Competition Between Networks: A Study of the Market for Yellow Pages

Marc Rysman*
Boston University

January 4, 2000

Abstract

This paper studies the welfare tradeoff between competition and monopoly in a market characterized by network effects. In particular, I estimate a structural model of the market for Yellow Pages directories. A competitive market structure may not be welfare-maximizing because a market with many small directories fails to take advantage of network effects between advertisers and consumers. I estimate a set of three simultaneous equations: consumer demand for usage of a directory, advertiser demand for advertising and a publisher's first-order condition (derived from profit-maximizing behavior). Parameter estimates show that for a given directory, the demand for advertising increases in the amount of consumer usage and that consumer usage increases in the amount of advertising, implying a network effect. I use these estimates of structural parameters to determine the optimal level of advertising in a given market and if the market would benefit from more or less competition. My model allows me to distinguish between the amount of forgone surplus that should be ascribed to standard problems with imperfect competition and the amount that should be ascribed to a failure to take advantage of network effects. The results show that there is under-entry in equilibrium and that network effects are important in driving a wedge between the social optimum and the market equilibrium.

1 Introduction

When a product is characterized by strong network effects, consumers may be best served by a market which is not very competitive. Network effects imply that there is a welfare gain to coordinating economic activity on the same standard, which in many real-world situations

*270 Bay State Rd., Boston, MA 02215, mrysman@bu.edu. I thank Ray Deneckere, John Kennan, Larry Samuelson, numerous seminar groups and especially Phil Haile for advice and encouragement. Also, this research would not be possible without the generosity of National Yellow Pages Monitor, Claritas, Inc., and the Yellow Pages Publishers Association. The Christensen Award in Empirical Economics provided important financial support.

implies using a single brand. Therefore, consumers may benefit from coordination even if doing so confers monopoly status on a producer. This paper studies this welfare tradeoff between competition and monopoly in a market characterized by network effects. Specifically, this paper estimates a structural model of the market for Yellow Pages and uses parameter estimates to determine if welfare increases in the number competing networks (in this case, directories).

The Yellow Pages industry is a striking example of how network effects can determine market structure. Yellow Pages publishers associated with phone companies exhibit very high profits. [1] Although phone companies make their listings available to competitors at reasonable prices, successful independent publishers are still quite rare. [2] In general, entrants are hindered by the fact that Yellow Pages are a network good - the value of the product depends (indirectly) on how many consumers choose to use the product. Consumers value directories based on how much information an advertising are in a directory. Meanwhile, retailers place more advertising in a directory if there are more consumers of a directory. The resulting positive feedback loop between usage and advertising makes it difficult for a new publisher to deliver a competitive product.

The welfare implications of monopoly are ambiguous in the market for Yellow Pages. While a market in which consumer usage and advertising are split between two directories might benefit from competition, a market in which usage and advertising are concentrated in a single book might create higher welfare because consumers and advertisers have wider access to each other. Determining which case holds in practice is the main objective of this paper.

Yellow Pages directories are an example of a good that exhibits *indirect* network effects and addressing this feature is central to the goals of this paper. Common examples of network

[1] A common way for people who work in the Yellow Pages industry to convey the profitability of their product is to compare it to the profitability of illicit narcotics. "We earn more money than anyone this side of the Cali drug cartel," said one person in an interview. Another said "We like to say that we are the second most profitable industry in the world." Statistical evidence of profitability is presented in Section 4.1.

[2] A U.S. Supreme Court ruling established that White Pages are not copyrightable. The expense of copying listings is not excessive for independent publishers so telephone companies just sell their listings to independents. *Feist Publications, Inc. v. Rural Tel. Service Co. 499 U.S. 340 (1991)*

goods are e-mail and fax machines, where the value to owning the product comes directly from the number of other people who own the product. Yellow Pages are different because the value of a directory to users comes through the number of advertisements and not directly from the number of directory users. In this sense, Yellow Pages exhibit indirect network effects. The concept of indirect network effects was introduced under the name *hardware/software paradigm* in Katz and Shapiro (1985), one of the earliest papers on network effects. Hardware becomes more valuable when more compatible software is supplied, and the amount of software available depends on the amount of hardware that consumers purchase. The names "hardware" and "software" should not restrict readers to thinking about the computer industry. Katz and Shapiro (1994) present a number of different examples. For instance, credit card networks fit the paradigm, where the card is the hardware and merchant acceptance is the software.

In the case of Yellow Pages, the "hardware" is the directory and the "software" is the advertising. I measure activity in the "hardware" market by consumer usage of directories and activity in the "software" market by the price and quantity of advertising. Explicitly modeling the interaction between the "hardware" and "software" markets sets this paper apart from previous empirical work. A prerequisite for this approach is to have an appropriate data set. Collecting a rich data set on a single market is often a difficult task, and this paper (or any paper which is interested in indirect network effects) requires collecting data on two markets. Fortunately, Yellow Pages are of interest to advertisers so high-quality data are available both on the prices and exposure of Yellow Pages directories.

My data set contains directory-level observations of the price and quantity of advertising, as well as the number of references that each directory receives. My basic data set consists of observations on 476 directories in 53 MSA's drawn from the year 1996. The relatively high number of observations is an obvious asset. Previous studies of network effects focus on industries in which most of the variation occurs across time, limiting the number of observations available. In contrast, the market for a Yellow Pages directory is a small geographic region (I observe directory boundaries at the level of 5-digit zip codes), allowing for a large number of observations in a cross-sectional data set. An especially interesting aspect of this industry is that, although telephone companies dominate most markets, independent publishers have established a presence in a number of regions. Therefore, I observe geographically distinct markets with different numbers of competitors. This feature is crucial for studying the welfare tradeoff between competition and standardization under network effects.

I aim to quantify the welfare losses due to network effects in equilibrium and determine the optimal number of directories for a given market. In order to do so, I estimate the demand for directory usage by consumers, the demand for advertising by retailers and a first-order condition derived from profit-maximizing behavior by publishers. In order to model the consumers' usage decision, I follow the approach of Berry (1994) for estimating demand in markets characterized by discrete choices and oligopoly. However, estimation is not straightforward because directories have distinct distribution areas with overlapping boundaries. A directory may be the only one available in a portion of its market but face competition in another. I adapt Berry's techniques to this environment. I cannot solve for unobservable terms explicitly so I nest a fixed point algorithm in my optimization routine, as suggested by Berry, Levinsohn and Pakes (1995) and Rust (1987). Because the system of equations makes non-linear restrictions on parameters, I estimate via the Generalized Method of Moments (Hansen, 1982; Davidson and Mackinnon, 1993).

To test for a network effect, I test whether consumer demand for usage increases in advertising and whether advertiser demand for advertising increases in usage. I then use the estimated structural parameters to analyze how equilibrium outcomes differ from the market optimum. First, I measure how much do equilibrium quantities of advertising differ from the optimum. I calculate the associated welfare losses and determine how much forgone surplus should be ascribed to standard problems with imperfect competition and how much should be ascribed to the network effect. Secondly, I take profit-maximizing directories as given and ask if there is too much or too little entry in equilibrium.

I find that the amount that consumers use a directory increases in the directory's level of advertising. I also find that retailer demand for advertising in a directory increases in the amount that consumer use a directory. Together, these results imply that a network effect exists. I find that forgone surplus due to a failure to take advantage of network effects is large relative to the amount of deadweight loss resulting from imperfect competition and relative to the amount of surplus realized in the market equilibrium. However, network effects are not so strong as to counteract the benefits of entry. The results show that the entry of independent publishers improves welfare, so rules which force telephone company publishers to facilitate entry are welfare-enhancing. The results also show that independent directories create more social benefit than they capture in private benefit. Thus, I find that entry is less than optimal in this market.

This paper is of independent policy relevance because the decision about whether or not to open the Yellow Pages market to competition is now in question. Although independent publishers have always been allowed in the United States, recent policy strengthens their position. In addition to the *Feist* decision mentioned in footnote 2, the 1996 Telecom Act has a provision requiring telephone companies to sell their listings to independents for a "reasonable rate". Similar policies have recently been enacted in Europe and Australia. My results suggest these policies are socially beneficial, as more competitive markets are preferable to monopoly markets.

The methodology developed here is also interesting because it could be applied to other industries. The methodology is relevant to any industry characterized by indirect network effects and incompatible networks. A particularly topical example is the competition between Microsoft Windows and its rivals in the market for operating systems. Like Yellow Pages directories, computer operating systems exhibit indirect network effects. In the terminology of the hardware-software paradigm, the operating system plays the role of "hardware" and accessory software is "software". Also, directories and operating systems are both incompatible networks. An advertisement placed in one directory confers no benefit on the user of another directory, just as software designed for one operating system confers no benefit on the user of a different operating system. Central to Microsoft's defense of its large market share is the claim that the benefits from standardization on its Windows system outweigh the costs of its monopoly power. That claim is exactly what the methodology in this paper is designed to test.

2 Related Literature

Network effects, and in particular indirect network effects, are the subject of increasing attention in the economics literature as well as in policy discussions. In the economics literature, many theoretical papers have appeared since Katz and Shapiro (1985), although the first formal models of indirect network effects were not developed until Chou and Shy (1990) and Church and Gandal (1992). Chou and Shy (1990) and Economides and Flyer (1997) consider entry in models of network effects and find that the welfare effects are ambiguous. Surveys of the literature on network effects are Katz and Shapiro (1994) and Economides (1996).

The two most closely related empirical papers are Park (1997) and Gandal, Kende and Rob (1997). Park studies the competition between the VHS and Beta standards for VCR's. Park deals with difficult issues of how consumers form expectations of the size of incompatible networks which are in flux. Park's paper is different from this one in that it does not address the issue of indirect network effects, although the interaction between sales of VCR's and sales of pre-recorded movies is potentially important.

Gandal, Kende and Rob (1997) study the entry decisions of the producers of compact disk players and the producers of compact disks. Like my paper, their paper explicitly models an indirect network effect. However, they study firms producing for a compatible standard. Because the standard is non-proprietary, there are no benefits to having a small number of firms so the central issue studied in my paper does not arise. Other empirical papers which study network effects include Augereau (1998), Economides and Himmelberg (1995), Gandal, Greenstein and Salant (1997), Saloner and Shepard (1995), Gandal (1994) and Greenstein (1993).

The two empirical models that are most similar to the one here are by Rosse (1970) and Berry and Waldfogel (1999). Although Rosse's model is designed to identify the cost curve of a newspaper as opposed to measure network effects, his model does allow him to measure the feedback between readership and advertising. However, one would expect readers' valuation of newspaper advertisements to be ambiguous. [3] One reason that I choose to focus on Yellow Pages directories are that they are valuable explicitly because of the advertisements. Note that Rosse's model is not derived from individual maximization problems as is the case in this paper. Berry and Waldfogel analyze the effects of entry by radio stations. They lack some station specific data (e.g. listenership, price, geographic coverage) and so are constrained from addressing some of the issues explored here.

3 Model

This section presents a model of competition in the Yellow Pages industry. The model explicitly captures the interaction between directory advertising and directory usage. Estimates of the parameters used here appear in Section 6. The model is a one period, simultaneous move game. There are n publishers that each produce a directory. The distribution areas of the directories may or may not overlap, and may do so for only a portion of their area. The amount of overlap is taken to be exogenous to this game. Each publisher i, $i = 1, ..., n$, simultaneously chooses its quantity of advertising A_i to sell in its directory. The solution concept is Nash equilibrium.

The price that a publisher receives depends on how many consumers use its directory. I assume that there is some exogenous number of times M that consumers require information of the type that can be found in the Yellow Pages. Consumers can use one of the directories in their area or an outside option. [4] The average number of references to book i per consumer in i's distribution area is denoted by $U_i(A_1, ..., A_n)$, $i = 0, ..., n$. The function $U_0(A_1, ..., A_n)$ is the average number of references per consumer to the outside option. The functions $U_i(.)$ represent consumer demand for usage. I expect to find $\partial U_i/\partial A_i > 0$ and $\partial U_i/\partial A_j \leq 0$, $j \neq i$.

[3] In Rosse's paper, the parameter estimate which captures the effect of advertising on readership is positive but not significant.

[4] Industry sources feel they have no close competitors. Many Internet sources seem to substitute for the Yellow Pages but the closest competitor is probably word-of-mouth.

Only consumers that live in the distribution area of a directory may use that directory, so $\partial U_i / \partial A_j = 0$ if the distribution areas of i and j do not overlap. Publishers do not collect revenue from consumers directly but the amount of consumer usage of a directory affects advertisers' factor demand for advertising in that directory. I now show how the factor demand of individual advertisers aggregates into the inverse demand curve for advertising.

3.1 Deriving the Inverse Demand Curve for Advertising

This section derives the inverse demand curve for advertising from the advertisers' profit maximization problem, modeling the advertisers as price-takers. Simple assumptions about the advertisers' problem imply specific restrictions on the resulting inverse demand curve.

A representative advertiser chooses the amount of advertising a_i (a continuous choice variable) to place in directory i. The advertiser acts as a price taker and chooses its optimal level of advertising, $a_i(P_1, ..., P_n)$. Total advertising in book i is A_i, so $int_0^{\bar{k}} a_i(P_1, ..., P_n) dk = A_i(P_1, ..., P_n)$, where $\bar{k}$ measures the size of the market of advertisers. By inverting the aggregate demand curve, we can construct an inverse demand curve that the publisher faces.

In order to derive the factor demand for advertising $a_i(P_1, ..., P_n)$, I use the following set-up: Let the number of people who use book i and look at the representative advertisement be $L_i = L(a_i, U_i, A_i)$. The function increases in a_i and U_i. The function decreases in A_i because a given advertisement is less likely to be seen in a large book. Some portion of the consumers who look at an advertisement will make contact with the advertiser, and these consumer contacts will generate some level of profit. Profit for the advertiser is:

$$\Pi = \pi(L(a_1, U_1, A_1), ..., L(a_n, U_n, A_n)) - P_1 a_1 - ... - P_n a_n. \tag{1}$$

The function π represents the profit to the advertiser from the number of people who look at the advertisement, net of the cost of serving those people. Now I make two assumptions:

A1 Consumers use at most one directory per information requirement.

