Market Structure and Product Quality in the U.S. Daily Newspaper Market*

Ying Fan†

July 2, 2009

Abstract

This paper studies the effects of market structure changes on newspaper quality, subscription prices and advertising rates. It provides a framework to quantify the welfare effect of ownership consolidations taking into account endogenous quality choice. I develop a structural model that captures key features of the U.S. daily newspaper market and propose an estimation strategy that allows me to study product choice with continuous characteristics in an oligopoly market. A new data set on the U.S. newspaper market is collected to identify the demand for newspapers, demand for advertising and the cost structure of newspaper production. Two sets of counterfactual simulation exercises are conducted. The first is a case study of a merger of two newspapers in the Minneapolis market that was blocked by the Department of Justice. The simulation suggests that if it were allowed, readers' welfare would have declined by 6 dollars per household on average, and 15 dollars in the county that would have been affected most adversely. The second exercise quantifies the welfare implications of ownership consolidations in duopoly and triopoly markets and analyzes the correlation between the effect of an ownership consolidation in a market and the underlying market structure of this market. The results show that reader's welfare loss is positively correlated with taste for newspapers in general and overlapping of the newspapers in an ownership consolidation; and negatively correlated with the asymmetry of pre-merger newspaper circulation and the number of competitors.

Keywords: endogenous product choice, ownership consolidation, multiple product firms, multiple discrete choice, advertising, daily newspaper market

JEL Classification: L0, L1, L2, L8, M3

*I am indebted to my advisors Steve Berry, Hanming Fang and Philip Haile for their continual guidance, support and encouragement. I also benefited from comments by Ruediger Bachmann, Matthew Gentzkow, Justine Hastings, Alvin Klevorick, Joshua Lustig and Jesse Shapiro. All remaining errors are mine.

†Department of Economics, Yale University, 28 Hillhouse Ave., New Haven, CT 06511; ying.fan@yale.edu

1 Introduction

This paper studies how ownership consolidation affects product quality and welfare in the U.S. daily newspaper market. The last 25 years witnessed an overall decline of 55% in the number of independently owned daily newspapers. The decline occurred in almost every state and cut across all circulation sizes. What are the consequences of such a change in market structure on newspaper quality and social welfare? Specifically, upon consolidation, do newspaper publishers increase or decrease the space devoted to news? Do they enlarge or shrink the opinion-oriented section of a newspaper? Do they provide more staff-written stories or utilize more material from news agencies? How do the newspaper price and the advertising rate change? What are the implications for welfare? Standard merger analyses generally study price effects only and ignore changes in product quality. As will be shown, this is an important omission in examining newspaper markets because newspapers make substantial quality adjustment after merger.

This paper provides a framework to study empirically how market structure affects newspaper quality and prices and, in turn, social welfare. I set up a structural model to capture three key features of the U.S. daily newspaper market.[1] First, a newspaper publisher's revenue comes from both selling newspapers and selling advertising space. The demand for advertising depends on the number of readers. Therefore, product choice and the newspaper price not only directly affect circulation revenue, but also indirectly affect advertising revenue.[2] Second, a household may subscribe to more than one newspaper,[3] which requires a multiple discrete choice model on the demand side. Third, since not only prices (i.e. the newspaper subscription price and the advertising rate), but also characteristics of newspapers are chosen by publishers, I use a two-stage game, where newspaper publishers choose characteristics in the first stage and prices in the second stage.

Methodologically, incorporating these features of the newspaper market into the empirical model requires several extensions to the existing estimation techniques. On the demand side, I use a model of multiple discrete choices based on Hendel (1999)[4] with three differences: I allow for lower utility from the second choice; the model ensures that a household buys only one copy of a newspaper; and micro data is not needed for estimation. I thus show that the Berry, Levinsohn and Pakes (1995) (henceforth, BLP) methodology can be generalized to this context. On the supply side, I allow both prices and product quality to be endogenous. However, endogenizing product choice introduces new

[1] Other examples in this literature are Rosse (1967), Ferguson (1983), Genesove (1999), Gentzkow and Shapiro (2006) and George (2007), the last of which is most closely related to this paper. George also studies market structure and product differentiation in the daily newspaper industry. She regresses measures of product variety on ownership concentration and finds a positive correlation between them. Since the concept of market structure is difficult to capture by a simple index, in this paper, I take the stance of modeling it explicitly.

