of making a linear parametric specification for these externalities, we specify it with a function to be able to capture the variation in network externality with variation in the number of advertisements.

Evans and Schmalensee (2008) and Argentesi and Ivaldi (2005) point out that failure to account for network externalities in two-sided platforms can lead to serious errors in antitrust analysis. Correct specification of these network effects is important for the same reason. However, none of the aforementioned papers on the magazine industry allows for nonlinear and nonmonotone network effects. It is straightforward to see that in case of nonlinearity or non-monotonicity of the network effect functions, the results of Kaiser and Song (2009) and Kaiser and Wright (2006) regarding the elasticities or mark-ups, would not be the same.[2] Rysman (2004) adopts a nonlinear specification for network effect functions.

[2]Suppose that the network effects are specified as second order polynomials in Kaiser and Song (2009), so that the mean utility is given by: $\delta_{jt} = X_{jt}\beta + \theta_0 + \theta_1 Ad_{jt} + \theta_2 Ad_{jt}^2 - \alpha p_{jt} + \gamma_t + \eta_j + \xi_{jt}$ instead of $\delta_{jt} = X_{jt}\beta + \theta Ad_{jt} - \alpha p_{jt} + \gamma_t + \eta_j + \xi_{jt}$. Then the advertising elasticity of demand would not only depend on θ_1 but also on $\theta_2 Ad_{jt}$, thus the magnitude as well as the effect would be ambiguous in case these parameters are estimated to have different signs.

However, his specification does not allow for non-monotonicity. It can be easily shown that, if non-monotonicities exist in the network effect functions, the market equilibrium condition he derives would be harder to satisfy, i.e. it may not be satisfied for all values of the advertisements. Thus, his results may not hold anymore if the network effect of advertisers on consumers are nonmonotone. We demonstrate with a simple numerical example in *Section 5* that a linear specification instead of a quadratic one results in 27% difference in the mark-up of the magazine on the reader's side.

In this semiparametric model specification, we make as few parametric approximations as we can. First of all, we make no assumption on the family of distribution functions of net benefits of readers and advertisers. Secondly, we assume that network effects are given by some unknown functions, $\varphi(N^a)$ and $\psi(N^r)$. Finally we specify the threshold benefit functions $\underline{b}^r$ and $\underline{b}^a$ as linear functions of network externalities and platform characteristics. Then, the system of demand equations are given by the following:

$$N^r = S^r(\varphi(N^a) + X\beta + U) \tag{8}$$

$$N^a = S^a(\psi(N^r) + W\gamma + V) \tag{9}$$

We are interested in estimating the network externality between the two sides, hence for simplicity, we consider just one platform characteristic, price, whose coefficient is normalized to one for identification. The reason to use just one characteristic is that since we use kernels in estimation, increasing the dimension of instruments and/or exogenous variables complicates the estimation process. Note that this way of specification does not allow us to analyze the elasticities however in this paper we are interested in identifying the shape of the network effect functions and we can do this normalization without loss of generality. This final specification gives us the equations we are going to estimate:

$$N^r = S^r(\varphi(N^a) + P^r + U) \tag{10}$$

$$N^a = S^a(\psi(N^r) + P^a + V) \tag{11}$$

where P^r is the cover price of the magazine and P^a is the ad rate of a single page.[3]

We are going to estimate the functions of interest, namely, S^r, S^a, φ and ψ nonparamet-

[3] In the magazine industry there are different price levels for the readers depending on the way to obtain the magazine, i.e., buying from kiosk or having a subscription. However, we do not observe this in our sample, and we use only one price, cover price. Borsenberger et al. (2010) observe also the subscribers in their data set and use this information. They find that the subscribers are less price sensitive. Moreover in our data set all advertisements are assumed to be one single page and the ad rate is the average price of giving one page of advertisement in the magazine

rically.

3.2 Identification

Before proceeding to discussion of identification, let us introduce our variables and some notation.

N^r, $N^a \in \mathbb{R}$ are the endogenous market shares of the magazine on the readers' and advertisers' side, respectively. The platform characteristics for readers and advertisers, which we denote by X and W in equations (8) and (9), can be endogenous or exogenous. These characteristics generally include prices on both sides, number of content pages, magazine segment and frequency of the magazine. Since we are interested in estimating the network effect functions $\varphi(.)$ and $\psi(.)$, for simplicity, we include just one platform characteristic in each function, namely cover price (P^r) and ad rate (P^a). Thus, $P^r, P^a \in \mathbb{R}$, are endogenous. $Z^r, Z^a \in \mathbb{R}^2$ are instruments for each equation and $Z = \{Z^r, Z^a\}$. For the moment we assume that we have valid instruments. We introduce the instruments in *Application*. Finally, unobservable characteristics for each side, U and V are scalars as well.

We assume that (N^r, N^a, P^r, P^a, Z) generate a random vector, Ξ, which has a cumulative distribution function F. Then for each F, we can define subspaces of our variables as $L^2_F(N^r), L^2_F(N^a), L^2_F(P^r)$, $L^2_F(P^a)$ and $L^2_F(Z)$ which belong to a common Hilbert space. That is, $L^2_F(Y)$ denotes the subspace of L^2_F of real valued functions depending on Y only. In the sequel, we use the notation L^2_Y to denote the $L^2_F(Y)$.

Now we can state the needed assumptions for identification.

Assumption 1 ***Strict monotonicity.*** *The survival function S^q, $q \in \{r, a\}$ is strictly decreasing in its arguments.*

As S is a survival function, we know that it is decreasing, however, by making the assumption of "strictly decreasing", we guarantee to have an inverse of it, which we will use for identification and estimation. Thus, using assumption (1), we can rewrite the system of equations in (10) and (11) as:

$$H^r(N^r) = \varphi(N^a) + P^r + U \tag{12}$$

$$H^a(N^a) = \psi(N^r) + P^a + V \tag{13}$$

where $H^q(N^q) = (S^q)^{-1}(N^q)$ for $q = a, r$.

Assumption 2 ***Conditional mean independence.*** $\mathbb{E}[U|Z^r] = 0$ *and* $\mathbb{E}[V|Z^a] = 0$

Assumption 3 ***Completeness.*** *(N^r, N^a) are strongly identified by Z^q for $q = a, r$, i.e.:*

$$\forall m(N^r, N^a) \in L^2_{N^r} \times L^2_{N^a}, \quad \mathbb{E}[m(N^r, N^a)|Z^q] = 0 \quad \Rightarrow \quad m(N^r, N^a) = 0 \quad a.s \quad forq \quad \in \{a, r\}$$

Assumption 4 ***Measurable separability.*** *N^r and N^a are measurably separable:*

$$\forall m \in L^2_{N^r}, l \in L^2_{N^a}, m(N^r) = l(N^a) \Rightarrow m(.) = l(.) = constant$$

Assumption 5 ***Normalization.***

$$\forall l \in L^2_{N^a}, l(.) = constant \Rightarrow constant = 0$$

For simplicity, we will assume that $\varphi(.)$ and $\psi(.)$ are normalized by the conditions $\mathbb{E}(\varphi(N^a)) = 0$ and $\mathbb{E}(\psi(N^r)) = 0$. Under this assumption, the parametric space we consider is:

$$\mathcal{E}_0 = (H, \phi) \in L^2_{N^r} \times L^2_{N^a} \quad such \quad that \quad \mathbb{E}[\phi] = 0$$

Assumption(2) is just a conditional mean independence condition and it holds whenever Z^r and Z^a are exogenous. *Assumption(3), Completeness* is the nonparametric counterpart of the rank condition in parametric IV estimation and it is also referred as *complete statistic* in the statistics literature. It is a condition on the power of Z^r and Z^a to identify $H^r(.)$ and $\varphi(.)$ and $H^a(.)$ and $\psi(.)$, respectively. Intuitively, it means that there is no function of N^r and N^a that is not correlated with any function of Z^r and Z^a. The completeness assumption is an assumption on the distribution of the endogenous variables conditionally on the instruments. Although, in this paper we take it as given, further reference on the primitive conditions for the completeness can be found in DHaultfoeuille (2011), Andrews (2011) and Hu and Shiu (2011). As it is shown by Hu and Shiu (2011), our completeness condition on the multivariate distribution function of the endogenous variables can be obtained by the completeness on univariate distribution functions and thus it is not very restrictive. *Assumption(4), Measurable separability*, is needed to distinguish $H^r(.)$ and $\varphi(.)$ and $H^a(.)$ and $\psi(.)$. It says that two functions $m(N^r)$ and $l(N^a)$ can be equal only if they are equal to a constant. In other words, it states that there is no exact relation between N^r and N^a. So, for the reader demand equation, *Assumption(4)* holds whenever $P^r + U$ varies independently of N^a and for the advertiser demand equation, it holds whenever $P^a + V$ varies independently of N^r. It is not hard to verify measurable separability in our model. In the model, we already state that reader demand depends on the share of advertisers as well as some observable and unobservable magazine characteristics for readers. Hence, we do not make an unrealistic assumption by measurable separability, as it is natural to expect that some magazine characteristics for readers vary independently of the share of advertisers. Moreover, with the same reasoning, it is natural to expect that the magazine characteristics for advertisers vary independently of the share of readers. Finally, *Assumption(5)* is just a normalization assumption for identification.

Theorem 1 *Under the assumptions 1-5, the functions S^r, S^a, φ and ψ are identified.*

In this paper we are considering a fully recursive system of equations. For this reason we now state some extra assumptions that guarantee the existence of the reduced form solution of the system. In other words, we show that N^r and N^a can be explained by the other variables of the system for given cover price and ad rate. Blundell and Matzkin (2010) also discuss the existence of reduced form solution to the system of equations where they investigate the control function approach in nonparametric nonseparable simultaneous equations models.

Remember that our structural system is given by equations (10) and (11)

$$N^r = S^r(\varphi(N^a) + P^r + U)$$

$$N^a = S^a(\psi(N^r) + P^a + V)$$

Let us now state some more assumptions:

Assumption 6 *For all values of N^r, N^a, P^r, P^a, U and V, the functions S^r and S^a are continuously differentiable.*

This assumption requires that the net benefits of readers and advertisers are distributed continuously so that the survival functions will be continuously differentiable on the interval $(0, 1)$.

Assumption 7 *For all values of N^r, N^a, P^r, P^a, U and V, we have:*

- $\frac{\partial S^r}{\partial N^a}\frac{\partial S^a}{\partial N^r} < 0$ *or equivalently*
- $\frac{\partial S^r}{\partial \varphi}\frac{\partial \varphi}{\partial N^a}\frac{\partial S^a}{\partial \psi}\frac{\partial \psi}{\partial N^a} < 0$

We know that $\partial S^r/\partial \varphi$ and $\partial S^a/\partial \psi$ are negative since the survival function is decreasing in its arguments. So the assumption reduces to

$$\frac{\partial \varphi}{\partial N^a}\frac{\partial \psi}{\partial N^r} < 0$$

This actually means that in the equilibrium, the network externalities have the opposite effects.

Now, we can state the theorem for the existence of a reduced form solution:

Theorem 2 *Under assumptions 1,6, 7 and for given cover price and ad rate, there exist unique functions h^r and h^a representing the structural model in equations (10) and (11), such that:*

$$N^r = h^r(P^r, P^a, U, V) \quad \textit{and} \quad N^a = h^a(P^r, P^a, U, V)$$

Remark 3 *In this section we present the identification of a separable model. Identification of a nonseparable model can also be obtained but for this we need stronger assumptions, i.e., instead of conditional mean independence, we need to assume full independence between the unobservable random terms and exogenous variables. Moreover, solutions for this kind of models lead to a nonlinear inverse problem, which is even harder to solve. In addition, we cannot identify the network effect functions separately in a nonseparable model. For these reasons, we decided to maintain a separable model for the application. Nonetheless, we present our model under a nonseparable framework and discuss its identification and estimation in Appendix A.*

3.3 Nonparametric Estimation of Semiparametric Transformation Equations

We estimate equations (12) and (13) with the nonparametric instrumental variable estimator for semiparametric transformation models. As already stated, nonparametric IV regression has gained a lot of attention recently. The model we present in this section is different from those found in the existing literature in the sense that it covers a very general case. We allow for nonparametric specification on the right and left hand side of the transformation model, as well as for endogeneity and use the mean independence condition rather than full independence. Although Horowitz (1996) and Linton et al. (2008) study the nonparametric estimation of transformation models, none of them considers the case of nonparametric specification on both sides of the regression equation. Our method is very similar to that of Florens and Sokullu (2012), since both are nonparametric estimation of semiparametric transformation models which allow for nonparametric specification on both sides of the equation. The difference between the two models is that in our model all explanatory variables are endogenous. Nonetheless, the estimation method we present below follows the approach of Florens and Sokullu (2012) closely.

We follow a limited information approach while presenting nonparametric IV estimation for transformation equations. In other words, we estimate the demand system given by equations (12) and (13) equation by equation. For the rest of this section, we continue to present the estimation method using equation (12) only. Extension to equation (13) is straightforward.