A2 Advertiser profit per look is constant.

The first assumption says that each of the M times that a consumer requires information, the consumer uses the outside option or a single directory. The consumer does not open up multiple directories simultaneously. The first assumption is already implicit in Equation 1. I write π as a function of the number of looks an advertisement receives from people who use each book. I could also have included terms capturing the number of looks an advertisement receives from people who use two books, i.e. L terms for each combination of books i and j, $i \neq j$. It would then be natural to include terms for people who use three books, etc. Assumption *A1* eliminates all of those terms. [5] The second assumption says that the function π is linear in the remaining arguments.

Note that these two assumptions imply that π is separable in a_i. Intuitively, the first assumption implies that consumers observe only the advertisements in one book, so advertising in one book is not a substitute or a complement for advertising in another book. As a result, there is no demand-side reason why the choice of advertising at one book should affect the choice at another book. The second assumption says that having many customers as a result of one advertisement does not affect the cost or benefit of serving customers generated by another advertisement. An implication is that there is no cost-side reason why the choice of advertising at one book affects the choice at another book. Separability of the profit function follows. Π can be rewritten as:

$$\Pi = \pi_1 L(a_1, U_1, A_1) - P_1 a_1 + ... + \pi_n L(a_n, U_n, A_n) - P_n a_n.$$

While the following results can be worked out for general demand functions with suitable restrictions, nothing is lost by introducing functional forms now in anticipation of the estimation

[5] This assumption can be motivated by more than just casual observation. My data show that when consumers pick up a directory, they contact an advertiser 82.1% of the time, but make more than one contact only 36.2% of the time. These data fit with consumers who use only one directory, although it is possible that consumers use multiple books and make a single contact in each. I present a statistical test of Assumptions *A1* and *A2* in Section 6.

section. I assume that $L(a_i, U_i, A_i)$ has the Cobb-Douglas form, so $L_i = a_i^{\gamma_1} A_i^{\gamma_2} U_i^{\alpha_1}$. The parameter γ_1 is expected to lie between 0 and 1, and captures decreasing returns to large advertisements. I expect that parameter γ_2 will be negative capturing the business stealing effect, i.e. the fact that an advertisement might get lost in a large directory. The parameter α_1 should be positive because more usage of a directory increases the likelihood of consumers looking at a given advertisement. Let $\pi_i = \pi_i e^{\hat{\nu}_i}$. The variable π_i captures directory specific factors such as the number of people that the directory covers. The term $\hat{\nu}_i$ represents factors which are specific to the relationship between directory i and the advertiser, such as the geographic scope of the directory and the location of the advertiser, or the nature the of the advertiser's business. [6]

Now the profit function can be written as:

$$\Pi = \pi_1 a_1^{\gamma_1} A_1^{\gamma_2} U_1^{\alpha_1} e^{\hat{\nu}_1} - P_1 a_1 + ... + \pi_n a_n^{\gamma_1} A_n^{\gamma_2} U_n^{\alpha_1} e^{\hat{\nu}_n} - P_n a_n.$$

The advertiser picks a_i to maximize $\pi_i a_i^{\gamma_1} A_i^{\gamma_2} U_i^{\alpha_1} e^{\hat{\nu}_i} - P_i a_i$, given price P_i. The representative advertiser is too small to affect A_i and takes it as given. The optimal a_i is:

$$a_i = \left(\frac{P_i}{\gamma_1 \pi_i A_i^{\gamma_2} U_i^{\alpha_1} e^{\hat{\nu}_i}} \right)^{\frac{1}{\gamma_1 - 1}}.$$

Let $\nu_i = (1 - \gamma_1) \ln(\int_0^{\bar{k}} e^{-\hat{\nu}_i/(\gamma_1 - 1)} dk)$. Then [7]

$$A_i = \int_0^{\bar{k}} a_i dk = \left(\frac{P_i}{\gamma_1 \pi_i A_i^{\gamma_2} U_i^{\alpha_1} e^{\nu_i}} \right)^{\frac{1}{\gamma_1 - 1}}.$$

Solving for P_i, we obtain the inverse demand curve:

$$P_i(A_i, U_i) = \gamma_1 A_i^{\gamma_1 + \gamma_2 - 1} U_i^{\alpha_1} \pi_i e^{\nu_i}. \tag{2}$$

[6] As I discuss below, it is straightforward to introduce advertiser heterogeneity via the parameter $\hat{\nu}_i$.

[7] Note that it is straightforward to obtain these results with advertiser heterogeneity. Let there be a continuum of advertisers indexed by $k \in [0, \bar{k}]$ distributed $f(k)$. Denote the choice of advertiser k at book i as a_{ik}, so $A_i(P_1, ..., P_n) = \int_0^{\bar{k}} a_{ik}(P_1, ..., P_n) f(k) dk$. Denote π_i as $\pi_{ik} = \pi_i e^{\nu_{ik}}$ and let $\nu_i = (1 - \gamma_1) \ln(\int_0^{\bar{k}} e^{-\nu_{ik}/(\gamma_1 - 1)} dk)$. This change of structure still leads to Equation 2.

There are several important features of this demand curve. First, it should increase in usage (I expect $\alpha_1 > 0$). This feature, along with the fact that usage increases in advertising, represents the network effect. Second, the demand curve decreases in advertising both because there are decreasing returns to individual advertisers from large ads (I expect $\gamma_1 < 1$) and because advertisements will be lost in a large book (I expect $\gamma_2 < 0$), but these effects will not be distinguishable in estimation.

Third, the price of advertising and the amount of usage in another book do not affect the amount of advertising in book i directly. Holding usage at book i constant, advertisers are willing to pay the same amount to advertise in book i regardless of what book j does. This result follows from the fact that users of book i do not look at book j and the advertisers have constant returns to scale. The price at i does not depend on the quantity at j directly, but usage at i depends on the amount of advertising in all directories so A_j can affect P_i indirectly. Thus, the amount that one directory overlaps with another directory matters only through usage. The inverse demand curve $P_i(A_1, ..., A_n, U_0(A_1, ..., A_n), ..., U_n(A_1, ..., A_n))$ can be rewritten as $P_i(A_i, U_i(A_1, ..., A_n))$. This feature makes structural estimation significantly easier and can be tested.

3.2 Equilibrium

Given these inverse demand curves, publishers choose advertising simultaneously. Publisher i faces two demand curves, the demand for advertising $P_i(A_i, U_i)$ and the demand for usage $U_i(A_1, ..., A_n)$. As stated previously, the publisher obtains no revenue directly from the demand for usage but the level of usage chosen affects the price of advertisements. The network effect is represented by the combination of partial derivatives, $\partial P_i/\partial U_i > 0$ and $\partial U_i/\partial A_i > 0$. Also, we have the "scarcity effect", $\partial P_i/\partial A_i < 0$, that the demand curve slopes down when ignoring the network effect.

Separating the network effect from the scarcity effect is important both for understanding how the publisher chooses advertising and for measuring consumer surplus. Figure 1 demonstrates these issues. Figure 1 shows the demand for publisher i's directory holding the competitors' choices fixed, so reference to competitors are suppressed for clarity. The bottom panel shows the demand for usage, $U_i(A_i)$. Usage levels u_i' and u_i'' correspond to advertising levels A_i' and A_i''. The upper panel shows how demand differs from willingness-to-pay. $P_i(A_i, U_i(A_i))$ is the demand curve, the price that the publisher receives for any given level of advertising. Willingness-to-pay curves are demand curves with usage held constant. $P_i(A_i, u_i')$ and $P_i(A_i, u_i'')$ represent willingness-to-pay curves for the usage levels u_i' and u_i''. On each willingness-to-pay curve $P_i(A_i, u_i)$, there is one point which is also on the demand curve. That point on the willingness-to-pay curve is the one at which the amount of advertising implies the level of usage in question, i.e. A_i such that $u_i = U_i(A_i)$. The locus of these points makes up the demand curve $P_i(A_i, U_i(A_i))$. The slope of each willingness-to-pay curve is measured by $\frac{\partial P_i}{\partial A_i}$. The slope of the demand curve which the publisher faces is $\frac{\partial P_i}{\partial A_i} + \frac{\partial P_i}{\partial U_i}\frac{\partial U_i}{\partial A_i}$. [8]

The model has n sets of three endogenous variables, A_i, P_i and U_i, and two demand curves for each set: the inverse demand for advertising $P_i(A_i, U_i)$ and the demand for usage $U_i(A_1, ..., A_n)$. The model is closed by the publishers' first-order conditions. Let $C_i(A_i)$ be the cost to i of producing a directory with advertising level A_i. Publisher i solves:

$$\max_{A_i} P_i(A_i, U_i(A_1, ..., A_n))A_i - C_i(A_i).$$

The Appendix assumes a logit structure for usage (as is used in estimation) and establishes that an equilibrium in pure strategies exists when network effects are not "too large" and one always exists when the directory distribution areas perfectly overlap. An equilibrium in pure strategies exists for the parameter estimates found in this paper.

[8] Economides and Himmelberg (1995) derive a similar chart to the top panel in Figure 1 and point out that in many cases, network effects imply an upward sloping demand curve for low levels of quantity. Under strong network effects, no one will pay for a product which no one else uses.

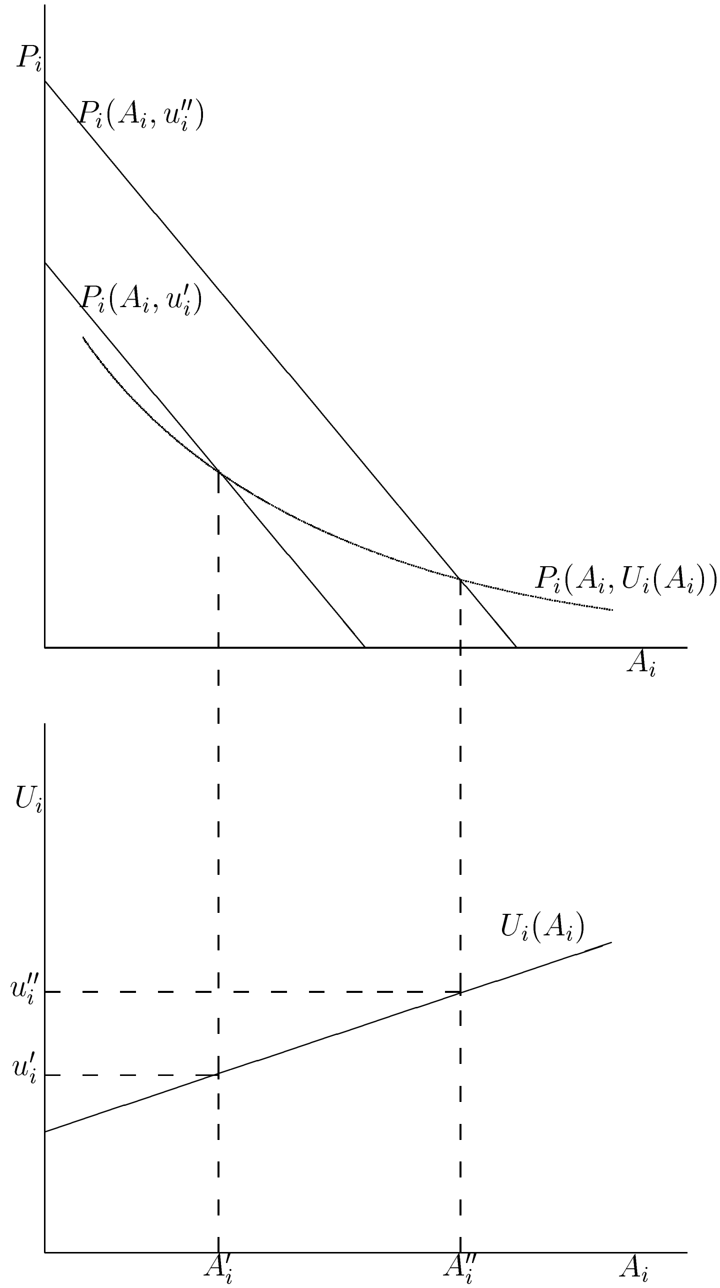


Figure 1: The interaction between two demand curves.

3.3 Efficiency

Consumers obtain directories for free and consumers do not have the option to pay for a directory with extra advertisements. The lack of prices for the consumer side of the market makes it difficult to convert consumer utility into units which are easily comparable to retailer surplus. I will analyze surplus formally only from the point of view of advertisers but I will discuss consumer surplus when I am able to do so.

To see how the equilibrium choice of advertising differs from the socially optimal choice, compare the first-order conditions solved by the publisher and the social planner. Publisher i picks A_i to solve:

$$P_i + \frac{\partial P_i}{\partial A_i} A_i + \frac{\partial P_i}{\partial U_i}\frac{\partial U_i}{\partial A_i} A_i - MC(A_i) = 0. \tag{3}$$

In contrast, a social planner would choose the set of advertising levels A_i, $i = 1, ..., n$, simultaneously to maximize $\sum_{i=1}^{n} \int_0^{A_i} P_i(s, U_i(A_1, ..., A_n))ds - C(A_i)$. The first-order condition for the social planner is:

$$P_i + \sum_{j=1}^{n} \int_0^{A_j} \frac{\partial P_j(s, U_j(A_1, ..., A_n))}{\partial U_j} \frac{\partial U_j}{\partial A_i} ds - MC(A_i) = 0.$$

These first-order conditions will rarely coincide. [9] It cannot be said for sure which regime will result in more advertising. The social planner accounts for the value of the network effect to the entire set of advertisers and the advertisers in other directories. The publisher takes into account the value of the network effect on its marginal purchaser. However, the publisher also takes into account the effect of downward sloping demand on marginal revenue and in most reasonable cases, we have the standard result that the size of the network is too small. By "reasonable," I mean that the network effect of advertising on a publisher's own price is stronger than the

[9] I have assumed that the publishers can set only one price. If a publisher can first-degree price discriminate, it could achieve the socially optimal level of advertising. This result is the case in Liebowitz and Margolis (1994) who assume that a monopolist faces identical purchasers, which implies that the monopolist can charge consumers their valuation with a single price.