[2] See Rysman's (2004) study of the Yellow Pages market for a similar setup.

[3] For instance, in 84 county/years in the sample, total newspaper circulation is larger than the number of households in the county.

[4] Other examples in this literature are Nevo, Rubinfeld and McCabe (2005) and Gentzkow (2007).

computational challenges. In particular, players in the first-stage decision take into account the impact of product choice on the equilibrium price in the second stage. But computing equilibrium prices for each possible product choice is burdensome. I overcome this by using the observation that it is sufficient to know the gradient of the equilibrium price function at the data points to formulate the optimality conditions for the observed product characteristics. This gradient is obtained from the total derivative of the first order condition for prices. This approach allows me to develop a tractable estimation routine, whereas nesting an equilibrium-solving procedure in an estimation algorithm is computationally prohibitive. This estimation strategy can be used in studying product choice problems in general. Existing papers in the literature either directly specify a profit function that is not derived from demand (such as Mazzeo (2002)), or focus on monopoly industries (such as Crawford and Shum (2006)), or markets with a naturally finite and discrete product choice set (such as Draganska, Mazzeo and Seim (2007)).

To estimate the model, I collect a large set of new data on newspaper characteristics, subscription prices, advertising rates, circulation and advertising linage for all U.S. daily newspapers between 1997 and 2005.[5] Based on the estimates of the model parameters, I analyze a blocked merger that two newspapers in the Minneapolis market proposed but the Department of Justice blocked. The simulation results show that if the merger had occurred, readers' welfare in the market would have declined by 6 dollars per household on average, and by 15 dollars in the county that would have been affected the worst. This welfare loss would have resulted from by a combination of increased subscription prices (10% on average) and a reduction in the opinion section staff as well as in the number of reporters in the smaller party to the merger, by 5% and 6% respectively.

I also quantify the welfare implications of ownership consolidation in duopoly markets and triopoly markets for which I have data. The simulation results show that the median loss in readers' surplus in duopoly mergers is 16 dollars per household and in triopoly mergers[6] 5 dollars. In 94% of the markets simulated, total welfare unambiguously falls. Ignoring the newspapers' quality adjustment typically yields an underestimation of the loss in readers' welfare. The median bias is 4 dollars per household in duopoly mergers and 2 dollars in triopoly mergers. The distribution of the welfare effects across markets is also used to study the correlation between the welfare effect of ownership consolidation in a market and that market's underlying structure.

The rest of the paper is organized as follows. A structural model is presented in Section 2. Estimation equations are also derived in this section. The data is described in Section 3 and the estimation is explained in Section 4. The estimation and simulation results are in Sections 5 and 6, respectively. Section 7 concludes.

[5] Data on advertising linage is available for only a subset of newspaper/years.

[6] In a triopoly merger, the publisher of the largest newspaper is assumed to buy the second largest.

2 The Model

The profit of a newspaper comes from both selling newspapers to readers and selling advertising space to advertisers. In this section, I describe the demand for newspapers, the demand for advertising, and the supply side of the model. Estimation equations are also derived.

A Brief Road Map

The demand on readers' side is described with a multiple discrete choice model based on Hendel (1999) with three differences: I allow for decreased utility from the second choice; the model ensures that a household buys only one copy of a newspaper; and the model can be estimated with aggregate data. Demand for newspapers is derived in equation (7) below. Following Rysman (2004), I derive the demand for advertising from a representative advertiser's decision. Advertising demand depends on circulation and the advertising rate of a newspaper as shown in equation (10).

The supply side is modeled as a complete information two-stage game in which newspaper publishers choose characteristics in the first stage and prices in the second stage. The two-stage structure is used to capture in a simple way that newspaper publishers have a longer decision horizon when they make quality decisions than they do for prices.[7] The three basic elements of this game — the set of players, timing and information, and payoffs — are described in the three subsections of Section 2.3. The revenue function is determined by the two demand systems. In Section 2.3.3, I then describe the cost structure: the marginal cost of increasing circulation (equation (12)), the marginal advertising sales cost (equation (13)), and the marginal cost of increasing a particular newspaper characteristic (equation (14)). Together with the two demand functions, they give the profit function that is relevant for the second-stage price decision (equation (15)), and the profit function that is relevant for the first-stage quality decision (equation (16)).