By the assumptions 1 and 2, we can write our estimation problem as:

$$\mathbb{E}[H^r(N^r) - \varphi(N^a) - P^r|Z^r] = 0$$

From now on, for ease of presentation we will work with operators. Let us define the following operator:

$$T^r : \mathcal{E}^r = \left\{ L^2_{N^r} \times \tilde{L}^2_{N^a} \right\} \mapsto L^2_{Z^r} : T^r(H^r, \varphi) = \mathbb{E}[H^r - \varphi|Z^r]$$

where

$$\tilde{L}^2_{N^a} = \left\{ \varphi \in L^2_{N^a} : \mathbb{E}(\varphi) = 0 \right\}$$

We use this projected space in order to satisfy the normalization assumption for identification, *Assumption 5*. Without this constraint on the space we cannot identify the function φ. The inner product is defined as:

$$\langle (H^r_1, \varphi_1), (H^r_2, \varphi_2) \rangle = \langle H^r_1, H^r_2 \rangle + \langle \varphi_1, \varphi_2 \rangle$$

The adjoint operator of T^r, T^{r*} satisfies:

$$\langle T^r(H^r, \varphi), \xi \rangle = \langle (H^r, \varphi), T^{r*}\xi \rangle$$

for any $(H^r, \varphi) \in \mathcal{E}^r$ and $\xi \in L^2_{Z^r}$. From the equality above it follows immediately that

$$T^{r*}\xi = (\mathbb{E}[\xi|N^r], \mathbb{P}\mathbb{E}[\xi|N^a])$$

where $\mathbb{P}$ is the projection operator.[4]

Now we can rewrite our estimation problem using the operator notation:

$$T^r(H^r(N^r), \varphi(N^a)) = f^r \tag{14}$$

where $f^r = \mathbb{E}(P^r|Z^r)$. This equation is a *Fredholm integral equations of the first kind* and the solution needs the inversion of the operator T^r. Thus it causes an ill-posed inverse problem. More precisely, this equation violates one of the definitions of the well-posedness

[4]It should be noted that we make a normalization, we need a projection of the non-normalized function on the normalized space. See Florens and Sokullu (2012).

of a problem.[5] The solution does not continuously depend on the data, so it is unstable. The reason is that the operator T^r has infinitely many eigenvalues in the neighborhood of zero, which makes $(T^r)^{-1}$ discontinuous. As a result, a very small change in the value of f^r may lead the solution to explode. Intuitively, in the finite dimensional case, this situation amounts to having a matrix M with zero eigenvalues and thus noninvertable.

To solve this ill-posed problem, we need to regularize it, i.e., we need to modify the operator such that the solution is not unstable, and such that this amount of modification approaches to zero as the sample size increases. For this, we choose the *Tikhonov Regularization* scheme. Basically, under this regularization scheme, the norm of the solution is controlled by a penalty term, γ, which is called *regularization parameter*. The *Tikhonov Regularization* requires the inversion of an n-by-n matrix, where n is the sample size, so it is not practical to use with very large samples. Since we have a relatively small sample, we use this scheme which is easier to apply in our case. The ill-posed inverse problem literature offers other regularization schemes, as well. See Carrasco et al. (2007) and Carrasco (2008).[6]

The regularized solution to the identifying relation in (14) is given by the following minimization program:

$$\min_{H^r,\varphi} \|T^r(H^r,\varphi) - f^r\|^2 + \gamma_n^r \|(H^r,\varphi)\|^2$$

where $\gamma_n^r > 0$ and γ_n^r converges to zero at a suitable rate. Hence,

$$(H^r(N^r), \phi(N^a))' = (\gamma_n^r I + T^{r*}T^r)^{-1} T^{r*} f^r \tag{15}$$

where I is the identity operator in $L^2_{N^r} \times L^2_{N^a}$. We can write the solution in (15) as follows:

$$(\gamma_n^r I + T^{r*}T^r)(H^r(N^r), \phi(N^a)) = T^{r*} f^r$$

Equivalently:

$$\begin{pmatrix} \gamma_n^r H^r + \mathbb{E}\left[\mathbb{E}(H^r|Z^r)|N^r\right] - \mathbb{E}\left[\mathbb{E}(\varphi|Z^r)|N^r\right] \\ -\gamma_n^r \varphi + \mathbb{PE}\left[\mathbb{E}(H^r|Z^r)|N^a\right] - \mathbb{PE}\left[\mathbb{E}(\varphi|Z^r)|N^a\right] \end{pmatrix} = \begin{pmatrix} \mathbb{E}[\mathbb{E}(P^r|Z^r)|N^r] \\ \mathbb{PE}[\mathbb{E}(P^r|Z^r)|N^a] \end{pmatrix} \tag{16}$$

To explain the implementation of the defined method, suppose that we have an i.i.d.

[5]As defined in Engl et al. (1996), a problem is well-posed if the definitions below hold:
(i) For all admissible data a solution exist.
(ii) For all admissible data the solution is unique.
(iii) The solution continuously depends on the data.

[6]For large samples, instead of *Tikhonov Regularization*, *Landweber-Friedman Regularization* can be used, which is an iterative scheme and does not require the inversion of an n-by-n matrix.

sample of $(N_i^r, N_i^a, P_i^r, P_i^a, Z_i)$, $i = 1, .., n$. As we do not know the true distribution of our variables, we need to replace the conditional expectations with their empirical counterparts. The multivariate kernel that is used in the implementation is defined as follows (See Florens et al., 2009):

Definition 1 *For all $w = (w_1, ..., w_q) \in \mathbb{R}^q$, K is a multiplicative kernel of order τ, i.e., $K(w) = \prod_{j=1}^q g(w_j)$ where g is a univariate, continuous, bounded function such that*

$$\int g(u)du = 1 \quad \int u^j g(u)du = 0$$

for all $j = 1, ..., \tau - 1$, and there exist two finite constants s_K^τ and C_K such that

$$\int u^\tau g(u)du = s_K^\tau, \quad \int g(u)^2 du = C_K$$

Then we can rewrite the system of equations with kernels. For the system in (16), let A_{Z^r} be the matrix whose (i,j)th element is:

$$A_{Z^r}(i,j) = \frac{K_{Z^r,h_{Z^r}}\left(Z_i^r - Z_j^r\right)}{\sum_j K_{Z^r,h_{Z_r}}\left(Z_i^r - Z_j^r\right)}$$

Let A_{N^r} and A_{N^a} be the matrices with the (i,j)th elements:

$$A_{N^r}(i,j) = \frac{K_{N^r,h_{N^r}}\left(N_i^r - N_j^r\right)}{\sum_j K_{N^r,h_{N^r}}\left(N_i^r - N_j^r\right)}$$

$$A_{N^a}(i,j) = \frac{K_{N^a,h_{N^a}}\left(N_i^a - N_j^a\right)}{\sum_j K_{N^a,h_{N^a}}\left(N_i^a - N_j^a\right)}$$

for some bandwidth parameters h_{N^r}, h_{N^a} and h_{Z^r}. Moreover let A_p be the matrix with $\frac{n-1}{n}$ on the diagonal and $-\frac{1}{n}$ elsewhere. We can rewrite the system in (16) as:

$$\begin{pmatrix} \gamma_n^r \hat{H}^r + A_{N^r}A_{Z^r}\hat{H}^r - A_{N^r}A_{Z^r}\hat{\varphi} \\ -\gamma_n^r\hat{\varphi} + A_pA_{N^a}A_{Z^r}\hat{H}^r - A_pA_{N^a}A_{Z^r}\hat{\varphi} \end{pmatrix} = \begin{pmatrix} A_{N^r}A_{Z^r}P^r \\ A_pA_{N^a}A_{Z^r}P^r \end{pmatrix}$$

Then the estimated functions are given by:

$$\begin{pmatrix} \hat{H}^r \\ \hat{\varphi} \end{pmatrix} = \begin{pmatrix} \gamma_n^r I + A_{N^r}A_{Z^r} & -A_{N^r}A_{Z^r} \\ A_pA_{N^a}A_{Z^r} & -(\gamma_n^r I + A_pA_{N^a}A_{Z^r}) \end{pmatrix}^{-1} \begin{pmatrix} A_{N^r}A_{Z^r}P^r \\ A_pA_{N^a}A_{Z^r}P^r \end{pmatrix} \tag{17}$$

Equation (17) is a system of $2n$ equations in $2n$ unknowns which means that we can recover

$\hat{H}^r$ and $\hat{\varphi}$, hence $\hat{S}^r$. For the estimation of the second equation of our system the procedure is exactly the same and $\hat{H}^a$ and $\hat{\psi}$ are given by:

$$\begin{pmatrix} \hat{H}^a \\ \hat{\psi} \end{pmatrix} = \begin{pmatrix} \gamma_n^a I + A_{N^a} A_{Z^a} & -A_{N^a} A_{Z^a} \\ A_p A_{N^r} A_{Z^a} & -(\gamma_n^a I + A_p A_{N^r} A_{Z^a}) \end{pmatrix}^{-1} \begin{pmatrix} A_{N^a} A_{Z^a} P^a \\ A_p A_{N^r} A_{Z^a} P^a \end{pmatrix} \quad (18)$$

3.4 Consistency and Rate of Convergence

Consistent estimation of the functions of interest depends on the consistent estimation of the defined operators $T^{r*}T^r$ and $T^{r*}f^r$ (*and for the second equation* $T^{a*}T^a$ and $T^{a*}f^a$). To prove the consistency, we are going to make a set of assumptions about the operators.[7]

Assumption 8 ***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 $\Phi = (H^r, \varphi)$, $\{\lambda_j, \phi_j, \psi_j\}$ is the singular system of operator T^r.[8],[9]

With this assumption we can define a regularity space for our functions. In other words, we can say that the unknown value of $\Phi_0 = (H_0^r, \varphi_0)$ belongs to the space Ψ_ν where

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

In fact, assuming that $\Phi_0 \in \Psi_\nu$ just adds a smoothness condition to our functional parameter of interest. It amounts to assuming that the functions H^r and φ have regularity $\nu > 0$.[10] This regularity captures both the properties of solution of Φ and the operator T^r, for that reason, in Hall and Horowitz (2005) and Horowitz (2009), two different assumptions are made instead of using the source condition. As was pointed out by Carrasco et al. (2007), this regularity assumption will give us an advantage in calculating the rate of convergence of the regularization bias.[11]

[7] In this part we continue to present our results based on the first equation of the demand system, namely, for H^r and φ, everything holds for the second equation, too.

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

[9] For singular value decomposition, see Carrasco et al. (2007).

[10] In Assumption 8, we assume that the functions H^r and φ belong to the same regularity space. Analysis with functions with different regularities is very complicated and left for future work.

[11] Tikhonov Regularization do not allow for regularity $\nu > 2$ (See Engl et al., 1996), for that reason in the analysis of convergence rate we are constrained by $\nu \leq 2$.

Assumption 9 *There exists $s \geq 2$ such that:*

- $\left\|\hat{T}^r - T^r\right\|^2 = O\left(\frac{1}{nh_n^3} + h_n^{2s}\right)$
- $\left\|\hat{T}^{r*} - T^{r*}\right\|^2 = O\left(\frac{1}{nh_n^3} + h_n^{2s}\right)$

where s is the minimum between the order of the kernel and the order of the differentiability of the joint density function of (N^r, P^r, Z^r).

Assumption 10

$$\left\|\hat{T}^{r*}\hat{f}^r - \hat{T}^{r*}\hat{T}^r\Phi\right\|^2 = O\left(\frac{1}{n} + h_n^{2s}\right)$$

Assumption 11

$$\lim_{n\to\infty} \gamma_n^r = 0$$

$$\lim_{n\to\infty} (\gamma_n^r)^2 n \to \infty$$

$$\nu \geq 2 \quad or \quad \lim_{n\to\infty} (\gamma_n^r)^{2-\nu} nh_n^3 \to \infty$$

$$\lim_{n\to\infty} h_n = 0$$

$$\lim_{n\to\infty} \frac{h_n^{2s}}{(\gamma_n^r)^2} = 0$$

$$\lim_{n\to\infty} nh_n^3 \to \infty$$

Assumptions (9) and (10) are assumptions on the rate of convergence of the operators. In fact, they can be obtained by imposing more primitive assumptions but here we do not go through this process. Assumption (11) presents the conditions needed to get the consistency.

Theorem 4 *Define $\Phi = (H^r(N^r), \varphi(N^a))$. Let s be the minimum between the order of the kernel and the order of the differentiability of f and ν be the regularity of Φ. Under assumptions 8 to 11:*

- $\left\|\hat{\Phi}_n^{\gamma_n} - \Phi\right\|^2 = O\left(\frac{1}{\gamma_n^2}\left(\frac{1}{n} + h_n^{2s}\right) + \frac{1}{\gamma_n^2}\left(\frac{1}{nh_n^3} + h_n^{2s}\right)\left(\gamma_n^{\min\{\nu,2\}}\right) + \gamma_n^{\min\{\nu,2\}}\right)$
- $\left\|\hat{\Phi}_n^{\gamma_n} - \Phi\right\| \to 0$ *in probability.*

The speed of convergence given in Theorem (4) can not be considered as optimal without any further assumptions. However, under some assumption on γ and s, *the minimax rate* is attainable. Corollary (5) states this.

Corollary 5 *Under assumptions 8 to 11, if:*

- *$\nu < 2$ and $s \geq \frac{3}{2}\frac{\nu+3}{\nu+1}$,*
  or if:

- *$\nu \geq 2$ and $s > 3$,*

Then, for $\gamma \sim n^{-\frac{1}{\min\{\nu,2\}+2}}$, there exist bandwidth choices satisfying the minimax rate:[12]

$$\left\| \hat{\Phi}^{\gamma_n} - \Phi \right\|^2 = O_P\left(n^{-\frac{\min\{\nu,2\}}{\min\{\nu,2\}+2}}\right)$$

3.5 Empirical Analysis of German Magazine Industry

In this section we make an application of the nonparametric estimation defined in the previous section. Using the specification in *Section 3.1* and data on the German magazine industry, we estimate the network effect functions nonparametrically. Estimation results show that the model is supported by the data, since we managed to get well-behaved demand functions without imposing any constraint on estimation. The found network effects are neither linear nor monotone.