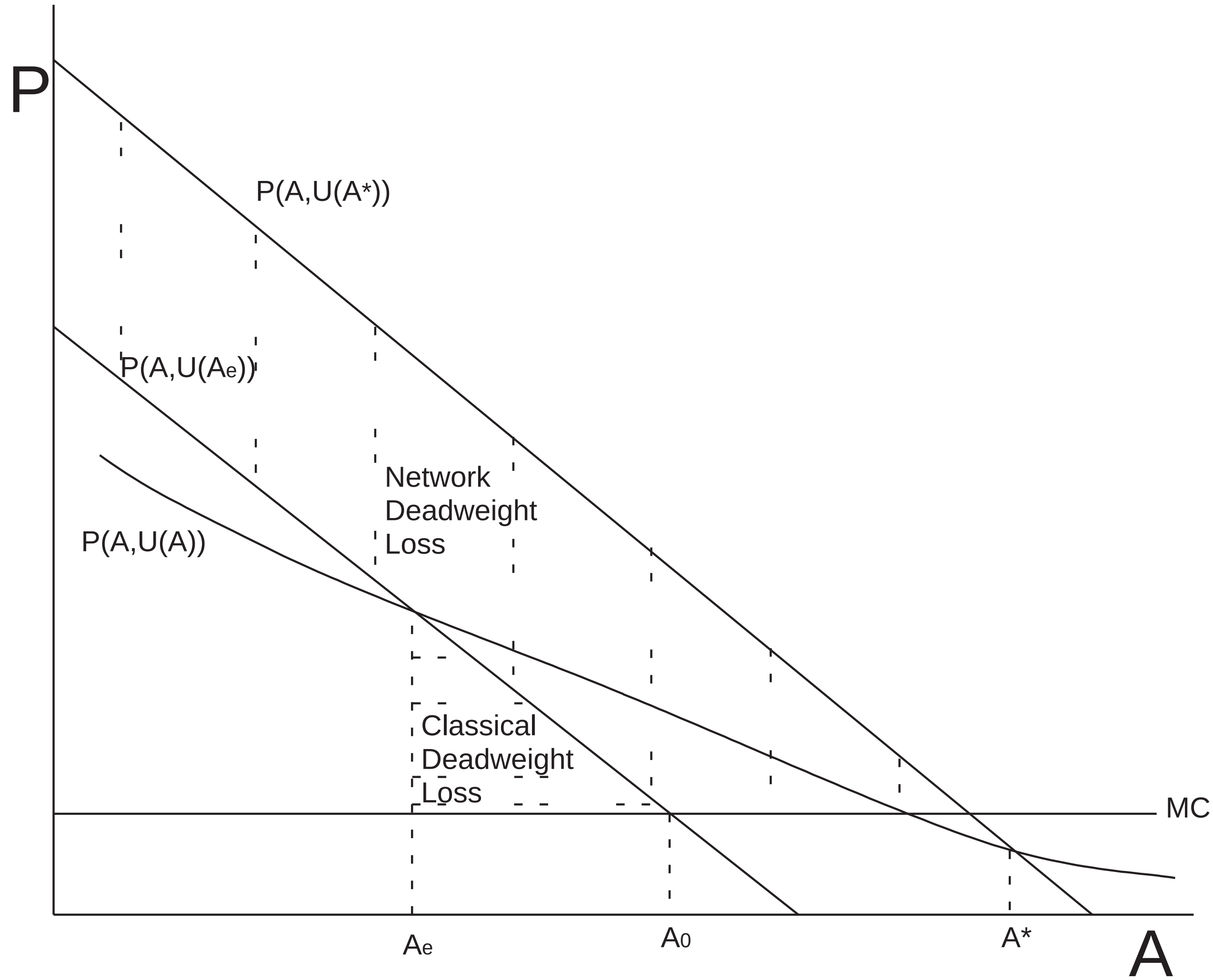


Figure 2: Deadweight Loss

negative effect on its competitors' prices, so $\sum_{j=1}^{n} \frac{\partial P_j}{\partial U_j} \frac{\partial U_j}{\partial A_i} > 0 \; \forall A_i, \; i = 1, ..., n$. In this case, the social planner picks A_i such that price is less than marginal cost. "Reasonable" also means that demand is downward sloping, even accounting for the network effect, so $\frac{\partial P_i}{\partial A_i} + \frac{\partial P_i}{\partial U_i} \frac{\partial U_i}{\partial A_i} < 0$ at the Nash equilibrium, which would imply that price is greater than marginal cost in the Nash equilibrium. I will be able to test for these conditions in the data and I return to them in Section 6.

I will be able to estimate the demand for advertising and use those estimates to measure total surplus at the social optimum and at the Nash equilibrium. Note that in equilibrium, some of the forgone surplus is a result of publishers exercising market power while some is due to the market's failure to account for the effects of aggregate advertising on the profits of individual advertisers. I will differentiate between these two sources of inefficiency to get a measure of network deadweight loss, the deadweight loss specifically due to the market's failure to take account of the network effect. Using a model which distinguishes between the two effects that a change in quantity has on price, the scarcity effect and the network effect, is important for this exercise.

Figure 2 demonstrates. I fix the competitors' outcomes at their Nash equilibrium choices and ask what a social planner would choose for book i. Let the equilibrium choice of advertising for publisher i be A_e. I define *classical deadweight loss* to be the difference in surplus between what the Nash equilibrium achieves and what a social planner would achieve if there was no network effect. Without network effects, the social planner would take the demand curve to be the willingness-to-pay curve $P(A, U(A_e))$. Denote the social planner's choice of advertising level as A_0. Classical deadweight loss is:

$$\int_{A_e}^{A_0} P_i(s, U_i(A_e, A_{-i}))ds - C_i(A_0 - A_e).$$

A_{-i} is the vector of levels of advertising chosen by competing directories in equilibrium. A_0 is the choice the social planner would make if usage were fixed at $U_i(A_e, A_{-i})$. The term $C_i(A_0 - A_e)$ represents the cost of the extra production, assuming constant marginal cost (as is assumed in estimation).

The definition of *network deadweight loss* is the difference between the amount of surplus generated when the social planner takes into account the network effect and when it does not. If a social planner raised advertising from A_e to A_0, usage would rise which means that advertising would generate more surplus, so the efficient level of advertising would be even higher. As stated previously, it will often be efficient for the willingness-to-pay of the last purchaser to be below marginal cost because of the network effect. Denote the socially efficient level of advertising by A_*. Network deadweight loss is defined as:

$$\int_0^{A_*} P_i(s, U_i(A_*, A_{-i}))ds - \int_0^{A_0} P_i(s, U_i(A_e, A_{-i}))ds. - C_i(A_* - A_0).$$

The point A_* is the efficient level of advertising for directory i given that the competitors outcomes are the Nash equilibrium levels. In the figure, the space between the efficient willingness-to-pay curve and the equilibrium willingness-to-pay curve is network deadweight loss. I use estimates of structural parameters to measure network deadweight loss, classical deadweight loss and equilibrium consumer surplus. Clearly, it is important to use a model which distinguishes between $\frac{\partial P}{\partial A}$ and $\frac{\partial P}{\partial A} + \frac{\partial P}{\partial U}\frac{\partial U}{\partial A}$ in order to perform this calculation. I return to these issues in the Section 6. [10]

4 Industry and Data Characteristics

As stated in the introduction, the Yellow Pages industry is an excellent one for study both because of the structure of the market and because of the availability of excellent data. I first discuss industry characteristics and how they motivate important assumptions. I then review the data that I work with.

4.1 Industry Characteristics

The Yellow Pages generated \$11.5 billion in sales in 1997 (Elliott, 1998). Yellow Pages directories published by telephone companies earn exceptional profits. A directory serving 240,000 people averages over \$6,000 per page in revenue from display advertisements alone. The average size of such a book is 621 pages so the book brings in around \$3.8 million in revenue. Industry sources estimate that variable costs of production for a book of this size would

[10]Another reasonable choice for the definition of A_o would be the level of advertising that sets demand $P(A, U(A))$ equal to marginal cost. This choice has some appeal but using $P(A, U(A))$ clearly involves network effects. Because I want to isolate network effects as much as possible, I define A_o with the willingness-to-pay curve, so usage is held constant.

be less than $1 million. Indeed, Yellow Pages industry sources estimate that profits represent 35-45% of revenue. Judge Harold Greene, in overseeing the break-up of AT&T, recognized the supra-normal profits associated with Yellow Pages production. He assigned the publishing companies to local phone companies in order to hold down local rates. Many public service commissions provide estimates of the increase in access line charges that would be required if local phone companies did not receive Yellow Pages profits. Most states that compute such a number estimate it to be between $1.80 and $3.50 per line per month for every line in their state (NARUC, 1994-1995).

While the Yellow Pages industry has traditionally been dominated by telephone companies who never enter each other's markets, there is some competition from independent publishers. Table 5 shows that the median person receives two directories from separate publishers. Almost 60% of the population in this data set receives 2 or 3 directories and more than 50% receive directories from more than 1 publishers. However, independent publishers have not been overly successful. Directories associated with telephone companies average 6.42 references per household per month, while the same number at independent directories is only 1.32. Directories associated with telephone companies are on average almost twice the size (746 pages compared to 413) and charge about twice as much ($2,014 to $1,221 for a double-quarter column advertisement) as at independent publishers. A better measure of the difference comes from comparing directories with distribution areas that perfectly overlap. Bell Atlantic and R.H. Donnelly both have directories with distribution areas that are exactly equal to the boundaries of Washington D.C. The phone company's (Bell Atlantic's) directory is 1,443 pages, charges $3,387 for a double-quarter column advertisement and collects 7.6 references per household per month. In contrast, the Donnelly directory is 947 pages, charges $2,352 for the same size advertisement and collects 1.4 references per household per month. Despite these numbers, independent publishers are becoming more important. In fact, 38% of the directories in my data set are published by independent publishers. In my data set, about 25% of the directories published by telephone companies face a competitor who takes 25% of the usage market or more. Each directory can be seen as a competing incompatible network, which makes this data set excellent for use in studying the tradeoffs between competition and standardization.

I make two assumptions about the behavior of publishers based on *a priori* observation of the industry. First, although Yellow Pages publishing is dominated by regulated utilities, I model publishers as profit maximizers. Publishers vigorously protect Yellow Pages profits, often transferring them to for-profit subsidiaries or otherwise hiding profits (White and Sheehan, 1992). For instance, the Ameritech phone company collects a fixed licensing fee from Ameritech Publishing which does not depend on the publisher's financial performance.

The second assumption is on the publisher's choice over the number of books to distribute. I assume that the number of books distributed is exogenous to the price- and quantity-setting process. I make this exogeneity assumption because practically all publishers choose to distribute one Yellow Pages directory to every phone line. Phone companies do so because they are required to distribute White Pages directories to every phone line in their area and most publishers choose to publish Yellow Pages in the same book as the White Pages. An industry source claimed that over 90% produce Yellow Pages and White Pages in the same book. Even for independent publishers, industry sources say that they always choose to distribute one book to every phone line in their area. Therefore, the number of Yellow Pages directories to distribute is determined not by a marginal condition but instead by the geographic scope of each directory. Geographic scoping is determined before the sale of advertising takes place, so I take it as exogenous to the price- and quantity-setting process. Population coverage is highly correlated with the price and quantity of advertising so it will be a powerful instrument for those endogenous variables.

4.2 Data

My data set draws on a number of independent data sets. The centerpieces come from the National Yellow Pages Monitor (NYPM) and Claritas Inc., both proprietary data firms, and the Yellow Pages Publisher's Association (YPPA), an industry trade group. NYPM collects usage data for individual directories. At the request of a client (normally a large Yellow Pages publisher), NYPM measures the number of references per household per month going to all directories in a given metropolitan statistical area. The data set takes each directory's distribution area as a unit of observation and presents the number of references to each directory in that area. Note that distribution areas do not perfectly overlap. A weakness in the NYPM data (for my purposes) is that it is possible to tell how many references go to Book A in Book A's area and how many go to Book B in Book B's area, but one cannot distinguish between how much usage goes to Book A in the area where it overlaps with Book B and the area where it does not. This issue will create a difficult estimation problem, to be discussed in Section 5. The NYPM data contains data on 476 directories.

I use the number of pages in a directory as a proxy for the quantity of advertising. The YPPA maintains a library of directories published by members and the Boston Consulting Group collected the number of pages in each directory (for a separate project), which I obtained. The data on page numbers includes a check for directories that are observably small. Industry sources say that advances in pagination techniques (eliminating empty page space) make this approximation inaccurate over time but useful in a cross-sectional study such as this one. YPPA membership covers over 6,500 directories including almost every directory in the United States. The YPPA claims that its membership accounts for 95% of Yellow Pages advertising sales.

To augment the usage and quantity data, I add advertising prices from the *Rate and Data* data set and directory statistics from the *Industrial Characteristics* data set, both from the YPPA. The pricing data is an especially rich source. The *Rate and Data* set contains prices for every size and color of advertisement at every directory in the YPPA. Directories in my data set offer an average of around 80 choices of advertisement size and color. Unfortunately, I have only one observation on quantity to match with each set of prices. In the next sub-section, I discuss issues of non-linear pricing in depth. The Industrial Characteristics data set contains directory information such as the number of people a directory covers, and the number of columns in a directory.