From this, five estimation equations are derived. The first two are the model implications of the two demand systems ([S](8), [ADV](11)) and the last three are the optimality conditions with respect to advertising rates [RFOC](17), subscription prices [PFOC](19), and newspaper characteristics [XFOC](20).

The detailed structure of the model follows. Throughout the paper, a symbol with a tilde represents a function and the same symbol without tilde is the value of the function at a point.

2.1 Demand for Newspapers

The demand for newspapers is derived from the aggregation of heterogeneous households' multiple discrete choices. A multiple discrete choice model is necessary to explain duplicate readership. In the model, I set the maximum number of newspapers that a household can subscribe ($\bar{n}$) to 2.[8]

[7] Estimating a fully dynamic model is beyond the scope of this paper.

[8] The model is generalizable to $\bar{n} > 2$.

Suppose all households in a county face the same choice set and the number of daily newspapers available in county c in year t is J_{ct}. A household i in this county gets utility u_{ijct} from subscribing to newspaper j in year t and utility u_{i0ct} from an outside choice. The household first compares u_{ijct} for $j = 0, 1, ..., J_{ct}$ and chooses j^* with the highest utility. If it chooses to subscribe to some newspaper, it can then subscribe to a second. In particular, if $j^* \neq 0$, it then compares $\max\{u_{ijct} - \kappa,\ j = 1, ...j^* - 1, j^* + 1, ..., J_{ct}\}$ to u_{i0ct} and again takes the choice with the highest utility. Here, κ is a positive parameter that captures the diminishing utility from subscribing to a second newspaper. The probability that household i subscribes to newspaper j is therefore the sum of the probability that j is the first choice and that j is the second choice:

$$\Pr\left(u_{ijct} \geq \max_{h=0,...,J_{ct}} u_{ihct}\right) + \sum_{j' \neq j, 0} \Pr\left(u_{ij'ct} \geq u_{ijct} \geq \max_{h=1,...,J_{ct}, h \neq j'} u_{ihct} \ \&\ u_{ijct} - \kappa \geq u_{i0ct}\right). \quad (1)$$

I assume that a household derives utility from some characteristics of a newspaper and that this utility is also affected by county-specific factors and individual-specific tastes. To be specific, the conditional indirect utility of household i in county c from subscribing to newspaper j in year t is assumed to be

$$u_{ijct} = p_{jt}\alpha + \boldsymbol{x}_{jt}\boldsymbol{\beta}_i + \boldsymbol{y}_{jct}\boldsymbol{\psi} + \boldsymbol{D}_{ct}\boldsymbol{\varphi} + \xi_{jct} + \varepsilon_{ijt},\text{[9]} \quad (2)$$

where p_{jt} is the annual subscription price, and $\boldsymbol{x}_{jt} = (x_{1jt}, ..., x_{Kjt})$ contains the endogenous newspaper characteristics that are chosen by the newspaper publishers. They are the news hole (the space of a newspaper devoted to news), the number of staff for opinion sections, and the number of reporters. The vector $\boldsymbol{y}_{jct} = (y_{1jt}, y_{2jt}, y_{3jct})$ includes the newspaper characteristics that are assumed to be exogenous in the model because they rarely change over time. The variable y_{1jt} measures the overall size of the market of newspaper j. The market size of a newspaper affects utility because, for example, 10 reporters covering a small region can write stories of different quality and in different quantities compared to 10 reporters serving a large area.[10] y_{2jt} is an edition dummy with value 1 when newspaper j is a morning newspaper and 0 otherwise. To captures readers' taste for local newspapers, y_{3jct} is the distance between county c and newspaper j's home county, where its headquarter is located. Households are assumed to have homogenous tastes for the exogenous

[9] Utility actually varies across i, j, t. The subscript c is redundant in u_{ijct}, as each household can be in only one county. I add the subscript c to emphasize that utility is affected by some county-specific taste, which is operationalized by county-level demographics.