3.5.1 Data

We use a data set on the magazine industry from Germany which is available online at *www.medialine.de*. It is an annual data on cover prices, ad prices, number of ad pages, number of content pages, and circulation numbers of German magazines for the year 2009.[13] The original source of their data is the *Informationsgemeinschaft zur Feststellung der Verbreitung von Werbetragern e.V.* This association ascertains, monitors and publishes information on magazine dissemination and circulation.

The unit of observation in our sample is a magazine and we have information on 171 magazines. In the original data set the magazines are grouped according to their content such as *actuality, DIY, women's, parents, tv*, etc. Our sample consists of magazines from 17 different segments. Table (1) shows how our sample is divided into different segments. 18% of the magazines in our data set belong to the *weekly women's magazines* segment and this is the most frequent segment in our sample. The least frequent segments are *DIY* and *animals* each comprising 1.7% of the observations.

The magazines in our sample are published by 25 different publishers. Some of the publishers own magazines only in one segment while some others are publishing for several

[12]Darolles et al. (2011) show that under the maintained assumptions, the rate we obtain is a minimax rate.

[13]The initial data set is quarterly, but we annualize it as the cover prices do not change within a year and ad rate data is supplied annually by the website.

Table 1: Magazine groups in the data

Group of Magazine	Number of titles	Percentage
Women's weekly	32	18.71
Women's monthly	17	9.94
TV	16	9.36
Gardening	13	7.60
Actuality	12	7.02
Photo-electronics-PC	11	6.43
Cars and Motors	11	6.43
Food	10	5.85
Economics	9	5.26
Music and Youth	8	4.68
Science-Geography	8	4.68
Life style-adult-erotic	6	3.51
Parents	6	3.51
Women's bimonthly	4	2.34
Sports	4	2.34
Do-it-yourself	2	1.17
Animals	2	1.17
TOTAL	171	100

different segments. Table (2) shows the publishers of the magazines in our data set. The most dominant publishers in the industry are *Bauer Media Group, Hubert Burda Media, Gruner and Jahr* and *Axel Springer*, publishing 19.3%, 18.1%, 11.7% and 9.9% of the magazines in our data set, respectively.

We assume that the advertisements are all the same size: one page. The mean advertising rate for the sample we use is $16,701$ Euros per ad, while the maximum and minimum rates are highly different from this mean value, at $64,004$ and $1,450$ Euros respectively. In contrast, the cover price range is not that wide. It has a minimum value of 0.45 Euros, a maximum value of 7.95 Euros and the average price is 2.61 Euros. In the data, the average circulation is $1,273,800$, the average number of ad pages is 543 and the average number of content pages is 2199 per year.[14] We also check the share of ad pages to total pages and the share of ad pages to content pages. The average share of ad pages to total pages is 20% while the average share of ad pages to content pages is 26.59%. Two magazines in the monthly women's magazines group, *Cosmopolitan* and *InStyle* have ad pages to content pages ratios larger than 1, showing the importance of adverts in women's magazines.

The share of readers is constructed by dividing the circulation of a given magazine by the population over the age of 14. For the share of advertisers, since we do not have the data on

[14] Circulation rate is composed of the following items: Copies delivered for retail sale, copies for subscribers, copies of the board, reading circle, other sales, free copies. Unsold copies are deducted from the sum of these items.

Table 2: Magazine groups in the data

Publisher	Number of titles	Percentage
Bauer Media Group	33	19.30
Hubert Burda Media	31	18.13
Gruner & Jahr	20	11.70
Axel Springer AG	17	9.94
Klambt Verlag	9	5.26
Gong Verlag	6	3.51
Jahreszeiten-Verlag	6	3.51
Motor-Presse-Stuttgart	6	3.51
Westdeutscher Zeitschriften-Verlag	5	2.92
Vision Media	4	2.34
Bayard Media	3	1.75
CHIP Communications	3	1.75
Cond Nast Verlag	3	1.75
MVG Medien Verlagsgesellschaft	3	1.75
Delius Klasing Verlag	2	1.17
Egmont Verlag	2	1.17
Family Media	2	1.17
Finanzen Verlag	2	1.17
IDG Enterainment Media	2	1.17
In Verlag	2	1.17
Jahr Top Special Verlag	2	1.17
Konradin Medien	2	1.17
SPIEGEL-Verlag Rudolf Augstein	2	1.17
Verlagsgruppe Handelsblatt	2	1.17
Weka Media Publishing	2	1.17
TOTAL	171	100.00

the total number of firms which use the magazines as an advertising medium, we construct it by taking the ratio of the number of ad pages to the total number pages of all magazines in a given segment. We assume that the market size for advertisements is the total number of pages of the magazines in a given segment. For example, for women's weekly magazines the market size is the sum of the number of pages of all women's weekly magazines.

In reality, a magazine with a high share of advertisers has generally a high share of ad pages over total pages. We check the correlation between our constructed market share and the ratio of ad pages over total pages and find that they are highly correlated with a correlation coefficient equal to 0.6. So, the magazines with a greater market share of adverts has generally greater ad page to total page ratio. Although an increase in advertising may decrease the cover price, it may also have a "congestion" effect for the readers. We check the correlation of cover price and share of advertisers and find that it is equal to 0.27. Although it is not high, its positive sign may reflect the fact that the two sides in our sample appreciate each other. In other words, readers are enjoying seeing adverts and the magazine is exploiting this fact.

Table 3: Summary Statistics of the data

Variable	Mean	Std Dev.	Min	Max
No. of Circulation	1273800	1467978	66039	9601297
No. of Ad pages	543.38	496.23	5	2664.1
No. of Content pages	2199.96	1405.99	79	6235.5
Cover price	2.61	1.57	0.45	7.95
Adrate	16701.38	11764.83	1450	67004
Ad pages/Total pages	0.20	0.09	0.04	0.52
Ad pages/ Content pages	0.2659931	0.1813923	0.0474383	1.10498

Remark 6 *We have a sample with* 171 *observations to estimate our model nonparametrically. In Appendix C, we present a simulation to show that the estimation method we propose performs well with a sample of* 170*.*

3.5.2 Estimation

In this section we estimate the network externality functions $\varphi(.)$ and $\psi(.)$ as well as the inverse survival functions $H^r(.)$ and $H^a(.)$.[15] We adopt the estimation technique defined in *Section 3.3*. Let us rewrite the system of equations that we estimate:

$$H^r(N^r) = \varphi(N^a) + P^r + U \tag{19}$$

$$H^a(N^a) = \psi(N^r) + P^a + V \tag{20}$$

where N^r is the share of readers who read the magazine, N^a is the magazine's market share of advertisers as defined in the previous subsection, P^r is the cover price, and P^a is the advertising rate.[16] Moreover, U and V are unobserved magazine characteristics.

All the explanatory variables in equations (19) and (20) are endogenous, so we need to use instruments. Following Kaiser and Wright (2006), we instrument the cover price with the average cover price of the publisher's other magazines. We assume that the common underlying costs will affect the cover price of all other magazines of the same publisher while they are going to be uncorrelated with the disturbances in a particular magazine's demand

[15] It should be noted that our main object is to estimate network effect on each side nonparametrically. For this reason, we give our attention to the functions of interest, φ and ψ, and leave a more complex analysis for future work.

[16] As already stated in *Section 2*, the reason to use the market share of the magazines on the advertisers' side instead of other measures of the advertising amount is to have a fully simultaneous system. We check robustness of our results by using other measure and find that the estimated φ is quite robust.

equation. We instrument the ad rate with the number of titles of the publisher. It is assumed that the number of the titles of the publisher will be correlated with the ad rate through its effect on cost factors via economies of scale. Moreover, again following Kaiser and Wright (2006), the share of readers in the advertising demand equation is instrumented with the average circulation rate of the publisher's other magazines and the share of advertisers in the reader demand equation is instrumented with the average number of advertising pages of the publisher's other magazines. We check the correlation of our instruments with the explanatory variables and they are all found to be reasonably correlated.

We use a *rule of thumb* to construct the bandwidth parameters. The regularization parameters γ_n^r and γ_n^a are chosen by the data-based selection rule of regularization parameter proposed in Florens and Sokullu (2012). Furthermore, since the survival function is monotonic, we monotonized our functions both by *isotonisation* and by *rearrangement*, after the estimation.[17] We report the results where the monotonization is done by rearrangement since it gives better results for the monotonization of probability distribution functions compared to other existing methods, as is pointed out by Chernozhukov et al. (2010). Finally, we construct pointwise bootstrap confidence intervals for the estimated functions. We obtain the confidence interval with 100 replications of *bootstrap in pairs*.[18] In estimation with the bootstrapped data, the values of the regularization and bandwidth parameters are fixed at the levels that we use in the original estimation.

The results are given in Figures (2) to (11). Figures (2) and (3) present the estimation results without monotonization. Figures (6) to (11) show the pointwise bootstrap confidence intervals for the estimated functions. First of all, it is worth noticing that even without monotonization, the estimated functions $\hat{H}^r$ and $\hat{H}^a$ seem to be downward sloping. We know that the survival function is decreasing and so is its inverse. The fact that we are able to obtain decreasing estimated functions without any restriction proves that the model is supported by the data. More concretely, the model we set up to explain the demand and the network externality in the German magazine industry fits the data very well. Figures (4) and (5) show that these results can be improved by monotonization by rearrangement. Moreover, obtaining well behaved inverse survival functions with our estimation method demonstrates also the strength of the method.

Secondly, the estimated network externality functions do not exhibit the same pattern. While the network externality function of advertisers on readers is nonlinear and nonmonotone, the network externality function of readers on advertisers seems quite linear. The

[17] For *isotonisation* and *rearrangement*, see Chernozhukov et al. (2009).

[18] Bootstrap in pairs is performed by resampling the data directly by replacement. We do bootstrap in pairs since all the explanatory variables in the model are endogenous. (See Freedman, 1981; Flachaire, 2005; Horowitz, 2001)

network effect of advertisers, φ is linear and decreasing up to 4% of ad share. Up to 1.5% ad share it takes positive values, and after that point, it takes negative values. In our set up, the share of agents joining the platform is given by the survival function of their net benefits and the survival function is decreasing in its arguments. So, it means that, up to an advertiser share of 1.5%, increasing share of advertiser in a magazine, increases the threshold benefit level and thus decreases the survival function. Hence the readers of the magazines in our sample do not like too few advertisements. They start to get benefit after 1.5% market share of the advertisers. This benefit keeps increasing until 4% and after that point it starts decreasing. So the benefits that the readers get from adverts starts to decrease after a threshold point, as expected. Our result are consistent with the previous literature. Borsenberger et al. (2010), Kaiser (2007) and Kaiser and Song (2009) conclude that readers of entertainment magazines like advertisements, however we also show that they do not like too few advertisements. Although this may sound strange, considering the fact that readers appreciate informative ads, this result means that they are not satisfied with too little information. An example can be given from the *Monthly high-price women's* magazines. Readers of these magazines would not be satisfied to see just the ad of one product of one fashion brand, but they would like to see the ads of other brands and of other products.

On the other side of the market, when we look at the estimated curves for the advertiser demand, figure (5), again we see that the network externality function ψ of readers changes its sign over the interval. Up to a readership share of 20%, ψ takes positive values meaning that it decreases the survival function. However, after 20%, the benefits of advertisers are increasing with the share of readers. So, we can conclude that advertisers do not benefit from advertising in a magazine with a very low level of readership. Another point worth noting is that, contrary to φ this network effect function follows a linear pattern. The benefit of advertisers keeps increasing with the readership of the magazine once it passes the threshold level.

Our results for both sides are intuitive. To give an example, women's magazines are mostly read for the advertisements they include and the adverts in these magazines are informative, see Kaiser and Song (2009). A women's magazine with the advertisement of only one brand and one type of product would not be appreciated by its readers. So, these magazines can have a large share of advertisements and at the same time charge high prices to their readers. *"Vogue"* is a perfect example of this. It is a highly priced women's magazine with a high level of advertising pages. The story holds for other types of entertainment magazines, as well. Readers of a photography magazine are interested in the ads of new products for photography, similarly readers of gardening magazines are interested in the ads of new gardening products. So, the readers of entertainment magazines would not be disturbed by seeing advertisements since ads have informative value for them. Furthermore, advertisers would not be interested in advertising in a magazine with low level of readership, since their aim is to reach as many people as they can.

We conduct three more estimations to check the robustness of our results: *(i)* We estimate the reader demand using *adpages/totalpages* as the explanatory variable instead of market shares of advertisers; *(ii)* we estimate both demands with a subsample which includes women's, tv, sports, gardening, food, music & youth and cars & motors magazines; and *(iii)* we estimate both demands with a subsample which excludes women's, tv and animal magazines.[19] We present the results in *Appendix D*. Results of all three estimations expose the same patterns such that the network effect functions of advertisers on readers are found to be nonlinear and the network effect functions of readers on advertisers are found to be linear.

4 Empirical Analysis of the Magazine Industry : A Parametric Approach

In this section we estimate the model for the magazine industry parametrically. When we make the parametric specification of network effect functions, we take into account the results obtained with our nonparametric specification. Using the same set of explanatory variables, we estimate the system of equations in (8) and (9) simultaneously using GMM. The results we obtain are consistent with the results of the nonparametric estimation.

4.1 Model Specification

To do our analysis with a parametric model, we need to do some specifications for the demand system given by equations (8) and (9). We need to specify *(i)* a distribution function for the net benefits of readers and advertisers, *(ii)* a functional form for the network effects, $\varphi(.)$ and $\psi(.)$ and *(iii)* a functional form for the threshold benefit level of readers and advertisers.