	Variable	Usage	Advertising	Rate (DQC)	Rate (Full Page)	Rate (Extra Line)	Population Cov.	Telephone Co.
MEAN		4.83	2,564.82	1,781.30	14,039.01	65.24	385.41	0.64
STD DEV		4.16	2,212.71	1,152.78	10,120.32	37.16	409.52	0.48
Obs.		431	431	431	404	430	431	431
Correlation	Usage	1.00	0.53	0.37	0.57	0.27	0.08	0.62
	Advertising	0.53	1.00	0.69	0.82	0.47	0.66	0.42
	Rate (DQC)	0.37	0.69	1.00	0.91	0.80	0.76	0.34
	Rate(FP)	0.57	0.82	0.91	1.00	0.65	0.71	0.52
	Rate (EL)	0.27	0.47	0.80	0.65	1.00	0.57	0.26
	Pop. Cov	0.08	0.66	0.76	0.71	0.57	1.00	0.06
	Telco	0.62	0.42	0.34	0.52	0.26	0.06	1.00
	% Urb. Pop	-0.14	0.17	0.14	0.17	-0.03	0.24	-0.04
	% Diff Cty.	0.19	-0.02	-0.11	-0.05	-0.10	-0.20	0.09
	% Diff State	-0.09	0.21	0.08	0.08	0.01	0.17	-0.06
	% Pub. Trans.	-0.01	0.20	0.30	0.30	0.09	0.27	0.01
	% Own Occ Hs	0.21	-0.07	0.01	-0.02	0.15	-0.17	0.16
	% HS Grad	-0.03	0.11	0.09	0.09	0.07	0.01	0.05
	% Col Grad	-0.11	0.16	0.17	0.17	0.03	0.11	0.01
	Per Cap Inc.	0.02	0.16	0.22	0.23	0.01	0.08	0.07
	Earnings	-0.08	0.17	0.27	0.24	0.08	0.21	0.01
	House Permits	-0.01	-0.02	0.05	-0.04	0.24	-0.06	0.10
	Growth Rate	-0.04	-0.10	-0.13	-0.17	0.03	-0.16	0.01
	Pop. Density	-0.08	0.15	0.23	0.25	0.04	0.27	-0.03

Table 1: Simple Statistics.

I obtained directory boundary data from Claritas, Inc. The boundary data came in the form of computer maps which I matched up with the population centers of 5-digit zip codes. I assume that if a population center of a zip code falls within the boundary of a directory's distribution area, then the publisher distributes the directory to the entire zip code. Zip codes seem to be a reasonably close approximation of directory boundaries, and in some cases coincide exactly. Claritas also supplied me with estimates of the number of households in each zip code so I am able to determine the shares of each directory's population affected by overlap. All data is from 1996 except for the boundary data which is from 1997. This difference creates some discrepancies between the boundary data and the rest of the data set. After matching all four data sets together, I have 431 observations.

I assign directories to their central or most populous counties and match them to demographic data from the *USA Counties CD ROM 1996*. Industry sources suggest that educated, relatively wealthy people who own their own home are likely to use the Yellow Pages, as are people who have recently moved. People who live in urban settings or regularly use public transportation use Yellow Pages less. Table 4 in the Appendix presents a description of demographic data which I use to capture these features.

Table 1 presents simple statistics for all of the variables. Note that there is a strong positive correlation between usage and the quantity of advertising, which suggests a network effect. [11] This simple correlation will be born out in the structural estimation to follow.

4.3 The Prevalance Non-Linear Pricing

One potentially problematic issue for my approach is the possibility of non-linear pricing. In estimation, I choose a single price to proxy for the price schedule at each directory. If the shapes of price schedules differ across directories in significant ways, my estimates are likely to be wrong. However, it turns out that incorporating non-linear pricing does not substantially change the results. I take the following approach to measure the degree of price non-linearity:

Directories vary in their offering of such options as "red and green whole page" and "half page with white knockout." Differences in value between these products would be difficult to capture given my data. But almost every book offers a range of sizes with no special colors. The most prevalent group of these choices are named by the number of quarter-columns that they take up and can be indexed as such; so I assign a 2 to a double quarter column advertisement, a 3 to to a triple quarter column, a 10 to a half page advertisement in a book with 5 columns and so on. Let q_{ij} be the size of option j at book i and p_{ij} be the associated price. At almost all books, I observe from 1 to 15 options, with 75% of the books having 7, 8 or 9 options. I run the log-linear regression $\ln(P_{ij}) = \ln(\tau_i^0) + \tau_i^1 \ln(q_{ij})$ at each book separately. I take τ_i^0 and τ_i^1 as observable variables which approximate the price schedule.

Table 2 presents descriptive statistics on τ_i^1, the coefficient on the size of the advertisement. A value of 1 implies linear pricing. About 65% of the observations are clustered within 0.08 of 1,

[11] There is already some evidence that consumers value Yellow Pages for the advertising. Laband (1986) and Mixon (1995) show that Yellow Pages advertisements are likely to be more informative if they are for products which are "search goods" (i.e. they are expensive and are purchased infrequently). Possibly more convincing is a recent Ameritech radio advertisement boasting that the Ameritech Yellow Pages have "the most ads and the most complete information."

Percentile	τ_i^1	Percentile	τ_i^1
100%	1.30	25%	0.92
99	1.25	10	0.78
95	1.10	5	0.73
90	1.07	1	0.57
75	1.04	0	0.40
50	0.96		
mean	0.96	std dev.	0.12
462 Observations			

Table 2: The Distribution of Price-schedule Curvatures

indicating that many directories are practicing something close to linear pricing. [12] These results suggest that assuming linear-pricing does not substantially impact the results. The Appendix discusses estimation of a model which captures non-linear pricing. The results are similar to those of the linear pricing model, presented in Section 6.

5 Estimation

This section presents the methods that I use for estimation as well as the assumptions on functional forms that I make on the model presented in Section 3. It also presents an alternative model which incorporates non-linear pricing. To review, my system consists of three equations: demand for usage by consumers ($U = U(A)$), inverse demand for advertising by retailers ($P = P(A, U(A))$), and a publisher first-order condition.

5.1 Consumer Usage

I follow methods presented in Berry (1994) for estimating parameters in markets characterized by discrete choice and oligopoly. As stated in Section 3, I assume there is some total number of times M that a household requires information of the kind that can be found in the Yellow

[12] Note that the log-linear representation of the price schedule is not a perfect one. Some directories display per-unit prices that are non-monotonic.

Pages. Consumers can choose to use a directory or use an unspecified outside option. I consider two models of the consumers' choice. The first one is a simple logit model, which assumes that the willingness of consumers to substitute between choices depends only on the size of the market share of each choice. This assumption is unlikely to be accurate. Clearly, if a consumer chooses to use a directory and that directory is removed from the consumer's choice set, the consumer is more likely to switch to another directory than to the outside option even if those two have the same market share. I also consider a nested logit model, where all directories appear in one nest and the outside option gets a nest of its own.

Estimation in the logit case is straightforward. Following Berry (1994), let the utility to consumer j from directory i be:

$$u_{ij} = \alpha_2 \ln(A_i) + X_i \beta_u + \xi_i + \epsilon_{ij}.$$

The variable A_i is the quantity of advertising at directory i. The vector X_i contains demographic variables associated with the central county of each directory. The parameter α_2 and β_u are to be estimated. The variable ξ_i is a directory specific variable which captures directory characteristics that are unobservable to the researcher. The unobservable variable ϵ_{ij} captures individual preference for a specific directory. The variable ϵ_{ij} captures issues such as the location of the individual relative to the boundaries of the directories. I assume that ϵ_{ij} is distributed Type I Extreme Value. Note that the standard logit assumption is that each agent in the population chooses only once. As the number of references a household makes per month ranges as high as 25, this assumption clearly does not hold in this case. However, as long as the number of times an individual needs information is uncorrelated with their choice of book, this issue is not a problem. I assume every household needs information the same number of times.

Let $\delta_i = \alpha_2 \ln(A_i) + X_i \beta_u + \xi_i$ be the mean level of utility derived from book i. Normalize the utility from the outside good such that $\delta_0 = 0$. Let s_i be the market share to directory i in its distribution region, so $U_i = M s_i$. Let s_0 be the share to the outside option in i's region (the i is suppressed as it will be obvious in context). Both s_i and s_0 are observable. In the simple

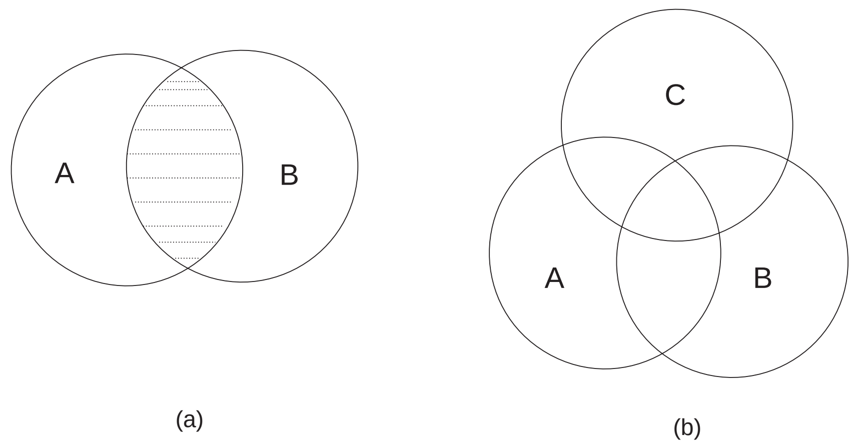


Figure 3: Overlapping markets

logit case where all products are available to all consumers, we have:

$$s_i = e^{\delta_i} s_0.$$

A potential problem is that in this industry, not all products are available to all consumers. One group of consumers in a given directory's distribution area may have two directories to choose from while another group may have only one. In figure 3(a), Directory A's market area overlaps with Directory B's. The people who live in the shaded area have two directories and everyone else has only one directory.

While this issue of overlapping markets with distinct boundaries causes serious difficulties in estimating a nested logit model, it does not affect the estimation of a standard logit model. Because of the assumption of the independence of irrelevant alternatives, we know that s_i/s_0 should be the same whether or not consumers have access to another good. Similarly, it is easy to show that s_i/s_0 is the same even if some portion of the consumers have access to another good, regardless of the size of the portion. Therefore, the relation $s_i = e^{\delta_i} s_0$ holds under the case of overlapping markets with distinct boundaries. [13] Estimation of the simple logit model for this situation is a straightforward application of techniques presented in Berry (1994). The

[13] I use the term "overlapping markets with distinct boundaries" to distinguish the problem from more standard

equation to be estimated is:

$$\ln(s_i) - \ln(s_0) = \alpha_2 \ln(A_i) + X_i\beta_u + \xi_i.$$

I assume that the unobservable term ξ_i is mean-independent of a matrix of instruments Z_i. Formally, I assume that $\mathrm{E}(\xi_i|Z_i) = 0$. I discuss elements of Z_i in Section 5.4.

The estimation problem is no longer so easy when the researcher allows for more complicated substitution patterns. I would like to consider a nested logit model where directories appear in one nest and the outside option appears in a nest of its own. Berry extends his technique to the case of the nested logit. In this case, define the utility to individual j from directory i to be:

$$u_{ij} = \alpha_2 \ln(A_i) + X_i\beta_u + \xi_i + \zeta_j + (1-\sigma)\epsilon_{ij}.$$

The variable ζ_j captures an individual's preference for Yellow Pages and ϵ_{ij} captures individual preference for a specific directory. The variable ζ_j is common to all Yellow Pages for each individual and is unobservable. I assume that ϵ_{ij} is distributed Type I Extreme Value and ζ_j is distributed such that $(1-\sigma)\epsilon_{ij} + \zeta_j$ is also distributed Type I Extreme Value. Berry (1994) discusses this issue and Cardell (1997) shows that ζ_j exists and is unique. The parameter σ is restricted to lie between zero and one, and captures the correlation between the amount of utility that consumers receive from different Yellow Pages directories. As σ approaches zero, correlation within the group goes to zero and the model approaches a standard logit model. The parameter σ will be estimated.

Let $s_{i|YP}$ be the share of people who choose directory i given that they choose to use a Yellow Pages directory. Berry (1994) shows that under the nested logit assumptions we have $s_i = e^{\delta_i} s_0 s_{i|YP}^{\sigma}$. This relation allows for the identification of α_2, β_u and σ in a log-linear model. If this relation held true at every observation, I could estimate α_2, β_u and σ by:

problems of market areas. For instance, two restaurants that are four miles apart may have overlapping customer bases but they do not have distinct boundaries. All consumers could choose to use either restaurant. The problem in this paper is one in which different groups of consumers have different choice sets. Consumers cannot use a directory unless they live in the directory's distribution area.

$$\ln(s_i) - \ln(s_0) = \alpha_2 \ln(A_i) + X_i\beta_u + \sigma \ln(s_{i|YP}) + \xi_i. \tag{4}$$

However, this function does not hold at most of the observations in my data set. Under the case of overlapping markets with distinct boundaries, this function is wrong because different consumers in a given directory's area have different choices in the Yellow Pages nest. Instead, the relation will be true in each "sub-market." I define a sub-market to be a region with a uniform set of directory choices. In Figure 3(a), there are two sub-markets in Directory A's market. In Figure 3(b), there are four sub-markets in Directory A's market, because the region where Directory C overlaps with A and B is counted as separate from where it overlaps with only A. In my data, 431 directories generate 664 sub-markets, with some sub-markets being served by as many as 8 directories.

In order to write down the correct relationship between s_i and δ_i, let $K(i)$ be the set of sub-markets in the region covered by directory i, let ψ_k be the number of households in sub-market k, let $\tilde{\psi}_i$ be the total number of households in directory i's market, let s_{0k} be the share of references to the outside option in sub-market k, and let $s_{i|YPk}$ be the share of Yellow Pages references in sub-market k which go to directory i. In this case, the share of usage going to directory i in its full market area can be written as:

$$s_i = e^{\delta_i} \sum_{k \in K(i)} \frac{\psi_k}{\tilde{\psi}_i} s_{0k} s_{i|YPk}^{\sigma}. \tag{5}$$

I do not observe $s_{i|YPk}$ or s_{0k}. However, the assumptions on functional form that are a part of the nested logit model imply:

$$s_{i|YPk} = \frac{e^{\delta_i/(1-\sigma)}}{\sum_{l \in C(k)} e^{\delta_l/(1-\sigma)}} \qquad s_{0k} = \frac{1}{1 + (\sum_{l \in C(k)} e^{\delta_l/1-\sigma})^{\sigma}}. \tag{6}$$

$C(k)$ is the set of directories which compete in sub-market k. These formulas are derived in Cardell (1997) and presented in Berry (1994). When the equations in (6) are plugged into Equation 5, the share s_i is a function of the mean utility of i and the mean utility of each directory j with which i competes. I cannot solve for the vector of mean utilities δ explicitly in this case. I use a fixed-point algorithm to solve this problem. Let s be the vector of observed market shares and let $s(\delta)$ by the vector of predicted market shares defined by Equation 5 (plugging in from (6) for s_{0k} and $s_{i|YPk}$). Define the function $g(.)$ by:

$$g(\delta) = \delta + s - s(\delta) \tag{7}$$

The Appendix shows that Equation 7 is a contraction mapping when σ is below a certain cut-off, ensuring that a unique δ exists and that it can be found. The Appendix derives the cut-off, which is always between 0.75 and 1. Following suggestions by Rust (1987), I nest a fixed point algorithm into my estimation routine which solves for δ at each evaluation of the estimation criterion function. I check that σ is less than the cut-off before each implementation of the fixed point algorithm.