- For the distribution of net benefits, we have chosen the log-logistic distribution function which is common in the literature of network diffusion models, see Larribeau (1993); Feve et al. (2008) Thus, the survival function will be given by:

$$S(X|m,\rho) = \frac{1}{1+(\frac{x}{m})^{\rho}}$$

[19] We have only two observations from animal magazines and one of it having a very high market share on the advertising side.

where m is the scale parameter and ρ is the shape parameter. These parameters can be estimated in advance or during the estimation of the other parameters. In this chapter, we assume that both m and ρ are equal to 1. Larribeau (1993) assumes that ρ is constant over time while m varies and she estimates both of the parameters before estimating the demand equations. On the other side, Feve et al. (2008) assume that they are constant and are equal to 1. For simplicity and without loss of generality, we also assume that both parameters are equal to 1.

- Secondly, we need to approximate the unknown functions $\varphi(.)$ and $\psi(.)$ by some parametric form. To do this, we use our nonparametric estimation results. We choose second order polynomial form for both reader and advertiser demand equations. Thus:

$$\varphi(N^a) = \alpha_0 + \alpha_1 N^a + \alpha_2 (N^a)^2$$

$$\psi(N^r) = \theta_0 + \theta_1 N^r + \theta_2 (N^r)^2$$

Taking into account the result of nonparametric estimation of advertiser demand, obtaining an insignificant estimate for θ_2 would not be surprising.

- Finally we need to choose a functional form for the threshold net benefit levels of the two sides. We decide to use an exponential function which will make the threshold benefit level nonnegative:

$$\underline{b}^r = f(N^a, X\beta, U) = exp(\varphi(N^a) + X\beta + U)$$

equivalently for the advertisers' side:

$$\underline{b}^a = f(N^r, W\gamma, V) = exp(\psi(N^r) + W\gamma + V)$$

Now, we need to specify the explanatory variables X and W. To be consistent with the nonparametric analysis, we used the same explanatory variables, cover price and ad rate of the magazines.

Then our simultaneous demand system to be estimated becomes:

$$N^r = \frac{1}{1 + exp(\alpha_0 + \alpha_1 N^a + \alpha_2 (N^a)^2 + \beta P^r + U)} \tag{21}$$

$$N^a = \frac{1}{1 + exp(\theta_0 + \theta_1 N^r + \theta_2 (N^r)^2 + \gamma P^a + V)} \tag{22}$$

4.2 Estimation

Using the specification above, we estimate the demand equations, (21) and (22) simultaneously by GMM. As we have mentioned before, to be consistent with the nonparametric estimation, we use just one explanatory variable in each equation, cover price and ad rate. Since, the prices, and the share of agents in each side is endogenous, we use instruments. The instruments for cover price and ad rate are the same as in the nonparametric analysis. So, we instrument them with the average cover price of the same publisher's other magazines and the number of titles published by the own publisher, respectively. For the share of readers and advertisers, we do not change the instruments, either, but we also use the competitors' averages.[20] Moreover, following Kaiser and Wright (2006), we use the natural logarithm of the number of magazine titles published by the same publisher and the natural logarithm of the total number of pages published by the same publisher. These are also assumed to be cost side instruments for cover prices and ad rates, while the latter is a proxy variable for the returns to scale, the former is considered as a proxy variable for the returns to scope. Other additional instruments we use are magazine group dummies. P-value of the Hansen's J-statistic is reported to be greater than 0.1 so that we can not reject the orthogonality of instruments and the error terms. The estimation results are given in Tables (4) and (5).

Table 4: Estimation Results for the Reader Demand

Variable	Parameter	Estimate	Standard Error	t-stat	p-value
constant	α_0	1.786042	0.2157	8.28	$< .0001$
share of advertisers	α_1	-113.34	30.2096	-3.75	0.0002
$(\textit{share of advertisers})^2$	α_2	819.8798	276.9	2.96	0.0035
cover price	β	0.903281	0.2013	4.49	$< .0001$

Table 5: Estimation Results for the Advertiser Demand

Variable	Parameter	Estimate	Standard Error	t-stat	p-value
constant	θ_0	4.243117	0.6937	6.12	$< .0001$
share of readers	θ_1	26.64147	9.1254	2.92	0.0040
$(\textit{share of readers})^2$	θ_2	-25.1534	16.6758	-1.51	0.1333
ad rate	γ	-0.13	0.016	-8.23	$< .0001$

For readability, estimated coefficient and standard error of ad rate (γ) is multiplied by 1000.

All of our estimated parameters are significant at 5% level and have the expected sign in the readers' demand equation. The fact that the coefficient of $(N^a)^2$ is found to be

[20] Similarly, assuming the fact that the underlying cost factors are the same across the same segments of the industry, more precisely, across the same magazine groups, Borsenberger et al. (2010), use competitors' average values of the related variables as the instruments for the endogenous variables.

significant leads to rejecting the null hypothesis that the network effects are linear. The coefficient of price is found to be positive, meaning that an increase in the cover price will increase the threshold benefit function, thus leading to a decrease in demand. The estimated constant is equal to 1.78 and is significant. It suggests that, a magazine with 0 price and 0 advertising pages will be bought or taken by the 14.5% of the readers. When we examine the estimated network effect function, $\hat{\varphi}$, we see that our results are very much in line with the nonparametric estimate of φ. According to the parametric estimation results, the readers start to get benefit from advertising only after a threshold level of approximately 1.8%. Moreover, their benefit from adverts is increasing up to around 7% of advertising share, then it starts to decrease. Note that with nonparametric estimation these values are estimated to be 1.5% and 4% respectively. In *Appendix E*, we present the graph of $\hat{\varphi}$ estimated parametrically. Moreover, we present a plot which shows parametric and nonparametric estimates of φ in one graph. It can be seen from the graph that the estimated values are similar with both methods for low levels of advertising share and the curves present almost the same patterns. According to the nonparametric estimation result, threshold level is reached at a lower rate of ads. The figure also suggests the possibility of using a third order polynomial in parametric approximation.

For the advertising demand equation, all of the estimated parameters, except the coefficient of $(N^r)^2$, are significant at 5% level. The estimated coefficient of $(N^r)^2$ is insignificant, so we fail to reject the null hypothesis that the network effect of readers are linear. Indeed this result is also in line with what we obtained in *Section 3.5.2*. However the estimated coefficients of ad rate and the readership share do not have the expected sign. According to the results, advertiser demand is decreasing in readership share and increasing in price of adverts. This counter-intuitive result may stem from incorrect functional approximations which brings us again to the debate of parametric vs. nonparametric estimation.

5 A Numerical Illustration

In this section we present a numerical example to demonstrate the importance of a correct specification of the network effects. To do this we perform two parametric estimations. In the first one the network effect function in reader demand is specified as a second order polynomial (Equation 21) and in the second one it is specified linearly as in equation (23) below. The network effect function in advertiser demand is specified linearly (Equation 24) in both models since we fail to reject linear network effects for advertisers in the previous section. Thus, the first model is given by equations (21) and (24) and the second model is given by equations (23) and (24):

$$N^r = \frac{1}{1 + exp(\alpha_0 + \alpha_1 N^a + \beta P^r + U)} \tag{23}$$

$$N^a = \frac{1}{1 + exp(\theta_0 + \theta_1 N^r + \gamma P^a + V)} \tag{24}$$

We estimate the equations in each model simultaneously by GMM and we use the same set of instruments as in the previous section for both models. The estimation results are presented in Tables (6) and (7) in *Appendix E*. All the estimated parameters are significant at 5% level. For both of the models, readers are found to be ad-lovers and an increase in the price levels of magazines decreases the reader demand. On the other side of the market, the results for the advertiser demand are again counter-intuitive.

We use the estimated parameters from both models to compute the mark-ups given in equations (6) and (7) at the mean values of demands.[21] We find that for the readers' side the mark-up estimated with linear network effects in 27% lower than the mark-up estimated with nonlinear network effects. For the advertisers' side, it is estimated 7% higher in the linear case.

This numerical illustration shows the importance of functional specifications in empirical analysis and that an analysis which is done under wrong specification will lead to erroneous conclusions. This is especially important from a regulatory point of view. If the analysis in the Argentesi and Filistrucchi (2007) is considered, where they analyze the market power of Italian newspapers, an error coming from this type of specification may reverse their results. The collusive behavior on the reader side would have been found to be competitive and/or the competitive behavior on the advertising side would have been found to be collusive if the network effects were allowed to be nonlinear.

6 Conclusion

This paper has developed a semi-nonparametric empirical model for two sided markets. We specify the network externalities nonparametrically to be able to capture the nonlinearities. The distribution functions of benefits of readers and advertisers are not specified and are left to be estimated, as well. The model is estimated with nonparametric IV estimation. We get two main results: First of all, the structural model is supported by the data since

[21] Although the results we obtain imply that the advertisers do not like readers so that they are subsidized by readers, since we are doing this exercise for illustrative purposes only, we think that we can still use the results.

we obtain well behaved demand curves by nonparametric instrumental variables estimation without any restriction. Secondly, network effects on the readers' side are nonlinear and nonmonotone.

Using our nonparametric estimation results, we make parametric approximations for the network effect functions and re-estimate the model by GMM. We find that with a second order polynomial specification for the network externality function on the reader side, we get very close to what we have obtained in nonparametric estimation.

This paper has many contributions. First of all, the nonparametric estimation method is a contribution as an extension to Florens and Sokullu (2012). Secondly, nonparametric specification and nonparametric tools have never been used in empirical two sided markets literature. Finally, it proves both by nonparametric and parametric estimations that the network effects in magazine industry are neither linear nor monotone. This result indeed may have important implications for policy analysis.

The estimations in the paper are done with cross-sectional data so a natural extension, nonparametric analysis with panel data is left for future work. Moreover, there is still no dynamic work in the empirical two sided market literature. A dynamic model based on network diffusion models is also a very interesting future research topic. Finally, the estimation method presented in this chapter can be applied to any industry with network externalities.

Appendices

A A Discussion on Nonseparable Case

In *Section 3* we discuss the identification and estimation of a semi-nonparametric model. We present the identification, estimation and consistency of the model in equations (10) and (11), where the threshold benefit function, $\underline{b}^i$, $i \in \{r, a\}$ is specified partially linear. In this section, we discuss the nonseparable case. More concretely, the case where we do not specify any form for $\underline{b}^i$ and we just assume that the survival functions of the readers' and advertisers' benefits are functions of (N^a, P^r, U) and (N^r, P^a, V), respectively. Before going into details of identification and estimation of a nonseparable model, let us state why we do not use this nonseparable model in our empirical analysis. First of all, our aim in this paper is to identify the network externality functions of the two sides that they exert on each other and a fully nonparametric nonseparable specification would not allow us to estimate φ and ψ. Secondly, the assumptions needed for the identification of our model is weaker and easier to satisfy than the nonseparable model. For example, conditional mean independence between U and Z^r and V and Z^a is enough for the identification of our model, whereas, for the nonseparable case full independence is needed. Finally, the estimation of the nonseparable case leads to a nonlinear ill-posed inverse problem which is much harder to deal with. Since the results we obtain with our model makes sense and since the method is easy to apply, we decide to use that one for our application.

Let us now rewrite the system of equations in (10) and (11) under the nonseparable framework:[22]

$$N^r = S^r(N^a, P^r, U) \tag{25}$$

$$N^a = S^a(N^r, P^a, V) \tag{26}$$

The two main assumptions for the identification of nonseparable models are *(i)* the independence of U and Z^r and V and Z^a, and *(ii)* the strict monotonicity of S^r in U and S^a in V. By its set up, our model satisfies these two assumptions. The identification results of this type of a model are already studied in Chernozhukov et al. (2007). Chernozhukov et al. (2006) examines estimation with Tikhonov Regularization and asymptotic properties of the same model.

Under the construction of nonseparable model we can identify *(i)* the network effect of one side on a particular quantile of the other side, namely, quantile treatment effect, and *(ii)*

[22] It should be noted that, in this case we no longer need the $\varphi(.)$ and $\psi(.)$ functions, as we do not make any specification for $\underline{b}^i$, $i \in \{r, a\}$.

network effects on the whole distribution (average structural function). However as already mentioned such a specification would not allow us to identify separate functional forms for the network externalities.

B Proofs of Theorems

B.1 Theorem 1

Proof. We will present the proof just for the reader side equation as the proofs are the same. By Assumptions (1) and (2)

$$\mathbb{E}[H^r(N^r) - \varphi(N^a) - P^r|Z^r] = 0$$

Let us recall two more functions $H^{r*}(N^r)$ and $\varphi^*(N^a)$. By Assumption (2) again, we can write:

$$\mathbb{E}[H^r(N^r) - \varphi(N^a) - P^r|Z^r] = 0 \quad \mathbb{E}[H^{r*}(N^r) - \varphi^*(N^a) - P^r|Z^r] = 0$$

If we take the difference of the two expectations:

$$\mathbb{E}[(H^r(N^r) - H^{r*}(N^r)) - (\varphi(N^a) - \varphi^*(N^a)) + (P^r - P^r)|Z^r] = 0$$

then by Assumption (3):

$$(H^r(N^r) - H^{r*}(N^r)) - (\varphi(N^a) - \varphi^*(N^a)) = 0$$

finally by Assumption (4):

$$(H^r(N^r) - H^{r*}(N^r)) = (\varphi(N^a) - \varphi^*(N^a)) = c$$

and by Assumption (5):

$$c = 0$$

then:

$$H^r(N^r) = H^{r*}(N^r) \quad and \quad \varphi(N^a) = \varphi^*(N^a)$$

■

B.2 Theorem 2

Proof. Assumption (1) guarantees the existence of the structural inverse system of differentiable functions g^r and g^a that satisfy:

$$U = g^r(N^r, P^r, N^a) \quad and \quad V = g^a(N^a, P^a, N^r)$$

and

$$N^r = S^r(N^a, P^r, g^r(N^r, P^r, N^a)) \quad and \quad N^a = S^a(N^r, P^a, g^a(N^a, P^a, N^r))$$

By Assumption (6), we can differentiate these equations with respect to N^r and N^a to get:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{\partial S^r}{\partial U}\frac{\partial g^r}{\partial N^r} & \frac{\partial S^r}{\partial \varphi}\frac{\partial \varphi}{\partial N^a} + \frac{\partial S^r}{\partial U}\frac{\partial g^r}{\partial N^a} \\ \frac{\partial S^a}{\partial \psi}\frac{\partial \psi}{\partial N^r} + \frac{\partial S^a}{\partial V}\frac{\partial g^a}{\partial N^r} & \frac{\partial S^a}{\partial V}\frac{\partial g^a}{\partial N^a} \end{pmatrix} \tag{27}$$

so:

- $\frac{\partial S^r}{\partial U}\frac{\partial g^r}{\partial N^r} = 1 \Rightarrow \partial S^r/\partial U = (\partial g^r/\partial N^r)^{-1}$
- $\frac{\partial S^r}{\partial N^a} = -\frac{\partial S^r}{\partial U}\frac{\partial g^r}{\partial N^a}$
- $\frac{\partial S^a}{\partial N^r} = -\frac{\partial S^a}{\partial V}\frac{\partial g^a}{\partial N^r}$
- $\frac{\partial S^a}{\partial V}\frac{\partial g^a}{\partial N^a} = 1 \Rightarrow \partial S^a/\partial V = (\partial g^a/\partial N^a)^{-1}$

Then we can write

- $\partial g^r/\partial N^r = (\partial S^r/\partial U)^{-1} < 0$
- $\partial g^a/\partial N^a = (\partial S^a/\partial V)^{-1} < 0$
- $\partial g^r/\partial N^a = -(\partial S^r/\partial N^a)(\partial S^r/\partial U)^{-1}$
- $\partial g^a/\partial N^r = -(\partial S^a/\partial N^r)(\partial S^a/\partial V)^{-1}$

By Assumption (7), this implies, $(\partial g^r/\partial N^a)(\partial g^a/\partial N^r) < 0$. Then it follows from Theorem 7 of Gale and Nikaido (1965) that the function $F(N^r, N^a) = (S^r(.), S^a(.))$ is univalent and hence there exist unique functions h^r and h^a such that for given P^r, P^a and for all U and V:

$$U = g^r(h^r(P^a, P^r, U, V), h^a(P^a, P^r, U, V), P^r)$$

$$V = g^a(h^r(P^a, P^r, U, V), h^a(P^a, P^r, U, V), P^a)$$

■

B.3 Theorem 4

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

$$\Phi = (\gamma_n^r I + T^{r*}T^r)^{-1}T^{r*}f^r$$

- For the proof of the rate of convergence, following Darolles et al. (2011), we decompose $\hat{\Phi}_n^\gamma - \Phi$ into three parts. Then we look at the rate of convergence of each part.

$$\hat{\Phi}_n^\gamma - \Phi = \underbrace{(\gamma_n I + \hat{T}^{r*}\hat{T}^r)^{-1}\hat{T}^{r*}\hat{f}^r - (\gamma_n I + \hat{T}^{r*}\hat{T}^r)^{-1}\hat{T}^{r*}\hat{T}^r\Phi}_{I}$$
$$+ \underbrace{(\gamma_n I + \hat{T}^{r*}\hat{T}^r)^{-1}\hat{T}^{r*}\hat{T}^r\Phi - (\gamma_n I + T^{r*}T^r)^{-1}T^{r*}T^r\Phi}_{II}$$
$$+ \underbrace{(\gamma_n I + T^{r*}T^r)^{-1}T^{r*}T^r\Phi - \Phi}_{III}$$

The first term (I) is the estimation error about the right hand side (f^r) 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 γ_n.

Now, let's first examine the first term:

$$I = (\gamma_n I + \hat{T}^{r*}\hat{T}^r)^{-1}\hat{T}^{r*}\hat{f}^r - (\gamma_n I + \hat{T}^{r*}\hat{T}^r)^{-1}\hat{T}^{r*}\hat{T}^r\Phi$$
$$I = (\gamma_n I + \hat{T}^{r*}\hat{T}^r)^{-1}\hat{T}^{r*}(\hat{f}^r - \hat{T}^r\Phi)$$
$$\|I\|^2 \leq \left\|(\gamma_n I + \hat{T}^{r*}\hat{T}^r)^{-1}\right\|^2 \left\|\hat{T}^{r*}\hat{f}^r - \hat{T}^{r*}\hat{T}^r\Phi\right\|^2$$

where the first term is $O\left(\frac{1}{\gamma_n^2}\right)$ by Darolles et al. (2011) and the second term is $O\left(\frac{1}{n} + h_n^{2s}\right)$ by Assumption (10).

Now, let us look at the second term II:

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

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

The first term in (II) is $O(\frac{1}{\gamma_n^2})$ by Darolles et al. (2011) while the second one is of order $O\left(\frac{1}{nh_n^3} + h_n^{2s}\right)$ as a result of relation $\left\|\hat{T}^{r*}\hat{T}^r - T^{r*}T^r\right\| = O\left(max\left\|\hat{T}^r - T^r\right\|, \left\|\hat{T}^{r*} - T^{r*}\right\|\right)$ by Assumption (9) and by Florens et al. (2009). Finally, the third is equal to $O(\gamma_n^{\min\{\nu,2\}})$ by Darolles et al. (2011).

The third term can be examined more straightforwardly:

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

$$= \Phi_n^\gamma - \Phi$$

and $\|III\|^2 = \|\Phi_n^\gamma - \Phi\|^2$ is $O(\gamma_n^{\min\{\nu,2\}})$ by assumption 8. Finally if we combine all what we have:

$$\left\|\hat{\Phi}_n^{\gamma_n} - \Phi\right\|^2 = O\left(\frac{1}{\gamma_n^2}\left(\frac{1}{n} + h_n^{2s}\right) + \frac{1}{\gamma_n^2}\left(\frac{1}{nh_n^3} + h_n^{2s}\right)\left(\gamma_n^{\min\{\nu,2\}}\right) + \gamma_n^{\min\{\nu,2\}}\right)$$

- 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:

$$\left\|\hat{\Phi}_n^\gamma - \Phi\right\| \leq \underbrace{\left\|\hat{\Phi}_n^\gamma - \Phi_n^\gamma\right\|}_{A} + \underbrace{\|\Phi_n^\gamma - \Phi\|}_{B}$$

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

$$\hat{\Phi}_n^\gamma - \Phi_n^\gamma = (\gamma_n I + \hat{T}^{r*}\hat{T}^r)^{-1}\hat{T}^{r*}\hat{f}^r - (\gamma_n I + T^{r*}T^r)^{-1}T^{r*}T^r\Phi$$

$$= \underbrace{(\gamma_n I + \hat{T}^{r*}\hat{T}^r)^{-1}(\hat{T}^{r*}\hat{f}^r - \hat{T}^{r*}\hat{T}^r\Phi)}_{I}$$

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

If we look at A term by term:

$$\|I\| \leq \left\|(\gamma_n I + \hat{T}^{r*}\hat{T}^r)^{-1}\right\| \left\|(\hat{T}^{r*}\hat{f}^r - \hat{T}^{r*}\hat{T}^r\Phi)\right\|$$

the first term is of order $O(\frac{1}{\gamma_n})$ by Darolles et al. (2011) and the second term is of order

$O(\frac{1}{\sqrt{n}}+h_n^s)$ by Assumption (10). So the the first term (I) converges to $\left(\frac{1}{\gamma_n}\left(\frac{1}{\sqrt{n}}+h_n^s\right)\right)$.

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

$$\|II\| \leq \left\|\gamma_n(\gamma_n I + T^{r*}T^r)^{-1}\Phi\right\|\left\|\hat{T}^{r*}\hat{T}^r - T^{r*}T^r\right\|\left\|(\gamma_n I + T^{r*}T^r)^{-1}\right\|$$

The first part is of order $O(\gamma_n^{\min\{\nu/2,1\}})$ by Darolles et al. (2011), the second term is of order $O(\frac{1}{\sqrt{nh_n^3}}+h_n^s)$ by assumption 9 and by Florens et al. (2009) and the last term is smaller than $O(\frac{1}{\gamma_n})$. So, the second term (II) of A converges to $\left(\frac{1}{\gamma_n}\left(\frac{1}{\sqrt{nh_n^3}}+h_n^s\right)\gamma_n^{\min\{\nu/2,1\}}\right)$. Then, we can conclude that $\left\|\hat{\Phi}_n^\gamma - \Phi\right\|$ converges to zero in probability under the conditions of Assumption (11).

■

B.4 Corollary 5

Proof. To prove Corollary 5, first it should be noted that the nonparametric estimation of conditional expectations does not cost too much in terms of speed of convergence. More precisely, the middle term in the convergence rate given in Theorem 4, is negligible in front of the two other terms. Optimal convergence rate is attained by the choice of optimal regularization parameter. Hence, it is obtained by equalizing the first and the third term of the convergence rate:

$$\frac{1}{\gamma_n^2 n} \sim \gamma_n^{\min\{\nu,2\}} \Rightarrow \gamma_n \sim n^{-\frac{1}{\min\{\nu,2\}+2}}$$

So, under the maintained assumption that h_n^{2s} is of order $1/n$, the optimal regularization parameter is proportional to $n^{-\frac{1}{\min\{\nu,2\}+2}}$. Now, we can work on the middle term to complete the proof. The rate of the middle term is given by:

$$\left(\frac{1}{nh_n^3}+h_n^{2s}\right)\gamma_n^{\min\{\nu-2,0\}} = O\left(\frac{\gamma_n^{\min\{\nu-2,0\}}}{nh_n^3}\right)$$

Under the assumptions 8 to 11 and under the condition that $\gamma_n \sim n^{-\frac{1}{\min\{\nu,2\}+2}}$, we can always find a bandwidth that satisfies:

$$\frac{1}{nh_n^3} = O\left(\frac{\gamma_n^{\min\{\nu,2\}}}{\gamma_n^{\min\{\nu-2,0\}}}\right)$$

$$\frac{1}{h_n^3} = \begin{cases} O\left(n^{\frac{\nu+1}{\nu+2}}\right) & if \nu < 2 \\ O\left(n^{\frac{1}{2}}\right) & if \nu \geq 2 \end{cases}$$

Combining this result with the assumption that $h_n^{2s} = O(1/n)$, we can write:

$$\frac{3}{2s} \leq \begin{cases} \frac{\nu+1}{\nu+2} & if \nu < 2 \\ \frac{1}{2} & if \nu \geq 2 \end{cases}$$

And, this completes the proof. ■

C A Simulation to See the Small Sample Performance of the Nonparametric Estimation

In order to show that the nonparametric estimation technique proposed in this paper performs well with small samples, we present below a simulation. The model considered in the simulation is:

$$H(Y) = \phi(X) + W + U$$

where $Y, X, W \in \mathbb{R}$ are endogenous and $U \in \mathbb{R}$ is the error term. H is specified as:

$$H(Y) = ln\left(\frac{1-Y}{0.1Y}\right)$$

Moreover $\phi(X) = X^2$. U is drawn randomly from a standard normal distribution. We draw random samples of ϵ_x, ϵ_w, i_1 and i_2 from a uniform distribution, $\mathcal{U}(0,1)$. Then the endogenous variables are constructed as the following:

$$\eta_x = -2U + \epsilon_x$$

$$\eta_w = U + \epsilon_w$$

$$X = i_1 + 0.2i_2^2 + \eta_x$$

$$W = -0.1i_2 + \eta_w$$

The sample size is chosen to be equal to 170 like the sample size of our dataset in the application. We present below the the real curves (green dots) and the fitted curves (red pluses) for the $H_y(.)$ and φ.