For a given σ, I can identify δ. With δ in hand, I can identify α_2 and β_u by linear means because $\delta_i = \alpha_2 \ln(A_i) + X_i\beta_u + \xi_i$. In order to identify σ, I use the fact that $s_{i|YP} = s_{i|YPk}$ and $s_0 = s_{0k}$ at markets with only one sub-market. Thus, I can combine information from Equation 4 and Equation 5. I use the following approach in order to construct ξ: Start with a given σ, α_2 and β_u. Use σ and Equations 5, 6 and 7 to obtain δ via the fixed point algorithm. Let $\kappa_i = 1$ if directory i has one sub-market and at least one competitor, and $\kappa_i = 0$ otherwise. (Note that markets with no competitors confer no information about σ). Define ξ_i by:

$$\xi_i = \begin{cases} \ln(s_i) - \ln(s_0) - \sigma \ln(s_{i|YP}) - \alpha_2 \ln(A_i) - X_i\beta_u & \text{if } \kappa_i = 1 \\ \delta_i - \alpha_2 \ln(A_i) - X_i\beta_u & \text{if } \kappa_i = 0 \end{cases}$$

Using this method, I estimate α_2, β_u and σ using the Generalized Method of Moments, to be discussed in Section 5.4. The directories for which $\kappa_i = 1$ identify σ. In my data, 121 directories out of 431 have a uniform set of directories offered across their entire market. Of those, 65 do not overlap with any other directories and 56 are completely overlapped by all competitors.

5.2 Advertiser Demand

I adapt Equation 2 presented in Section 3.1. Let $\gamma = \gamma_1 + \gamma_2 - 1$. Let $e^{T_i\beta_p} = \gamma_1\pi_i$. In practice, I estimate parameters of the inverse demand function:

$$P_i = A_i^{\gamma} U_i^{\alpha_1} e^{T_i\beta_p + \nu_i}. \tag{8}$$

The function is linear in logs. I assume that $\mathsf{E}(\nu_i|Z_i) = 0$. I discuss the instrument matrix Z in Section 5.4. I expect $\gamma < 0$, capturing both decreasing returns to the size of advertisements and the business stealing aspect of more advertisements. I expect $\alpha_1 > 0$ which along with $\alpha_2 > 0$ represents the network effect. The matrix T contains directory characteristics such as the size of the population covered by a directory and a dummy for whether or not a directory is published by a telephone company.

A benefit of this estimation procedure is that it will separately identify γ, α_1 and α_2. In this way, I can distinguish between the standard scarcity effect of an increase in quantity (which pushes price down - represented by $\gamma < 0$) and the network effect (which pushes price up - represented by $\alpha_1 > 0$ and $\alpha_2 > 0$).

5.3 Supply

The publisher producing directory i picks A_i to maximize profits across the directories that it owns. Let $C_i(A_i)$ be the cost of producing a directory with advertising level of A_i. [14] Let $J(i)$ be the set of other directories owned by the publisher which produces directory i. Therefore, the publisher picks A_i to maximize $P_i(A_i, U_i(A))A_i - C_i(A_i) + \sum_{j \in J(i)} P_j(A_j, U_j(A))A_j$, where A is the vector of advertising levels. The first order condition is:

[14] Industry sources claim that the largest cost in producing a Yellow Pages directory is the cost of selling advertisements, which suggests that modeling the cost as a function of the number of advertisements is a good choice.

$$P_i + A_i\frac{\partial P_i}{\partial A_i} + A_i\frac{\partial P_i}{\partial U_i}\frac{\partial U_i}{\partial A_i} + \sum_{j\in J(i)} A_j\frac{\partial P_j}{\partial U_j}\frac{\partial U_j}{\partial A_i} - MC_i = 0. \tag{9}$$

The term MC_i is the marginal cost of an advertisement and is assumed to be constant over the quantity of advertising. I assume that $\ln(MC_i) = W_i\beta_c + \omega_i$, where W is a matrix of cost shifters, β_c is to be estimated and ω_i affects the marginal cost of directory i but is unobservable to the researcher. I assume $E(\omega_i|Z_i) = 0$ and discuss the construction of Z_i in the next subsection.

5.4 Estimation Issues

Because of non-linear cross equation restrictions on parameters, I estimate this system simultaneously using the Generalized Method of Moments (Davidson and Mackinnon 1993, Hansen 1982). The problem involves three endogenous variables, A, P and U. I use an $(n \times l)$ matrix Z of instruments for which I assume E($[\xi\ \nu\ \omega]$|Z) $= 0$. The matrix Z includes all exogenous variables i.e. directory characteristics such as the population coverage of the book and a variable indicating whether or not a book is published by a phone company, as well as consumer demographics. Descriptions of the variables can be found in Table 4 in the Appendix.

The estimation problem is to choose parameters $[\alpha\ \beta\ \gamma\ \sigma]$ to minimize the criterion $[m'Wm]$, where W is a positive definite weighting matrix and:

$$m = \begin{bmatrix} Z_u'\hat{\xi} \\ Z_p'\hat{\nu} \\ Z_c'\hat{\omega} \end{bmatrix}$$

The $(n \times 1)$ matrices $\hat{\xi}$, $\hat{\nu}$ and $\hat{\omega}$ are estimates of ξ, ν and ω based on estimates of α, β, γ and σ. Hansen (1982) shows that this estimator is consistent for any positive definite W, and that the estimator is efficient if W is chosen to be the inverse of the correlation matrix of the vector m.

I use the rate for a double quarter column advertisement as a price. This rate is the most closely watched rate in the Yellow Pages industry and is available at practically all directories in my sample. When I used other rates, results did not change (note the high correlation between rates in Table 1). For the quantity of advertisements, I used the number of pages multiplied by the number of columns. I multiplied this number by 0.8 if the book was observably smaller than a standard directory.

5.5 Identification

Computationally, parameters on endogenous variables (γ, α_1, α_2) are identified by variables which appear in the corresponding Z_u or Z_p matrix but do not appear in the equation which defines the parameter. The variables used to instrument for each equation appear in the Appendix. These "exclusion restrictions" have intuitive explanations behind them. In order to identify the affect of advertising on usage, I use the size of the directory area as measured by the number of people in it. This number (population coverage) should be correlated with the quantity of advertising in a book but not with unexplained per-household usage (interpreted as region-specific usage effects or unobserved quality of the directory). Similarly, the number of people who recently moved identifies the effect of usage on advertising. Measures of how many people recently moved should be correlated with usage levels but not with unexplained demand for advertising. Local earnings levels identify the effect of the quantity of advertising on the price of advertising, as movement in local earnings levels should move the marginal cost curve but not the demand for advertising. [15] Also, industry sources indicate that almost the entire industry contracts its printing to a single firm, R.R. Donnelly. However, Bell South and GTE maintain their own printing facilities. I include dummy variables in the Z_p matrix for being one of these companies. There are no endogenous variables introduced by the cost equation so there are no instruments applied to that equation (i.e. $Z_c = W$).

[15] Note that per capita income is a control variable.

The identification of σ, which measures the correlation of utility within the Yellow Pages group (or equivalently, the differentiation of the Yellow Pages group from the outside option) comes from comparing the market share of a directory to the market share that directory gets within the Yellow Pages group. Consider the following example of the kind of information that would identify σ: Suppose we observe a market with one directory where the share to the outside option and the share to the directory was $(\frac{1}{2},\frac{1}{2})$, so the mean utility of the directory is 0. Now consider another market with directories that were identical to the one in the first market. Suppose the market shares of the outside option and the two directories iss $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$. The second directory took just as much share from the outside option as it did from the first directory. This observation suggests very little differentiation of the Yellow Pages from the outside option or $\sigma = 0$ (the standard logit model). Now suppose the observed market shares are $(\frac{1}{2},\frac{1}{4},\frac{1}{4})$. The share to the Yellow Pages group did not increase at all with the addition of a new directory suggesting high differentiation of Yellow Pages from the outside good, or $\sigma = 1$.

There is a potential problem of endogeneity in identifying σ. If there are important unobserved variables which determine usage, then there might be correlation between the number of firms in an area and unexplained usage at each firm. For instance, suppose we observe the two markets $(\frac{1}{2},\frac{1}{2})$ and $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$. If usage was high in the second market and attracted an entrant, we could see this outcome even if $\sigma = 1$. Ignoring this issue leads us to underestimate σ. This problem enters the model via equation 4: $\ln(s_i/s_0) = \delta_i + \sigma \ln(s_{i|YP}) + \xi_i$. If ξ_i is high, s_i will be high and there will be more entrants so $s_{i|YP}$ will be low. The estimate of σ will be biased downwards unless $s_{i|YP}$ is properly instrumented for. For this purpose, I use the square mileage of the distribution area of a directory as an instrument for the usage equation (i.e., it is included in Z_u). An empirical fact is that larger directories have less of their region overlapped by other directories and so should have a higher within-group share. The simple correlation between the number of directories in an area (weighted by population) and the size of the area is -0.09. Table 8 presents OLS regressions of each endogenous variable on their corresponding instruments. In the regression of $s_{i|YP}$, the coefficient on the variable *distribution area* is positive and significant. Including the variable in the instrument vector for the usage equation has a strong effect on the estimate of σ, raising it from 0.43 to 0.83.

6 Results

Coefficient estimates are presented in Table 6 for the case where households could potentially refer to a directory 75 times in a month ($M = 75$). [16] The maximum in any market in my data is 25 uses per household per month. The two parameters of most interest are the two that make up the network effect: the effect of advertising on usage and the effect of usage on advertising. In both the logit and nested logit results, both of the coefficients are positive and statistically significant. The third parameter of primary interest is the scarcity effect, the effect of the quantity of advertising on the price of advertising. In both models, this coefficient is estimated to be negative and significant. Most of the other coefficients are significant and of the expected sign. [17] Note the large effect that being a directory associated with a telephone company has on the demand for usage. There does not seem to be a similar effect in the demand for advertising.

I also estimated a model designed to test the accuracy of Assumptions *A1* and *A2*. Those assumptions drove the result that the price at one directory does not depend directly on the amount of advertising at another directory. If one of those assumptions were very wrong, we would expect the amount of advertising at competing directories to be an important excluded variable from the price equation. I construct an index of the amount of competition that a directory faces by averaging the total number of pages of advertising at other directories across sub-markets. I re-estimated the nested logit model with the natural log of this competition

[16] I tried estimating with lower values of M. The main results did not change substantially although the estimate of σ was lower.

[17] The coefficients on education are negative in usage, contrary to the claims industry sources. This result held up across different specifications and in simple statistics as well.

index in the price equation. Because I expect the amount of advertising to be endogenous to the system, I also included an extra variable in the instrument vector for the price equation: the percentage of people in a directory's area served by competing directories. As I argued earlier, I take the geographic scope of directories to be exogenous to the price- and quantity-setting process. The results support the original specification. The coefficient on the competition index was estimated to be 0.04 with a standard deviation of 0.06. The other coefficients showed very little change. This test fails to reject Assumptions *A1* and *A2*.

As another check, I compute the profits of directories published by telephone companies to be 46% of revenue. The number is very close to the industry quote of 35-45%.

For the purpose of computing summary statistics, I use the estimates from the nested logit case. With these parameter estimates, I calculate classical and network deadweight loss as discussed in Section 3.3. [18] Figure 4 draws the estimated demand curve for a single directory. The solid line ($P_i(A_i, Ue)$) is the willingness-to-pay curve for the equilibrium choice of advertising Ae. A social planner who did not take advantage of the network effect would choose advertising level Ao. The actual optimal choice is $A*$, where price ($P_i(A_i, U_i(A_i))$ - the dotted line) is below marginal cost. Choosing $A*$ places the market on the optimal willingness-to-pay curve $P_i(A_i, U*)$ (the dashed-and-dotted line). The space between $P_i(A_i, Ue)$ and $P_i(A_i, U*)$ represents network deadweight loss. Classical deadweight loss is the triangle between $P_i(A_i, Ue)$ and marginal cost MC, to the right of Ae. The results for an average market are computed in Table 3. For the nested logit parameters, the model predicts that the quantity of advertising will be 3,117 double-quarter columns. A social planner that did not take account of the network effect would choose 12,522 double-quarter column advertisements. A social planner who did

[18] Note that I estimated demand by comparing the price of a double-quarter column advertisement with the number of pages times the number of columns per page. In order to generate the surplus numbers in dollars, I must multiply my results by the number of double-quarter columns worth of display advertisements per column. I chose a medium-sized directory and calculated the number by random sampling from the pages and columns. The multiplier is 1.22.