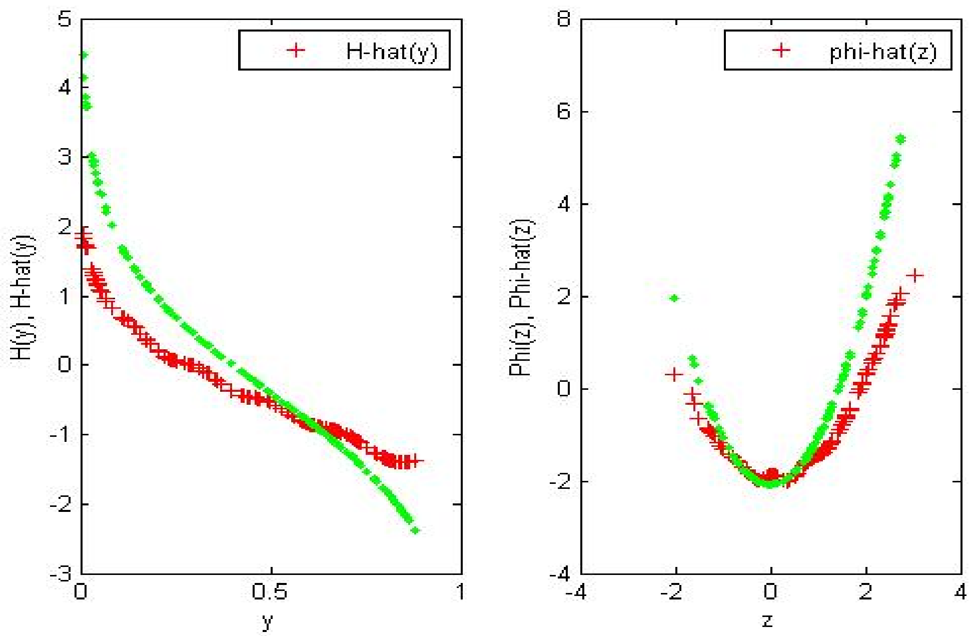

Figure 1: *Simulation for a sample of size 170*

D Estimation Results of the Semiparametric Model

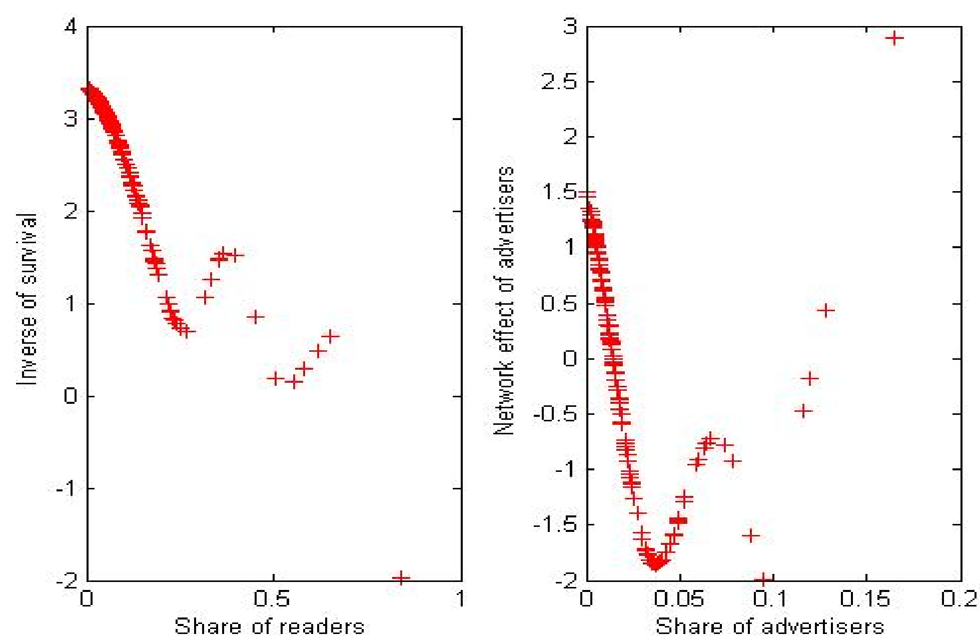

Figure 2: *Estimated functions for reader demand equation*

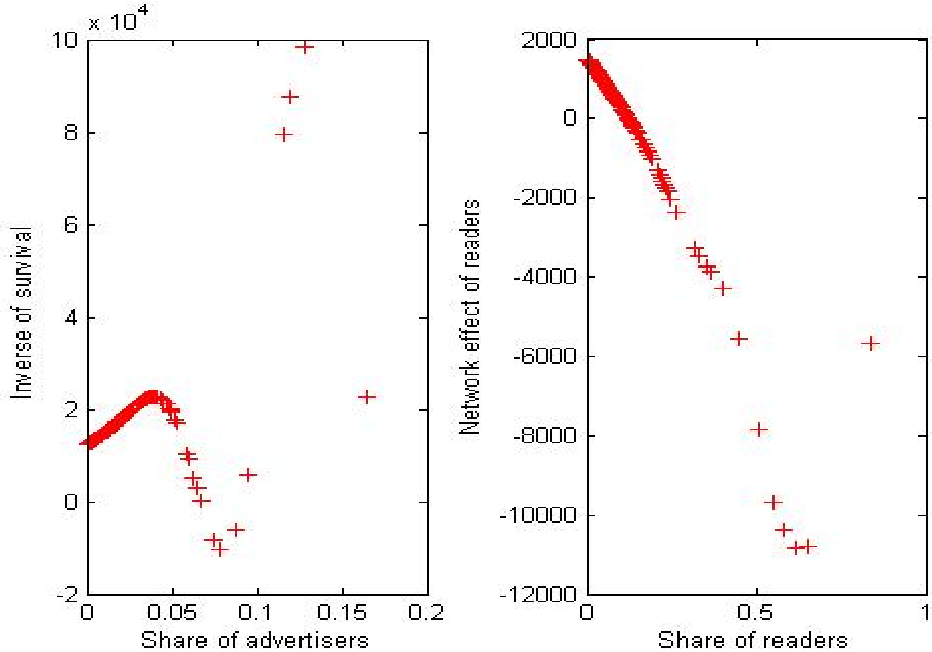

Figure 3: *Estimated functions for advertiser demand equation*

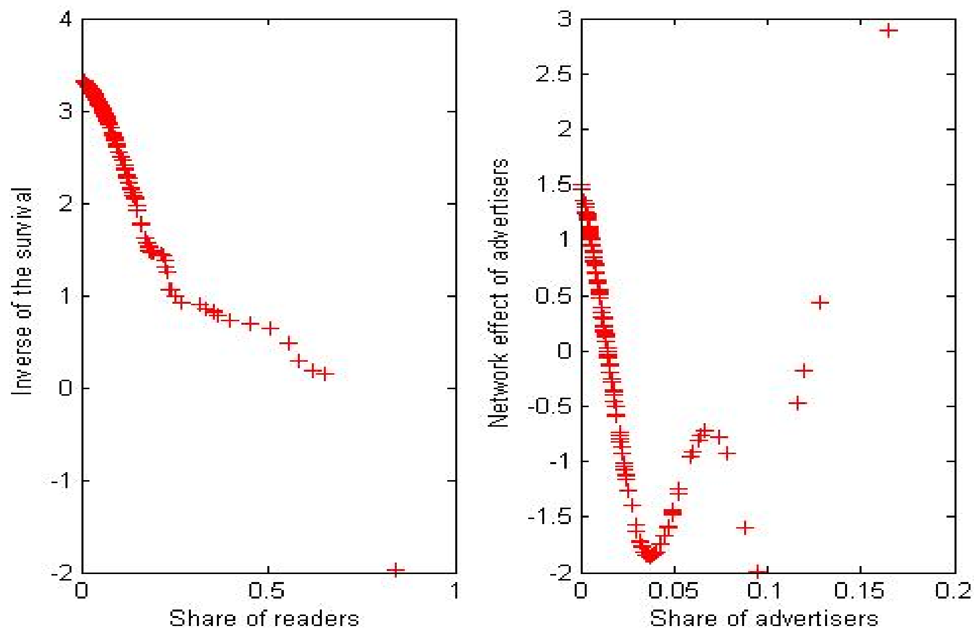

Figure 4: *Estimated functions for reader demand equation after monotonization with rearrangement*

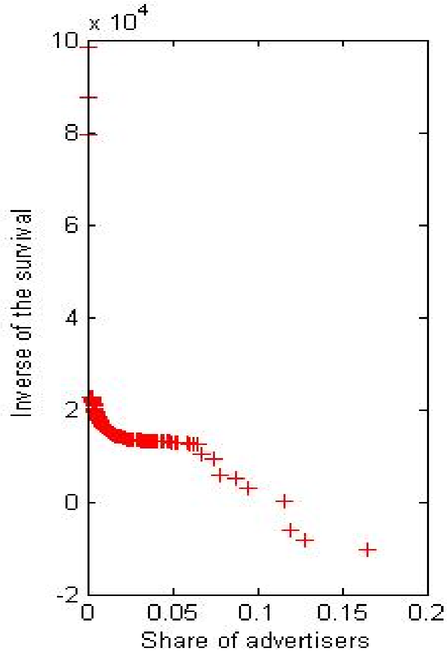

Figure 5: *Estimated functions for advertiser demand equation after monotonization with rearrangement*

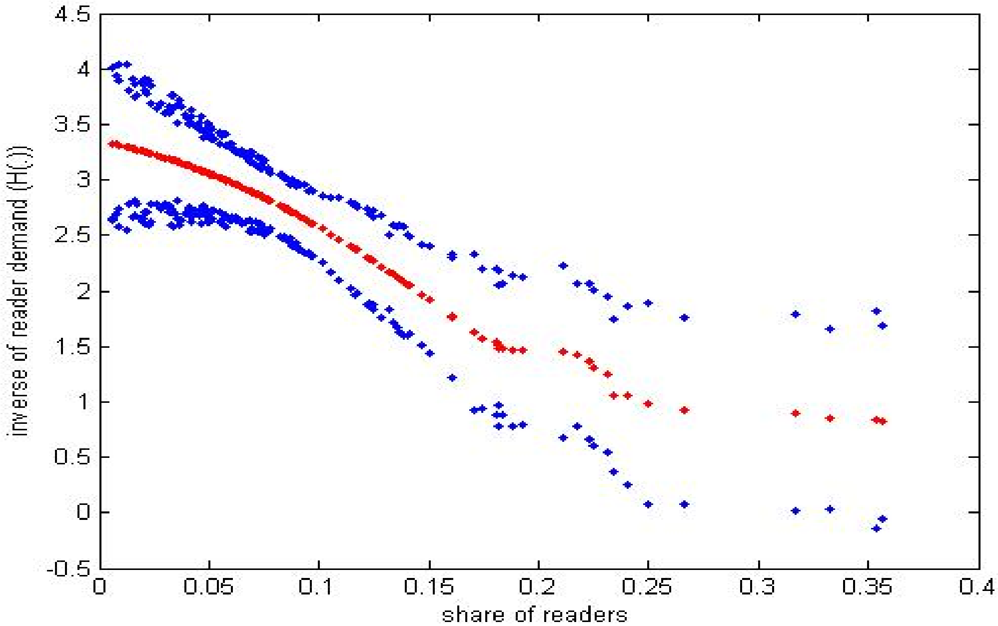

Figure 6: *Bootstrap confidence interval for function* $H^r(N^r)$

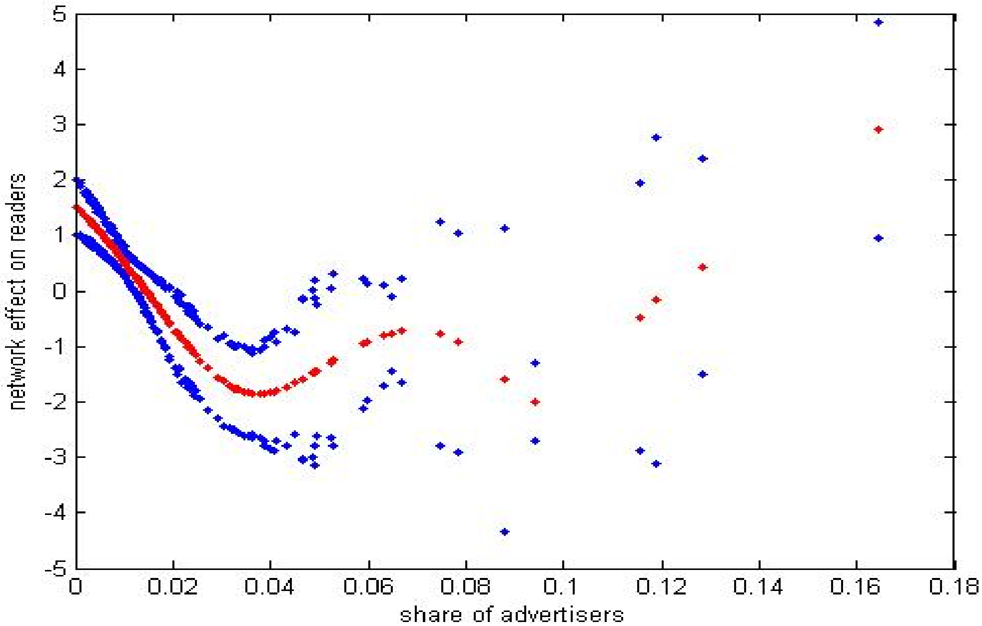

Figure 7: *Bootstrap confidence interval for function* $\varphi(N^a)$

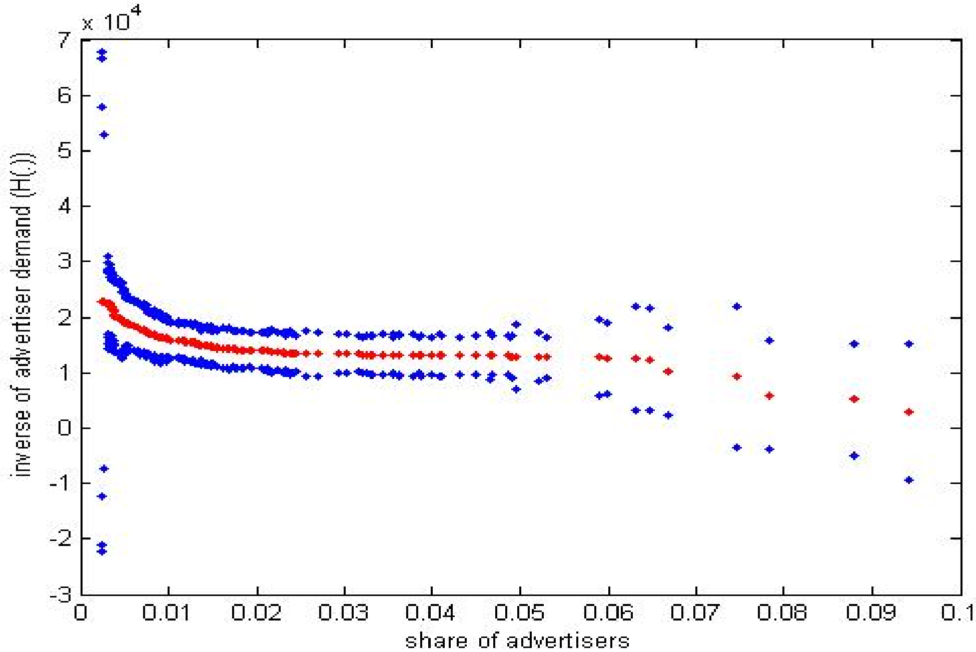

Figure 8: *Bootstrap confidence interval for function* $H^a(N^a)$

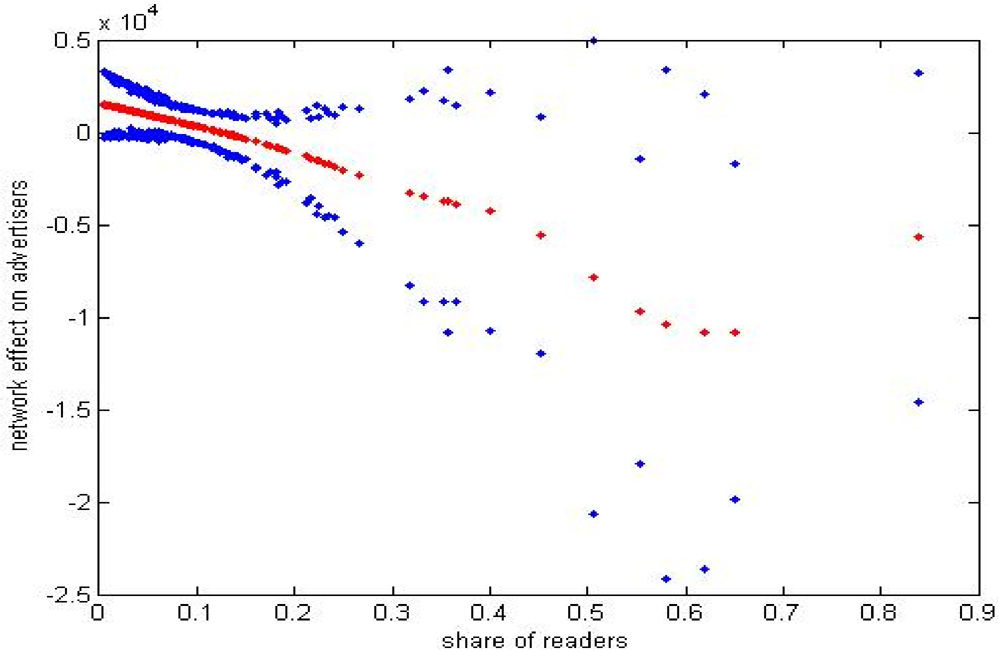

Figure 9: *Bootstrap confidence interval for function* $\psi(N^r)$

D.1 Estimation Results for Robustness Checks

In the first robustness check exercise we use a different definition of ad pages share in the reader demand equation. We construct a new variable $N^{ar} = adpages/totalpages$ and estimate the reader demand equation: $H^r(N^r) = \varphi(N^{ar}) + P^r + U$.

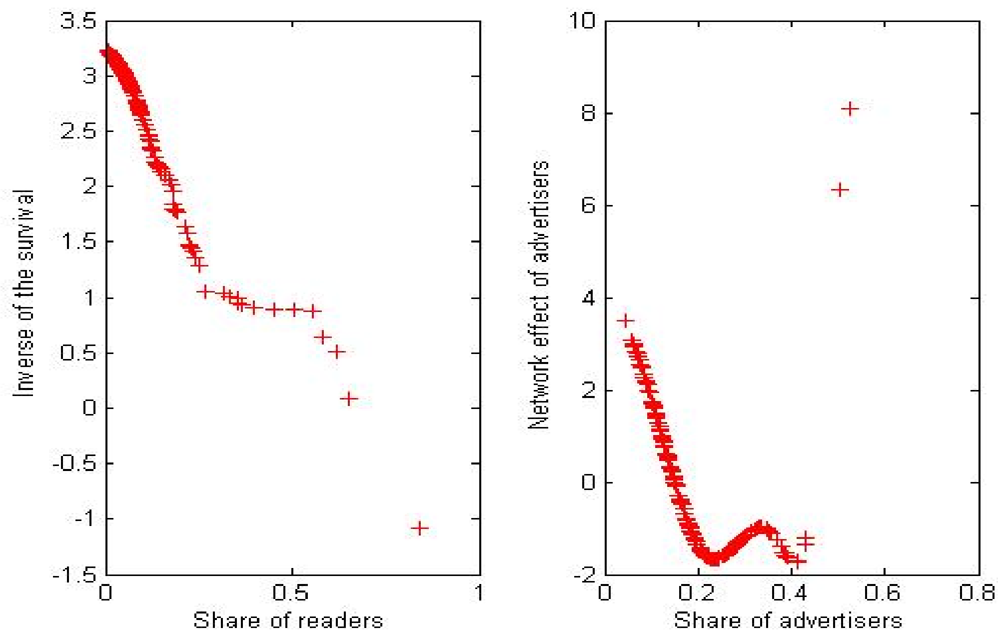

Figure 10: *Reader demand estimation with* $N^{ar} = adpages/totalpages$

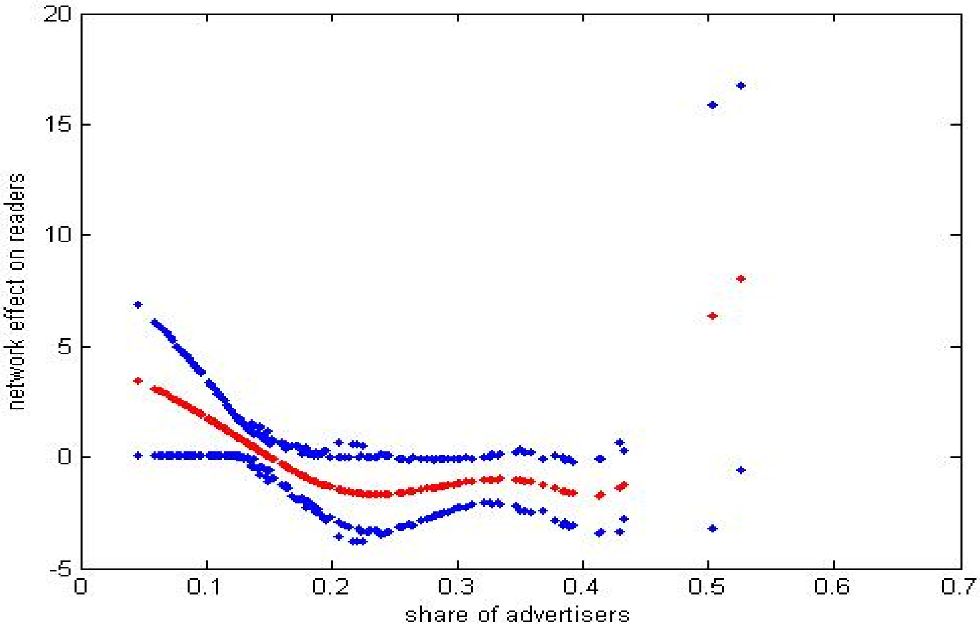

Figure 11: *Bootstrap confidence interval for function* $\varphi(N^{ar})$

In the second exercise, we construct a subsample that consist of observations from women's and hobby magazines, i.e. women's, tv, gardening, sports, cars and motors, music and youth, food. We estimate both the reader and advertiser demand functions.

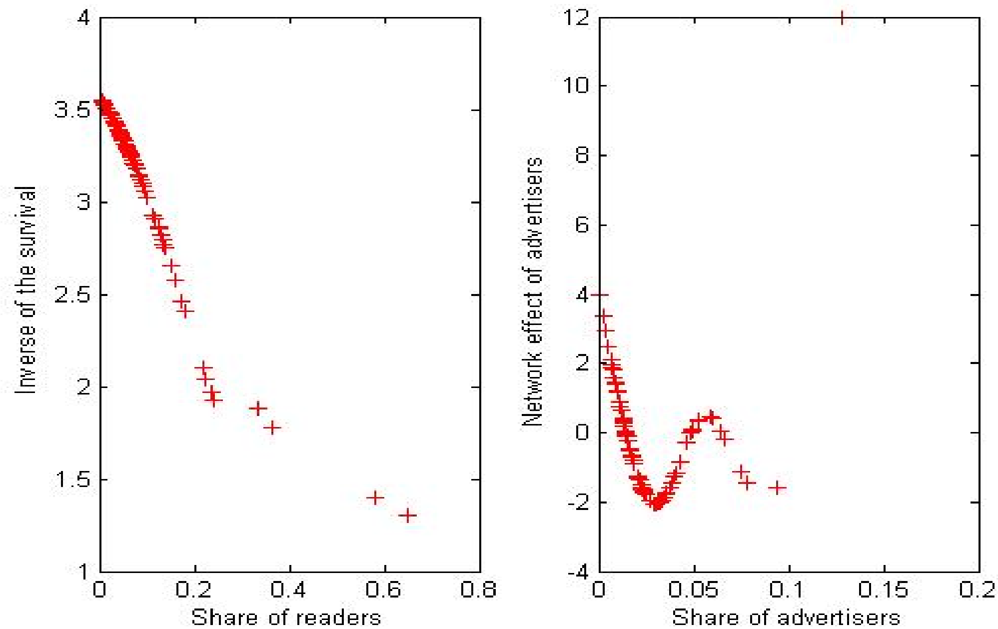

Figure 12: *Estimated functions for reader demand equation after monotonization with rearrangement. Estimation done with subsample of women's and hobby magazines. Sample size=115*

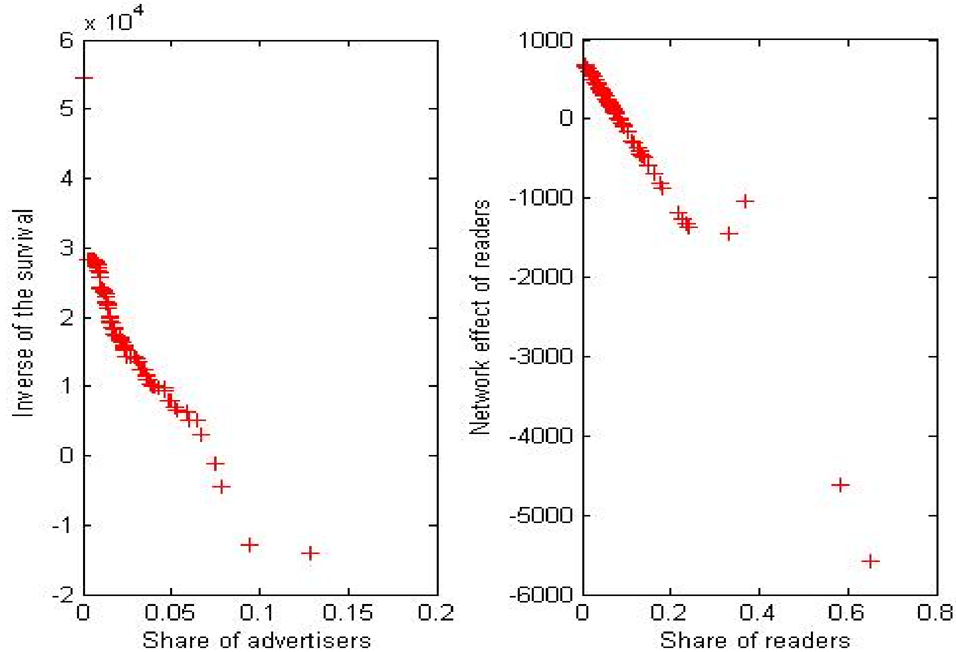

Figure 13: *Estimated functions for advertiser demand equation after monotonization with rearrangement. Estimation done with subsample of women's and hobby magazines. Sample size=115*

As a third robustness check, we drop all the observations of women's and tv magazines from our sample and estimate the reader and advertiser demand with this new sample.