Network Deadweight Loss		
Summary Variable	Result	Std. Err.
Equilibrium Advt (DQC's)	3,117	1,255
Classical Social Optimum	12,522	6,782
Social Optimum	17,881	9,903
Equilibrium Surplus ($000's)	18,589	16,472
Class. Soc. Opt. Surplus	22,943	19,672
Soc. Opt. Surplus	26,238	22,369
Classical Deadweight Loss	4,356	3,333
Network Deadweight Loss	3,294	3,531
Ratio of NDWL to CDWL	0.756	0.702
Ratio of Total DWL to Equ Surp.	0.412	0.088

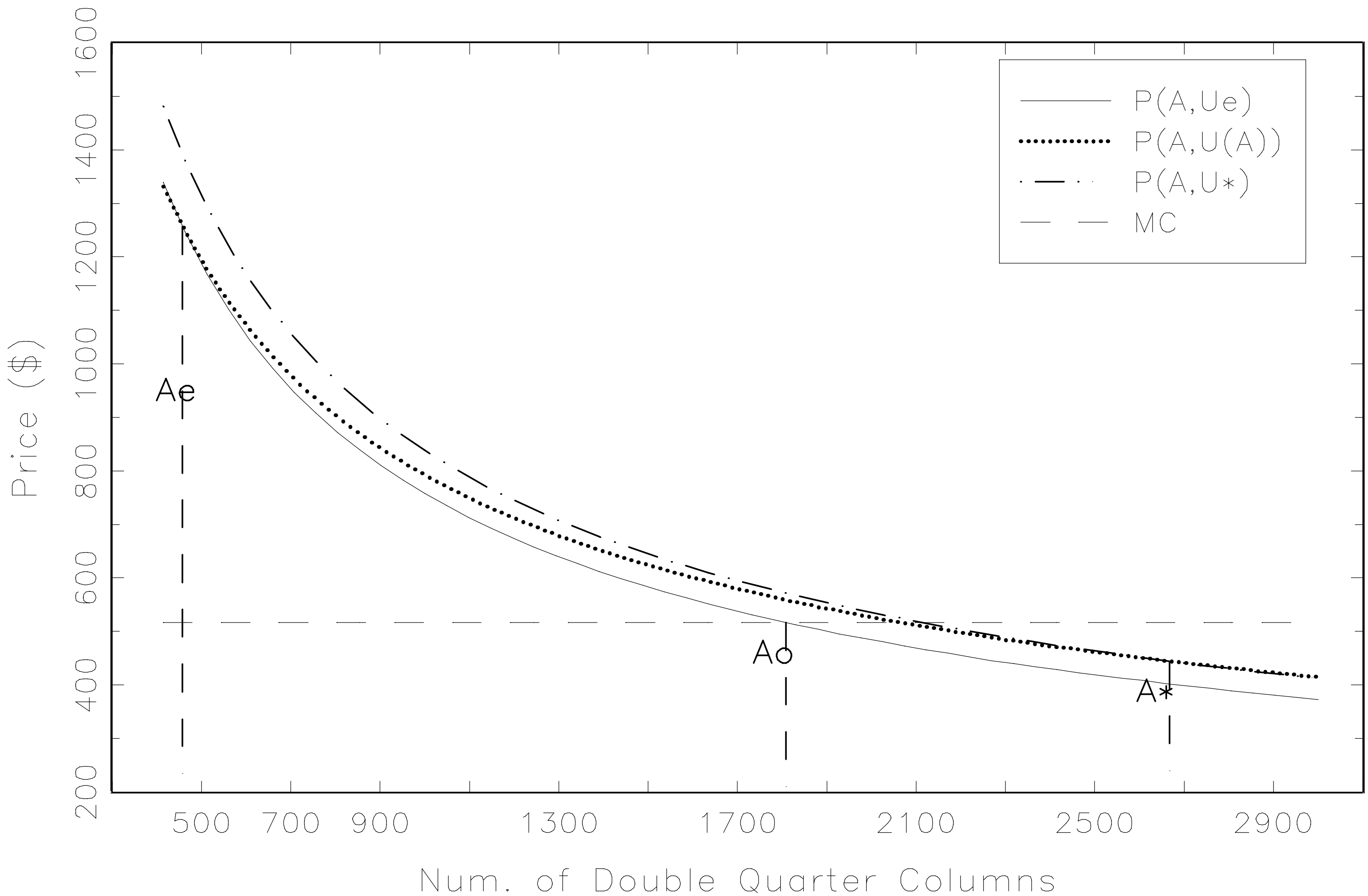
Table 3: Network Deadweight Loss

account for the network effect would choose 17,882 double-quarter column advertisements, which would imply a price below marginal cost. Network deadweight loss is $3,293,521 and classical deadweight loss is another $4,355,511. Therefore, the ratio of network deadweight loss to classical deadweight loss is 0.76 (note the large standard error: 0.70). Total deadweight loss equals 0.41 of equilibrium consumer surplus (with a standard error of only 0.09). In a market where producers clearly exhibit strong market power, network effects still create a large amount of deadweight loss. [19]

I can also calculate the effect of entry. The results from the entry experiment appear in Table 7. This table presents equilibrium outcomes for different numbers of competitors which perfectly overlap in an average market. Note that the independents are identical so the outcomes for each independent are the same for any given number of competitors. As each independent publisher enters, advertising at each independent shrinks so the benefits of the network effect are dissipated. However, total advertising and usage increase, reflecting the benefits of competition. The competitive effect is particularly striking in the incumbents' response. As the first few independent directories enter, the telephone company actually increases its quantity of

[19] The estimates of surplus and deadweight loss might be high for reasons outlined in Section 6.1. For this reason, I focus more on the ratios of one summary term to another, rather than the levels.

Figure 4
Deadweight Loss at Estimated Parameters



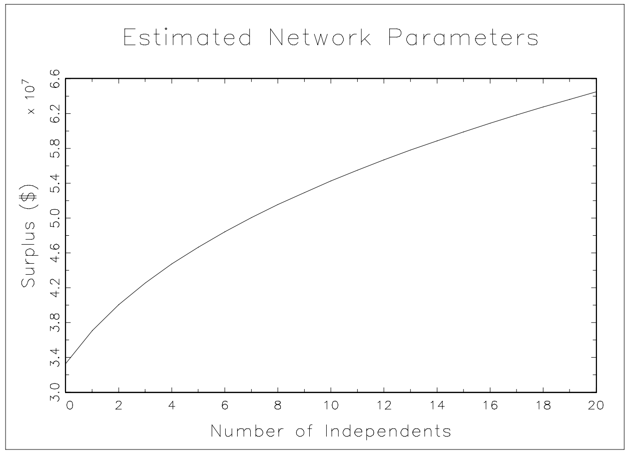
advertising, although price and usage at the telephone company continue to fall.

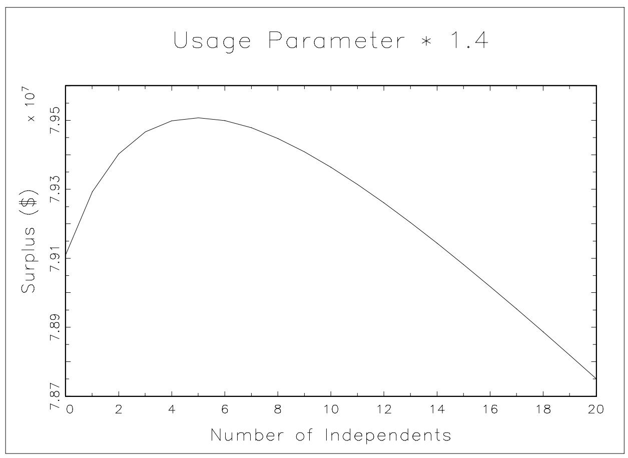
Table 7 also presents the amount of total surplus generated for each number of competitors. The results show that, ignoring fixed costs, welfare improves in the number of competitors. In this market, network effects are not strong enough to imply that the benefits of monopolization outweigh the benefits of competition. Note that the results of the model could have been different if the network effect had been estimated to be stronger. I tried raising the network parameter in the advertising equation (α_2, the coefficient on usage) and recalculating how surplus changed over the number of competitors. Figure 5 maps these results when the parameter is multiplied by 1.4 and 1.6. The figures show that for large network effects, the model implies that welfare decreases in the number of competitors or could even be hump-shaped. These results do not take account of any fixed costs.

For the actual parameter estimates, surplus increases in the number of competitors when there are no fixed costs to producing a directory. The crucial question for policy purposes is: how does the increase in surplus due to an entrant compare to the profits of the entrant? Note how poorly the entrant does. Table 7 shows that with one independent publisher, the independent collects only about 1 reference per household per month as compared to about 9 for the telephone company. The independent publisher sells about a sixth of the advertising of the telephone company for a lower price. Surplus from entrants is considerably higher than their profits, implying that there will be under-entry in equilibrium. This result suggests that current U.S. law which allows entry in the Yellow Pages market should be encouraged.

6.1 Equilibrium entry and the optimal number of directories

If we knew the fixed cost of producing a directory, we could calculate the number of entrants in equilibrium as well as the optimal number of directories for a market. While obtaining data on the fixed costs of setting up a Yellow Pages directory is difficult, I can obtain an estimate by imposing a zero-profit condition on independent directories. The estimate of the fixed cost is





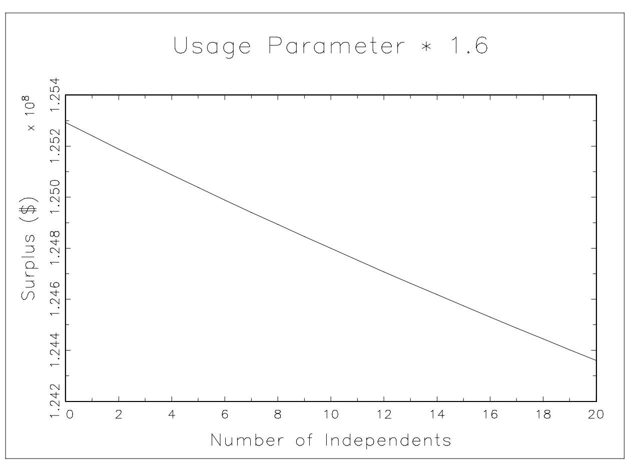


Figure 5: Surplus versus the Number of Competitors.

$1,001,022 (with a standard error of $717,198). Comparing this number of to the profits of independents in Table 7 shows that there would be no entrants in an average-sized market. This prediction seems largely accurate, as we observe most entry in bigger markets.

A fixed cost of this magnitude implies that the optimal number of directories is 16. Note that I never observe a sub-market with more than 8 directories in it. I take this result to mean that a more competitive market is strongly preferable to the current market structure. Remember that I do not take surplus on the consumer side of the market into account. At the estimated parameters, consumer utility would increase in the number of competitors because total usage does. Measuring consumer utility would mean that the benefits to entry would be even higher.

There are several explanations for the high optimal number of directories. One is that I am making predictions far out of sample and, despite the structural nature of the model, it is bound to have limitations. [20] A second explanation is that I have a poor measure of fixed costs. My estimate is a per-year per-directory fixed cost. True fixed costs might be substantially more complicated and larger.

Other explanations suggest limitations to the present model. The Cobb-Douglas framework implies that some advertisers are receiving arbitrarily high profits from purchase. I tried recalculating welfare but with willingness-to-pay capped at the price implied by advertising set at 5% of the equilibrium level. This technique removes the upper tip of the demand curve (surplus is reduced by $\int_0^{0.05A} P(A)dA - P(0.05A)0.05A$). Doing so reduces the optimal number of directories from 16 all the way to 9, suggesting that the tip of the demand curve (where there is very little data) is a major contributor to the high optimal number of directories. However, one would have to place the cut-off at a much higher level in order to actually reverse the result

[20] For instance, the logit set-up implies that each new directory is equally differentiated from all previous ones so that product space may not "fill up". If we could observe markets with 16 directories, we could estimate the level of differentiation (σ) which is relevant for that region. A model in which σ varied with the number of directories might address this issue, but I have been unable to estimate such a model in my data.

that more competitors are better. Even when the cut-off is 25% of equilibrium advertising, the optimal number of directories is 4 and the surplus caused by entry is still much higher than profits for entrants.

An area where the model might be improved lies in the interpretation of the error term. I interpret the Extreme Value error term in the consumers' utility function as capturing the relationship between the coverage area of the directory and the location of the consumer. However, in the final policy experiment, I restrict directories to be covering the same area. The estimate of correlation of the unobserved portion of utility from Yellow Pages directories (σ) comes from an environment where there is more differentiation among directories then there is in the final policy experiment. One possible solution to this problem is to exploit information about the geographic distribution of directories and include an explicit model of how consumers value directories with different distribution areas.

Moving differentiation from the error term into the observable portion of the utility function will presumably increase the estimate of σ. A higher σ would imply that adding directories in the entry experiment would bring in less usage from the outside option and reduce the benefits of entry. However, a coherent model of how consumers choose between books with different distribution areas is difficult to write and estimate given my data. Another possibility along these lines would be to estimate the model on a sample of directories which closely overlap each other. Then the estimation environment would match the assumptions in the policy experiment. Unfortunately, restricting the sample in this way leaves me with very few degrees of freedom, especially for the parameter σ, and my estimation routine fails to converge.

7 Conclusion

This paper examines the welfare trade-off between competition and standardization in a market characterized by network effects. The paper presents a model of the market for Yellow Pages which explicitly captures the relationship between advertising and consumer usage in a directory.

The model recognizes that an increase in the quantity of advertising affects the price of advertising in two ways: the "scarcity effect" forces the price down for standard reasons while the "network effect" forces price up through advertising's positive effect on usage. A model which differentiates between these two effects allows the researcher to determine how much market equilibrium differs from the social optimum because of standard problems with imperfect competition and how much of the difference results from the presence of network effects.

The paper estimates the model by extending the techniques of Berry (1994) to the case of overlapping markets with distinct boundaries. The paper also discusses how to estimate a model of non-linear pricing when the researcher observes many prices for different sizes on a product but only an aggregate quantity variable. The results show that, for a given directory, retailer demand for advertising increases in consumer usage and that consumer demand for directory usage increases in the amount of advertising, implying a network effect. In equilibrium, potential surplus that is not achieved due to unexploited network effects is high relative to the amount of surplus obtained in equilibrium and also relative to the deadweight loss due to imperfect competition. However, the results show that despite the network effect, a more competitive market structure is preferable to a more concentrated one. Strikingly, the paper finds that entrants improve welfare but are unprofitable in the average market. The results of the paper imply that encouraging competition in the Yellow Pages (as a number of recent policy changes do) improves welfare.