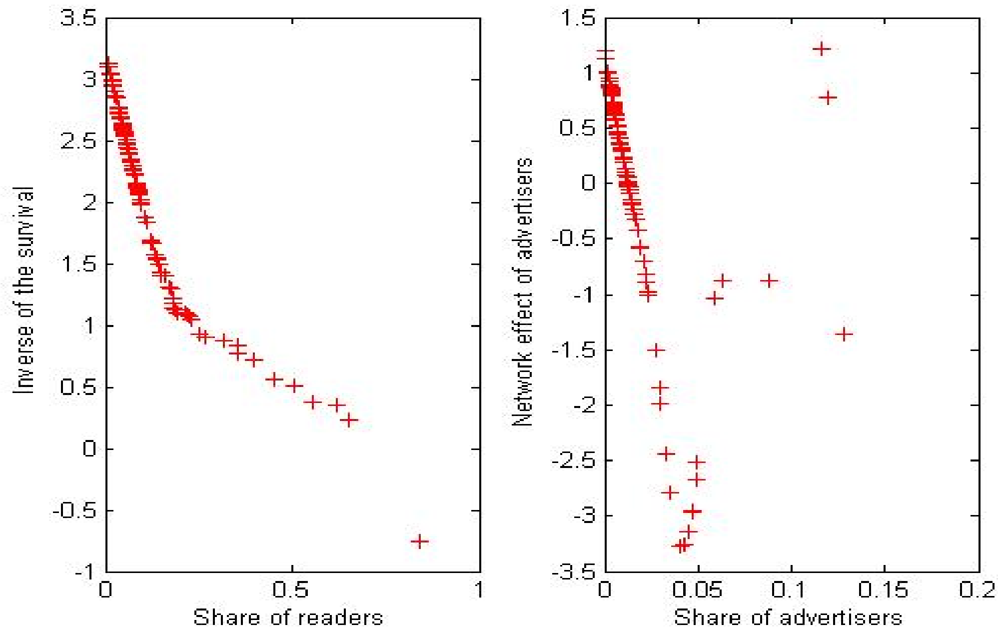

Figure 14: *Estimated functions for reader demand equation after monotonization with rearrangement. Estimation done with subsample without women's, tv and animal magazines. Sample size=100*

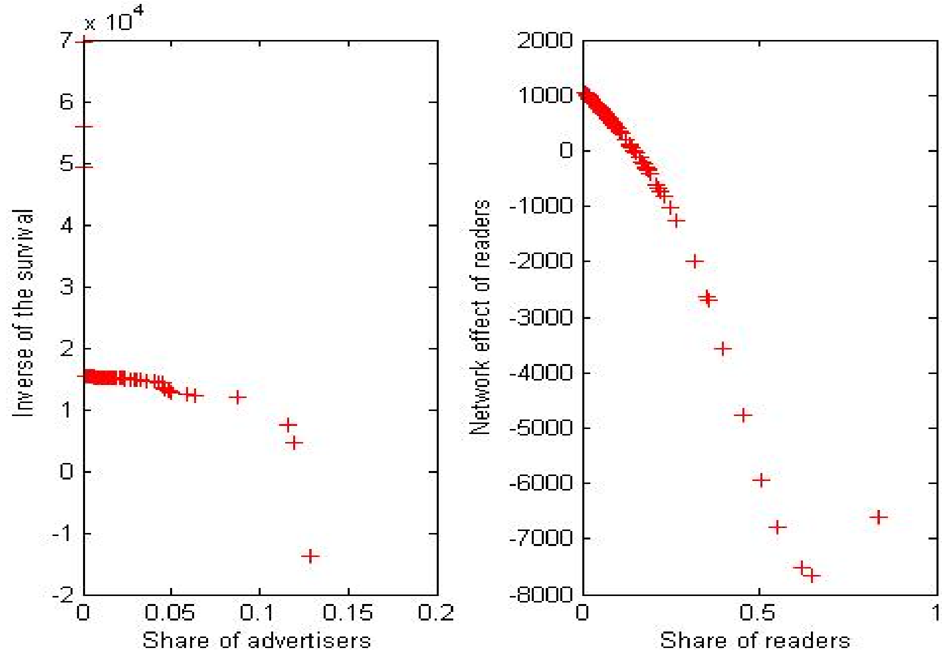

Figure 15: *Estimated functions for advertiser demand equation after monotonization with rearrangement. Estimation done with subsample without women's, tv and animal magazines. Sample size=100*

E Estimation Results of the Parametric Model

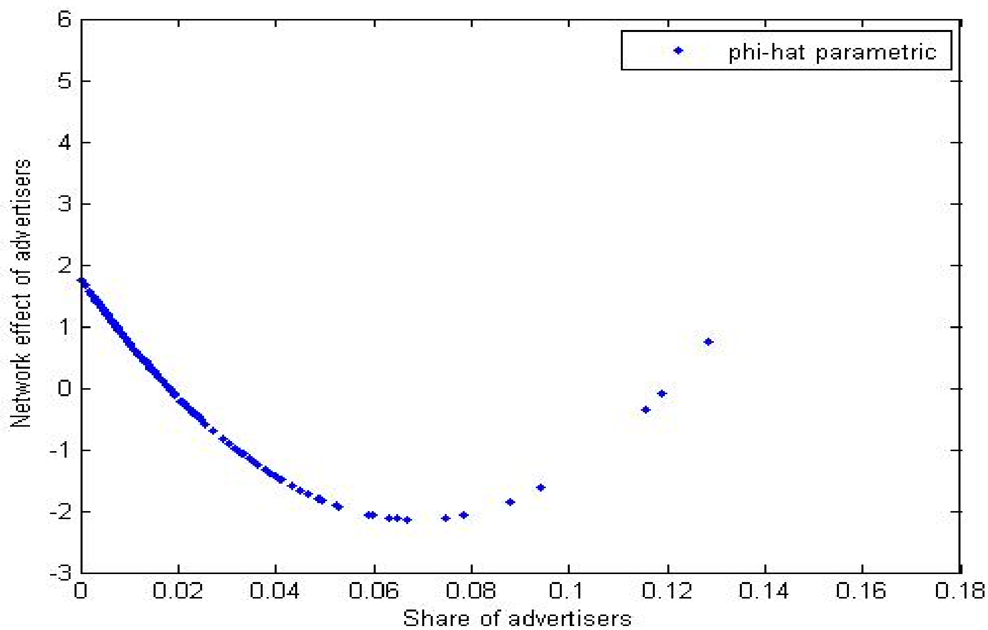

Figure 16: *Parametrically estimated* $\varphi(N^a)$

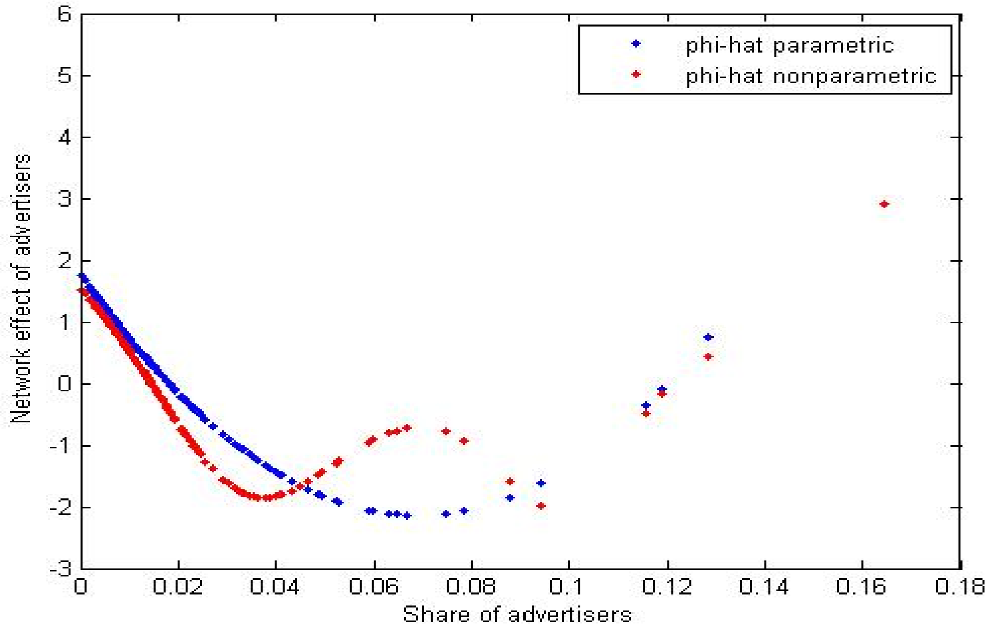

Figure 17: *Parametric and nonparametric estimates of* $\varphi(N^a)$

Table 6: Results for the model in equations (21) and (24)

Variable	Parameter	Estimate	Standard Error	t-stat	p-value
constant	α_0	1.732	0.214	8.08	$< .0001$
share of advertisers	α_1	-117.027	32.26	-3.63	0.0004
$(share\ of\ advertisers)^2$	α_2	850.027	291.4	2.92	0.004
cover price	β	0.941	0.213	4.41	$< .0001$
constant	θ_0	5.083	0.347	14.65	$< .0001$
share of readers	θ_1	14.623	2.510	5.83	$< .0001$
ad rate	γ	-0.00014	0.000013	-10.73	$< .0001$

Table 7: Results for the model in equations (23) and (24)

Variable	Parameter	Estimate	Standard Error	t-stat	p-value
constant	α_0	1.378	0.498	2.77	0.0062
share of advertisers	α_1	-103.988	41.196	-2.52	0.0125
cover price	β	1.296	0.402	3.23	0.0015
constant	θ_0	5.204	0.442	11.77	$< .0001$
share of readers	θ_1	11.198	2.826	3.96	0.0001
ad rate	γ	-0.00013	0.000016	-7.66	$< .0001$

References

Ai, C. and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

Andrews, D. W. (2011): "Examples of L^2-Complete and Boundedly-Complete Distributions," Cowles Foundation Discussion Papers 1801, Cowles Foundation for Research in Economics, Yale University.

Argentesi, E. and L. Filistrucchi (2007): "Estimating market power in a two-sided market: The case of newspapers," *Journal of Applied Econometrics*, 22, 1247–1266.

Argentesi, E. and M. Ivaldi (2005): "Market Definition in the Printed Media Industry: Theory and Practice," CEPR Discussion Papers 5096, C.E.P.R. Discussion Papers.

Armstrong, M. (2006): "Competition in Two-Sided Markets," *RAND Journal of Economics*, 37, 668–691.

Behringer, S. and L. Filistrucchi (2010): "Price Wars in Two Sided Markets: The Case of UK Quality Newspapers in the 90's," Tech. rep., Tilburg University.

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75, 1613–1669.

Blundell, R. and R. Matzkin (2010): "Conditions for the existence of control functions in nonseparable simultaneous equations models," CeMMAP working papers CWP28/10, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Borsenberger, C., M. Ivaldi, E. Malavolti-Grimal, and C. Vibes (2010): "La Presse: Un Marche Biface? Une Etude Empirique," Tech. rep., Toulouse School of Economics.

Carrasco, M. (2008): "A Regularization Approach to the Many Instruments Problem," Tech. rep., Universite de Montreal.

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, Elsevier, vol. 6 of *Handbook of Econometrics*, chap. 77.

Chernozhukov, V., I. Fernandez-Val, and A. Galichon (2009): "Improving point and interval estimators of monotone functions by rearrangement," *Biometrika*, 96, 559–575.

——— (2010): "Quantile and Probability Curves Without Crossing," *Econometrica*, 78, 1093–1125.

Chernozhukov, V., P. Gagliardini, and O. Scaillet (2006): "Nonparametric Instrumental Variable Estimators of Structural Quantile Effects," Swiss Finance Institute Research Paper Series 08-03, Swiss Finance Institute.

Chernozhukov, V., G. W. Imbens, and W. K. Newey (2007): "Instrumental variable estimation of nonseparable models," *Journal of Econometrics*, 139, 4–14.

Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011): "Nonparametric Instrumental Regression," *Econometrica*, 79, 1541–1565.

DHaultfoeuille, X. (2011): "On The Completeness Condition In Nonparametric Instrumental Problems," *Econometric Theory*, 27, 460–471.

Engl, H. W., M. Hanke, and A. Neubauer (1996): *Regularization of Inverse Problems*, Dordrecht: Kluwer Academic Publications.

Evans, D. S. and R. Schmalensee (2008): "Markets with Two-Sided Platform," *Issues in Competition Law and Policy*, 1.

Feve, F. and J.-P. Florens (2009): "The Practice of Non Parametric Estimation by Solving Inverse Problems: The Example of Transformation Models," Tech. rep., 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," Tech. rep., Toulouse School of Economics, Toulouse.

Flachaire, E. (2005): "Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap," *Computational Statistics & Data Analysis*, 49, 361–376.

Florens, J.-P., J. Johannes, and S. V. Bellegem (2009): "Instrumental Regression in Partially Linear Models," Tech. rep., IDEI.

Florens, J.-P. and S. Sokullu (2012): "Nonparametric Estimation of Semiparametric Transformation Models," Tech. rep., Toulouse School of Economics.

FREEDMAN, D. A. (1981): "Bootstrapping Regression Models," *The Annals of Statistics*, 9, 1218–1228.

GALE, D. AND H. NIKAIDO (1965): "THe Jacobian Matrix and Global Univalence of Mappings," *Mathematische Annalen*, 159, 81–93.

HALL, P. AND J. L. HOROWITZ (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *Annals of Statistics*, 32, 2904–2929.

HOROWITZ, J. L. (1996): "Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable," *Econometrica*, 64, 103–137.

——— (2001): "The Bootstrap," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Elsevier, vol. 5 of *Handbook of Econometrics*, chap. 52.

——— (2009): "Applied Nonparametric Instrumental Variables Estimation," Tech. rep., Northwestern University.

HU, Y. AND J.-L. SHIU (2011): "Nonparametric identification using instrumental variables: sufficient conditions for completeness," CeMMAP working papers CWP25/11, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

KAISER, U. (2007): "Do Media Consumers Really Dislike Advertising? An Empirical Assessment of a Popular Assumption in Economic Theory," CIE Discussion Papers 2007-08, University of Copenhagen. Department of Economics. Centre for Industrial Economics.

KAISER, U. AND M. SONG (2009): "Do media consumers really dislike advertising? An empirical assessment of the role of advertising in print media markets," *International Journal of Industrial Organization*, 27, 292–301.

KAISER, U. AND J. WRIGHT (2006): "Price structure in two-sided markets: Evidence from the magazine industry," *International Journal of Industrial Organization*, 24, 1–28.

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

LINTON, O., S. SPERLICH, AND I. V. KEILEGOM (2008): "Estimation of a Semiparametric Transformation Model," *Annals of Statistics*, 36, 686–718.

NEWEY, W. K. AND J. L. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

Rochet, J.-C. and J. Tirole (2003): "Platform Competition in Two-Sided Markets," *Journal of the European Economic Association*, 1, 990–1029.

Rysman, M. (2004): "Competition Between Networks: A Study of the Market for Yellow Pages," *Review of Economic Studies*, 71, 483–512.

Tirole, J. (1988): *The Theory of Industrial Organization*, vol. 1 of *MIT Press Books*, The MIT Press.

VanDalen, R. (2010): "Advertising, Market Power and Network Effects in Radio Broadcasting," Tech. rep., University of Groningen.