Directory Level Variables		
Name	Description	Instrument Vector
advertising	Number of pages times number of columns times 0.8 if observably smaller, logged. From YPPA library, collected by Boston Consulting Group.	
price	rate for a double-quarter column ad, logged. From *Rate and Data CD-ROM*.	
usage	Number of uses per household per month, logged. From National Yellow Pages Monitor Area Reports.	
Pop. Coverage	Number of people covered by a directory area, logged. From YPPA Industry Characteristics CD-ROM.	U,P,C
telco	Dummy, 1 if directory is associated with phone company. Constructed by observing publisher, books and some company contact.	U,P
GTE, Bell South	Telco Dummies	P,C
Distribution Area	Square mileage of distribution area From Claritas PowerPages CD-ROM	U
County Level Variables - From USA Counties CD-ROM 1996		
urban pop.	% of population in urban setting, 1990	U,P
diff county	% of population in different county 5 years previous, 1990	U,P
diff state	% of population in different state 5 years previous, 1990	U,P
have not moved	% of population in same house as 5 years previous, 1990	U,P
owner occ. house	% of housing that is owner occupied, 1990	U,P
income	per capita income, 1993	U,P
density	population in 1995 divided by square mileage	U,P,C
growth	rate of population increase from 1990 to 1995	U,P
public trans.	% of population regularly using public transportation (1995)	U,P
earnings	per capita earnings	U,P,C

Table 4: Explanation of Variables

8 Appendix

8.1 Equilibrium in the Publisher's Game

This section shows that there exists an equilibrium in pure strategies in the publishers' game. I show that each publisher's objective function is concave in its choice variable and appeal to Theorem 1.2 in Fudenberg and Tirole (1991), originally proven by Debreu, Glicksberg and Fan. Publisher i, $i = 1, ..., n$, chooses its quantity of advertising A_i to maximize profits: $P_i(A_i, U_i(A_1, ..., A_n))A_i - C_i(A_i)$. For this sub-section, I assume $P_i = A_i^{\gamma} U_i^{\alpha_1} X_i$. X_i captures directory level characteristics. In order to define the usage equation $U(.)$, I use a logit model adjusted to capture the effects of overlapping markets with distinct boundaries. To do so, I divide the industry into sub-markets so that each sub-market has a uniform set of directories (as described in Section 5). Let $K(i)$ be the set of sub-markets that are covered by directory i and let $C(k)$ be the set of directories that serve sub-market k. Let s_{ik} be the share of directory usage in sub-market k that goes to i and let ψ_k be the share of i's market represented by sub-market k (the i will be obvious in context). I assume the logit model holds in each sub-market. Therefore:

$$s_{ik} = \frac{A_i^{\alpha_2} X_i}{1 + \sum_{j \in C(k)} A_j^{\alpha_2} X_j} \qquad s_i = \sum_{k \in K(i)} \psi_k s_{ik}$$

Usage is defined by $U_i = M s_i$, where M is the size of the market. Now consider publisher i's first-order condition:

$$P_i + A_i \left(\frac{\partial P_i}{\partial A_i} + \frac{\partial P_i}{\partial U_i} \frac{\partial U_i}{\partial A_i} \right) = MC_i.$$

I assume marginal cost is constant at MC_i, as is done in estimation. For these functional forms, the first-order condition for publisher i is:

$$P_i \left(1 + \gamma + \frac{\alpha_1 \alpha_2}{s_i} \sum_{k \in K(i)} \psi_k s_{ik}(1 - s_{ik}) \right) = MC_i.$$

The term in large parentheses represents the price-cost markup. Allow Γ_i to equal the markup, so $P_i \Gamma_i = MC_i$. Concavity requires that the second derivative is negative:

$$\left(\frac{\partial P_i}{\partial A_i} + \frac{\partial P_i}{\partial U_i} \frac{\partial U_i}{\partial A_i} \right) \Gamma_i + P_i \alpha_1 \alpha_2 \frac{\partial}{\partial A_i} \frac{\sum_{k \in K(i)} \psi_k s_{ik}(1 - s_{ik})}{s_i} < 0. \tag{10}$$

The term in the first set of parenthesis is the slope of the demand curve and is assumed to be negative. The price-cost margin Γ_i is positive for any reasonable parameter values. In the next term, P_i, α_1 and α_2 are each positive. Unfortunately, the term $\frac{\partial}{\partial A_i} \frac{\sum_{k \in K(i)} \psi_k s_{ik}(1-s_{ik})}{s_i}$ is difficult to sign. In the case of perfect overlap, that term becomes $\frac{\partial}{\partial A_i}(1 - s_i)$ which is obviously negative, so there is a solution in pure strategies. In the case without perfect overlap, the term is:

$$\frac{\partial}{\partial A_i} \frac{\sum_{k \in K(i)} \psi_k s_{ik}(1 - s_{ik})}{s_i} = -\frac{[\sum_{k \in K(i)} \psi_k s_{ik}(1 - s_{ik})]^2}{s_i} + \sum_{k \in K(i)} (1 - 2s_{ik})(1 - s_{ik}) s_{ik}.$$

The first term on the right-hand side is always negative but the second term might be positive. I must assume that α_1 and α_2 are not "too large" to guarantee an equilibrium in pure strategies. Condition 10 easily holds at the parameters estimated in this paper.

8.2 Establishing when the Usage equation is a contraction

In order to establish that Equation 5 has a unique fixed point, I show that the equation $g : \Re^n \to \Re^n$ defined as $g(\delta) = \delta + s - s(\delta)$ is a contraction by showing that the function satisfies the conditions stated in the theorem in Appendix I of Berry, Levinsohn and Pakes

(1995). The important conditions to show are that $g(\delta)$ is continuous in δ, that $\partial g_i(\delta)/\partial\delta_j \geq 0$ for all i and j, and that $\sum_{j=1}^{n} \partial g_i(\delta)/\partial\delta_j < 1$. The function is continuous by construction. I show that $\partial g_i(\delta)/\partial\delta_i \geq 0$ last as this part of the proof requires conditions on σ.

First, I establish that $\partial g_i(\delta)/\partial\delta_j \geq 0$ for $i \neq j$.

$$\frac{\partial g_i(\delta)}{\partial\delta_j} = -e^{\delta_i} \sum_{k\in K(i)} \frac{\psi_k}{\psi_i} \frac{\partial}{\partial\delta_j} s_{0k} s_{i|YPk}^{\sigma}.$$

Following from Berry (1994), it is easy to show that $\partial s_{i|YPk}/\partial\delta_j = -s_{i|YPk}s_{j|YPk}/(1-\sigma)$ and that $\partial s_{0k}/\partial\delta_i = -s_{ik}s_{0k}$. Plugging in, we have that:

$$\frac{\partial g_i(\delta)}{\partial\delta_j} = e^{\delta_i} \sum_{k\in K(i)} \frac{\psi_k}{\psi_i} s_{0k} s_{i|YPk}^{\sigma} \left(\frac{\sigma}{1-\sigma} s_{j|YPk} + s_{jk}\right) > 0.$$

In order to show that $\sum_{j=1}^{n} \partial g_i(\delta)/\partial\delta_j < 1$, note that $\partial s_{i|YPk}/\partial\delta_i = (s_{i|YPk} - s_{i|YPk}^2)/(1-\sigma)$. Therefore:

$$\frac{\partial g_i(\delta)}{\partial\delta_i} = 1 - s(\delta) + e^{\delta_i} \sum_{k\in K(i)} \frac{\psi_k}{\psi_i} s_{0k} s_{i|YPk}^{\sigma} \left(\frac{\sigma}{1-\sigma}(s_{i|YPk} - 1) + s_{ik}\right)$$

$$\sum_{i=1}^{n} \frac{\partial g_i(\delta)}{\partial\delta_j} = 1 - s(\delta) + e^{\delta_i} \sum_{k\in K(i)} \frac{\psi_k}{\psi_i} s_{0k} s_{i|YPk}^{\sigma}(1 - s_{0k}) =$$

$$1 - e^{\delta_i} \sum_{k\in K(i)} \frac{\psi_k}{\psi_i} s_{0k}^2 s_{i|YPk} < 1.$$

Now I establish conditions that insure that $\partial g_i(\delta)/\partial\delta_i > 0$. It is equivalent to show that $\partial s_i(\delta)/\partial\delta_i < 1$. A sufficient condition is that $\partial s_{ik}(\delta)/\partial\delta_i < 1$ for all sub-regions k. Suppressing the subscript k, I show that:

$$\frac{\partial s_i}{\partial\delta_i} = \frac{\partial s_{i|YP} s_{YP}}{\partial\delta_i} = s_{i|YP} \frac{\partial s_{YP}}{\partial\delta_i} + s_{YP} \frac{\partial s_{i|YP}}{\partial\delta_i} < 1.$$

Following Berry (1994), we have that $\partial s_{i|YP}/\partial\delta_i = s_{i|YP}(1 - s_{i|YP})/(1-\sigma)$. Also, we have that $s_{YP} = \exp(x)/1 + \exp(x)$ where $x = (1-\sigma)\ln\sum_j \exp(\delta_j/(1-\sigma))$. The derivative reduces to $\partial s_{YP}/\partial\delta_i = s_{YP}(1 - s_{YP})s_{i|YP}$. Plugging into $\partial s_i/\partial\delta_i$ and solving for σ shows that we have sufficient conditions for a contraction whenever:

$$\sigma < \frac{1 - s_i(1 - s_i)}{1 - s_i(s_{i|YP} - s_i)}. \tag{11}$$

It is easiest to study this condition by converting $s_i = s_{i|YP}s_{YP}$, as the latter two terms can be moved independently of each other. The upper bound on σ decreases in s_{YP} and is convex in $s_{i|YP}$, reaching a minimum at $s_{i|YP} = 0.5$. The bound is always greater than 0.75 and less than 1. There is actually some intuition to the result that σ must be less than one. We are trying to show that $\partial s_i/\partial\delta_i$ is less than one. When σ is high, all of the randomness in utility is placed at the group level, which means that within-group choices are based almost entirely on mean utility. When mean utility (δ_i) moves slightly, it generates a big response and s_i rises too quickly. Of course, if few people choose the group, σ can be higher and s_i still will rise at a reasonable rate. And as is typical in logit models, the within-group derivative is highest when the within-group market share is close to 0.5.

I check for σ to satisfy (11) before searching for a fixed point. Note that I always start my estimation procedure with a guess of σ that is below 0.75, ensuring that a fixed point exists at my starting values. If the true σ is greater than the bound, I expect my estimate of σ to rise. When the estimate of σ crosses the bound, my estimation routine stops because estimates of δ might be meaningless at that point.

8.3 Estimation which incorporates non-linear pricing

Although I have only a single observation on quantity, I observe the entire set of prices that each directory chooses and would like to take advantage of this rich data set. I take the variables τ_i^0 and τ_i^1 (derived at the end of Section 4.3) as observable variables which represent the price schedule of each firm.

In this case, the analog to the advertiser maximization problem from Section 3.1 is to let advertiser k picks its level of advertising at book i, denoted a_{ik}, to maximize $\pi_i e^{\nu_{ik}} a_{ik}^{\gamma_1} A_i^{\gamma_2} U_i^{\alpha_1} - \tau_i^0 a_{ik}^{\tau_i^1}$. Following the same steps as above, and now letting $\gamma = \gamma_1 + \gamma_2$, we can derive a new demand curve to replace Equation 8:

$$\tau_i^0 = A_i^{\gamma - \tau_i^1} U_i^{\alpha_1} e^{T_i \beta_p + \nu_i}. \tag{12}$$

Instead of picking A_i and having the market return P_i, the publisher now picks A_i and τ_i^1 and the "inverse demand curve" gives τ_i^0. I cannot derive a publisher first-order condition in observable variables for the non-linear pricing model, but I can still estimate the two demand curves and compare the results to the linear pricing case to see if the results change substantially. [21] The results do not change substantially. Note that the double-quarter column rate and τ_i^0 have a simple correlation of 0.987, and that τ_i^1 is clustered around 1, so it is not surprising that incorporating these variables does not have a large effect.

9 Bibliography

Augereau, Angelique. "The Battle for 56K: Did network effects matter?" Mimeo, Kellogg Graduate School of Management, 1998.

Berry, Steven. "Estimating Discrete Choice Models of Product Differentiation." *Rand Journal of Economics*, 1994, 25, pp. 242-262.

[21] The first-order condition involves summing a non-linear function of advertising and price over each individual advertiser.

Berry, Steven; Levinsohn, James and Pakes, Ariel. "Automobile Prices in Market Equilibrium." *Econometrica*, July 1995, 63(4), pp. 841-890.

Berry, Steven and Waldfogel, Joel. "Free Entry and Social Inefficiency in Radio Broadcasting." *Rand Journal of Economics*, 1999, 30 (3), pp. 397-420.

Cardell, N. Scott. "Variance Components Structures for the Extreme Value of Logistic Distributions with Application to Models of Heterogeneity." *Econometric Theory*, 1997, 113(2), pp. 185-213.

Chou, Chien-fu and Shy, Oz. "Network Effects without Network Externalities." *International Journal of Industrial Organization*, June 1990, 8(2), pp. 259-70.

Church, Jeffrey and Gandal, Neil. "Network Effects, Software Provision, and Standardization." *Journal of Industrial Economics*, March 1992, 40(1), pp. 85-103.

Davidson, Russell, and Mackinnon, James G. *Estimation and Inference in Econometrics*. New York: Oxford University Press, 1993.

Economides, Nicholas. "The Economics of Networks." *International Journal of Industrial Organization*, October 1996, 14(6), pp. 673-99.

Economides, Nicholas and Flyer, Frederick. "Compatibility and Market Structure for Network Goods." Working paper, Stern School of Business at New York University, 1997.

Economides, Nicholas and Himmelberg, Charles. "Critical Mass and Network Size with Application to the US FAX Market." Working paper, Stern School of Business at New York University, 1995.

Elliott, Stuart. "Advertising." *New York Times*, April 17, 1998, pp. C7.

Feist Publications, Inc., v. Rural Telephone Service Company, Inc. *Supreme Court Reporter*, 111 S.Ct. 1282(1991).

Fudenberg, Drew, and Tirole, Jean. *Game Theory*. Cambridge: MIT Press, 1991.

Gandal, Neil."Hedonic Price Indexes for Spreadsheets and an Empirical Test of Network Externalities." *Rand Journal of Economics*, Spring 1994, 25(1), pp. 161-170.

Gandal, Neil; Greenstein, Shane and Salant, David. "Adoptions and Orphans in the Early Microcomputer Market" Working paper 2/95, Foerder Institute for Economic Research, Tel Aviv University, 1997.

Gandal, Neil; Kende, Michael and Rob, Rafael."The Dynamics of Technological Adoption in

Hardware/Software Systems: The Case of Compact Disk Players." Mimeo, Tel-Aviv University. 1997.

Greenstein, Shane M. "Did An Installed Base Give an Incumbent any (Measurable) Advantage in Federal Computer Procurement?" *Rand Journal of Economics*, Spring 1993, 24(1), pp. 19-39.

Hansen, Lars. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica*, 1982, 50, pp. 1029-1054.

Katz, Michael L. and Shapiro, Carl. "Network Externalities, Competition, and Compatibility." *American Economic Review*, June 1985, 75(3), pp. 424-40.

Katz, Michael L, and Shapiro, Carl. "Systems Competition and Network Effects." *Journal of Economic Perspectives*, Spring 1994, 8(2), pp. 93-115.

Laband, David N. "Advertising as Information: An Empirical Note." *Review of Economics and Statistics*, 1986, 68, pp. 517-521.

Liebowitz, S. J., and Margolis, Stephen E. "Network Externality: An Uncommon Tragedy." *Journal of Economic Perspectives*, 1994, 8(2), pp. 133-150.

Mixon, Franklin G. "Advertising as Information: Further Evidence." *Southern Economic Journal*, 1995, 61, pp. 1213-1218.

NARUC, *Compilation of Utility Regulatory Policy 1994-1995*, National Association of Regulatory Utility Commisioners. pp. 376.

Park, Sangin. "Quantitative Analysis of Network Externalities in Competing Technologies: The VCR Case." Mimeo, SUNY at Stony Brook, 1997.

Rosse, James N. "Estimating Cost Function Parameters without Using Cost Data: Illustrated Methodology." *Econometrica* 1970, 38(2), pp. 256-275.

Rust, John. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher ." *Econometrica*, September 1987, 55(5), pp. 999-1033.

Saloner, Garth and Shepard, Andrea. "Adoption of Technologies With Network Effects: An Empirical Examination of the Adoption of Automated Teller Machines." *Rand Journal of Economics*, Autumn 1995, 26(3) pp. 479-501.

White, Evan D. and Sheehan, Michael F. "Monopoly, The Holding Company, and Asset Stripping: The Case of Yellow Pages." *Journal of Economic Issues*, March 1992, 26(1), pp. 159-182.

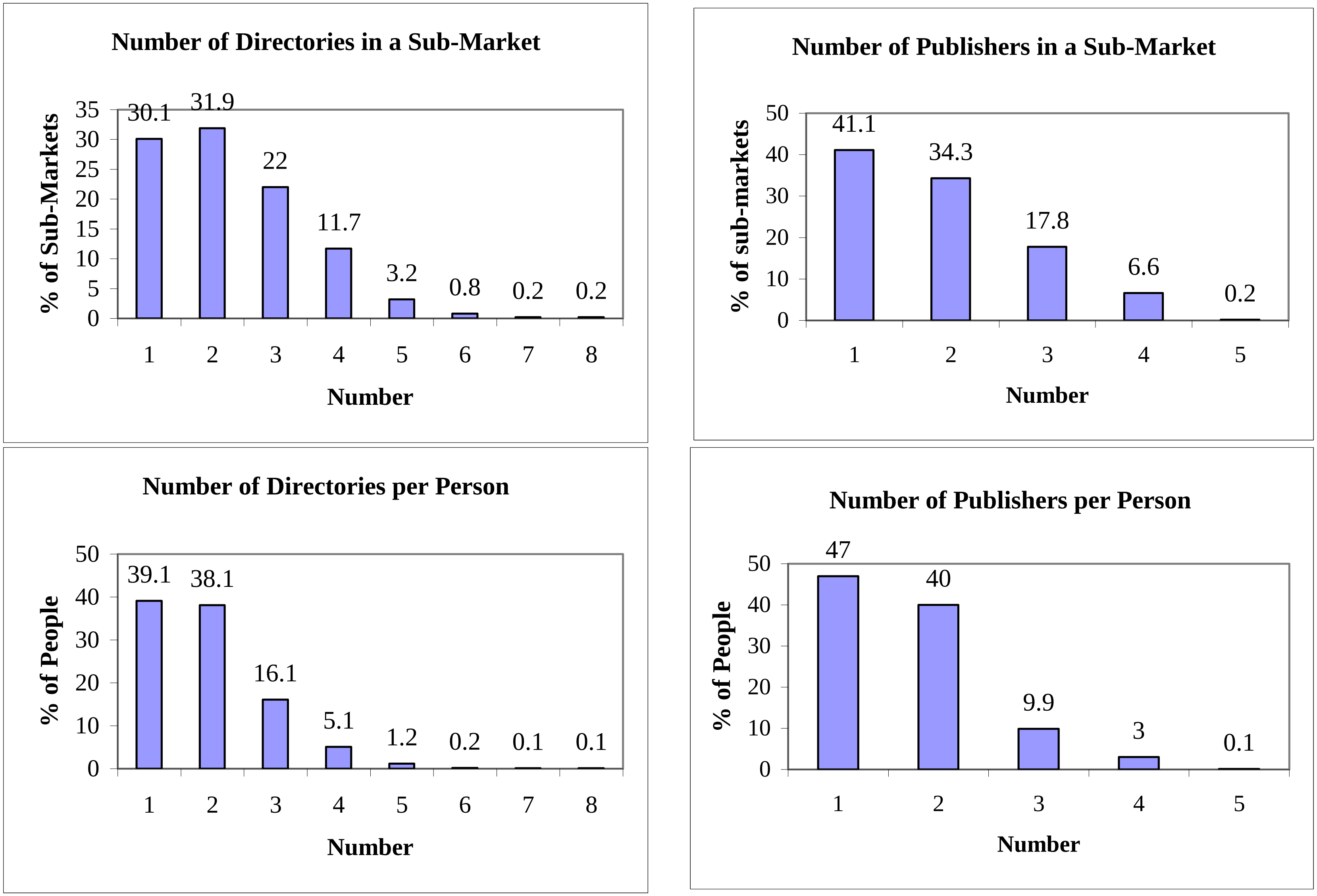

Number of Directories in a Sub-Market
% of Sub-Markets
Number
30.1
31.9
22
11.7
3.2
0.8
0.2
0.2
Number of Publishers in a Sub-Market
% of sub-markets
Number
41.1
34.3
17.8
6.6
0.2
Number of Directories per Person
% of People
Number
39.1
38.1
16.1
5.1
1.2
0.2
0.1
0.1
Number of Publishers per Person
% of People
Number
47
40
9.9
3
0.1

TABLE 5

Table 6: Results from the Generalized Method of Moments

	Nested Logit			Standard Logit	
	Mrg Efct	coef.	std error	coef.	std error
USAGE EQUATION					
advertising	0.001	0.116	0.056	0.658	0.065
constant		-2.564	0.764	-7.954	0.875
% urban population	0.042	0.005	0.002	-0.016	0.004
% lived in diff county	0.177	0.020	0.007	0.076	0.011
% lived in diff state	0.163	0.018	0.008	0.059	0.013
% have not moved	0.032	0.004	0.010	0.067	0.013
% own house	-0.015	-0.002	0.006	-0.020	0.010
% grad hi school	-0.146	-0.016	0.007	-0.034	0.011
% grad college	-0.152	-0.017	0.008	-0.029	0.013
per cap income	0.117	0.013	0.013	0.042	0.018
telco book	2.474	0.275	0.094	1.186	0.093
county pop. growth rate	0.040	0.004	0.008	0.013	0.013
% take public trans.	0.090	0.010	0.015	-0.017	0.025
pop. density	-4.89E-04	-5.43E-05	2.01E-05	-1.14E-04	3.32E-05
PRICE EQUATION					
advertising		-0.647	0.179	-0.718	0.186
usage		0.527	0.123	0.517	0.125
constant		6.384	0.671	6.458	0.683
population coverage		0.693	0.093	0.711	0.097
telco book		0.005	0.131	0.078	0.132
establishments per cap.		0.015	0.061	0.017	0.062
% urban population		-1.40E-04	0.002	1.59E-03	0.002
% grad college		0.008	0.008	0.007	0.008
% grad hi school		0.010	0.005	0.013	0.006
per cap. Income		0.015	0.014	0.014	0.014
pop. density		3.06E-05	1.41E-05	3.26E-05	1.41E-05
COST EQUATION					
constant		3.402	0.644	3.637	0.415
population coverage		0.463	0.110	0.536	0.060
earnings per worker		-0.007	0.015	0.001	0.009
pop. density		1.07E-04	3.82E-05	9.58E-06	1.77E-05
Bell South		-0.264	0.542	0.472	0.089
GTE		0.530	0.135	0.020	0.066
CORRELATION (sigma)		0.829	0.069		
p-value of exogeneity test, d.f		0.9912	6	0.9999	7
Number of Observations			431		431

Table 7
Equilibrium for Different Numbers of Competitors

# of	advertising (DQC's)		refs./HH/mth.		price ($)		profits ($)*		total	increase
independents	telco	ind.	telco	ind.	telco	ind.	telco	ind.	surplus ($)*	in surplus*
0	5,661	0	9.94	0.00	2194	0	8,148	0	33,277	
1	5,851	1,577	9.31	0.77	2076	1297	7,659	706	37,076	3,799
2	5,957	1,426	8.84	0.68	1997	1291	7,278	630	40,052	2,976
3	6,019	1,314	8.46	0.61	1939	1287	6,964	574	42,549	2,497
4	6,054	1,226	8.14	0.56	1892	1284	6,694	532	44,724	2,175
5	6,072	1,153	7.86	0.51	1854	1282	6,459	497	46,664	1,940
6	6,077	1,093	7.61	0.48	1822	1280	6,249	469	48,425	1,760
7	6,073	1,041	7.39	0.45	1795	1278	6,060	444	50,042	1,617
8	6,063	995	7.18	0.43	1771	1277	5,888	423	51,541	1,499
9	6,048	955	7.00	0.40	1749	1276	5,730	405	52,942	1,401
10	6,029	919	6.83	0.38	1730	1275	5,585	389	54,258	1,317
11	6,007	887	6.67	0.37	1713	1274	5,450	374	55,502	1,244
12	5,983	858	6.52	0.35	1697	1273	5,325	361	56,682	1,180
13	5,957	831	6.38	0.34	1683	1272	5,207	349	57,806	1,124
14	5,930	807	6.25	0.33	1670	1271	5,097	338	58,879	1,073
15	5,902	784	6.13	0.31	1657	1271	4,993	328	59,907	1,028
16	5,873	763	6.01	0.30	1646	1270	4,894	319	60,894	987
17	5,843	743	5.90	0.29	1635	1270	4,801	310	61,844	950
18	5,813	725	5.80	0.28	1626	1269	4,713	302	62,760	916
19	5,782	708	5.70	0.28	1616	1269	4,629	295	63,644	884
20	5,751	692	5.60	0.27	1608	1268	4,549	288	64,500	856

*Profits and surplus are in thousands.
*Profits and surplus are computed assuming there are no fixed costs of production.

Table 8: "First Stage" Regressions

These are OLS regressions of endogenous variables on their respective instruments.
Note that in GMM, there is no explicit "first stage".

Usage Equation Instruments

Dependent Variable:	Advertising		Within Group Share (SiYP)	
population Coverage	0.684	(0.038)	0.634	(0.069)
earnings per worker	-0.014	(0.010)	0.059	(0.016)
pop. with a competitor	-0.088	(0.083)	-0.077	(0.150)
distribution Area	-7.33E-06	(9.07E-06)	4.03E-04	(1.41E-04)
constant	2.79	(0.722)	-1.54	(1.570)
% urban population	0.003	(0.003)	0.004	(0.006)
% lived in diff county	0.043	(0.010)	0.037	(0.021)
% lived in diff state	0.037	(0.012)	-0.002	(0.025)
% own house	0.019	(0.009)	0.058	(0.017)
% grad hi school	-0.024	(0.010)	-0.102	(0.016)
% grad college	-0.016	(0.013)	-0.019	(0.027)
per cap income	0.061	(0.019)	0.037	(0.049)
telco book	0.702	(0.061)	1.729	(0.112)
county growth rate	-0.019	(0.011)	-0.039	(0.030)
% take public trans.	0.037	(0.027)	0.009	(0.050)
% have not moved	-0.012	(0.012)	-0.044	(0.030)
pop. density	-5.00E-05	(3.33E-05)	-1.88E-04	(5.15E-05)

Price Equation Instruments

Dependent Variable:	Advertising		Usage	
earnings per worker	-0.020	(0.009)	-0.020	(0.014)
Bell South	0.095	(0.142)	0.341	(0.217)
GTE	0.035	(0.085)	-0.175	(0.130)
% have not moved	0.011	(0.007)	0.040	(0.010)
% lived in diff county	0.057	(0.009)	0.087	(0.014)
% lived in diff state	0.048	(0.011)	0.066	(0.016)
constant	1.86	(0.57)	-2.50	(0.87)
% urban population	0.003	(0.00)	-0.005	(0.00)
% grad hi school	-0.017	(0.01)	-0.036	(0.01)
% grad college	-0.022	(0.01)	-0.039	(0.02)
per cap income	0.060	(0.02)	0.078	(0.03)
telco book	0.685	(0.06)	1.552	(0.09)
population coverage	0.696	(0.04)	0.341	(0.06)
establishments per cap	0.141	(0.10)	0.110	(0.15)
population density	1.41E-01	(9.97E-02)	1.10E-01	(1.52E-01)

Bold variables are "excluded".
Standard errors are in parenthesis.