[10] One way to understand why the market size of a newspaper affects utility is to recognize that both x_{kjt} and x_{kjt}/y_{1jt} (for example, both reporters and reporters per household) affect the quality of a newspaper. In the empirical implementation in Section 4, I specify the quality characteristics as $\log(1 + x_{kjt})$ and the overall size of the market of newspaper j as the logarithm of the number of households in its market, i.e. $\log(y_{1jt})$. Utility then depends on $\sum_k \beta_{ki} \log(1 + x_{kjt}) + \psi_1 \log(y_{1jt})$, which is equivalent to $\sum_k \left[(\beta_{ki} - \psi_{1k}) \log(1 + x_{kjt}) + \psi_{1k} \log(\frac{1+x_{kjt}}{y_{1jt}})\right]$, where $\sum_k \psi_{1k} = \psi_1$. The latter expression means that utility depends on, for example, reporters as well as reporters per household. The former expression, which is used in the model, is more flexible as it even allows utility to depend on $(1 + x_{kjt}) / y_{1jt}^b$ for $b \neq 1$. To see this, note that $\sum_k \beta_{ki} \log(1 + x_{kjt}) + \psi_1 \log(y_{1jt})$ is also equivalent to $\sum_k \left[\left(\beta_{ki} - \frac{\psi_{1k}}{b}\right) \log(1 + x_{kjt}) + \frac{\psi_{1k}}{b} \log(\frac{1+x_{kjt}}{y_{1jt}^b})\right]$.

characteristics $\boldsymbol{y}_{jct}$. The vector $\boldsymbol{D}_{ct} = (1, D_{1ct}, ..., D_{Lct})$ includes some demographics of county c and a constant. ξ_{jct} is the unobservable (to the econometrician) county/year-specific taste for newspaper j. It captures characteristics of the newspaper that are relevant for readers but unobservable to the econometrician and therefore not included in $\boldsymbol{x}_{jt}$ or $\boldsymbol{y}_{jct}$. It also captures a county-specific taste for newspapers that is not included in $\boldsymbol{D}_{ct}$. ε_{ijt} is an i.i.d. stochastic term representing an unobservable household specific taste. Finally, $\boldsymbol{\beta}_i = (\beta_{1i}, ..., \beta_{Ki})$ where $\beta_{ki} = \beta_k + \sigma_k \varsigma_{ki}$ is household i's specific taste for the k^{th} endogenous characteristic. β_k captures the mean taste, while σ_k is the standard deviation of the marginal utilities associated with characteristic k. ς_{ki} has an identically and independently distributed standard normal distribution across characteristics and households, and $\Phi(\cdot)$ represents the distribution function of $\boldsymbol{\varsigma}_i = (\varsigma_{1i}, ..., \varsigma_{Ki})$.

Instead of treating the utility from the outside good as fixed, I model it as a time trend, which is associated with the development of online news sources and the increase in internet penetration. Specifically, let

$$u_{i0ct} = (t - t_0)\,\rho + \varepsilon_{i0t} \tag{3}$$

be the utility from the outside choice, where t_0 is the first year in the data.

This concludes the description of the individual household's multiple discrete choice model. Aggregation follows. To derive the market demand for newspapers, I define the "relative" county mean utility, δ_{jct}, as the difference between the county mean utility from subscribing to newspaper j and the county mean utility of the outside choice:

$$\delta_{jct} = p_{jt}\alpha + \boldsymbol{x}_{jt}\boldsymbol{\beta} + \boldsymbol{y}_{jct}\boldsymbol{\psi} + \boldsymbol{D}_{ct}\boldsymbol{\varphi} + \xi_{jct} - (t - t_0)\,\rho, \tag{4}$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_K)$ is a vector of the mean utilities. Let $\vartheta_{ijt} = \sum_{k=1}^{K} \sigma_k x_{kjt} \varsigma_{ki}$. Then, $u_{ijct} = [\delta_{jct} + (t - t_0)\,\rho] + \vartheta_{ijt} + \varepsilon_{ijt}$. Individual utility is now expressed as the sum of mean utility $\delta_{jct} + (t - t_0)\,\rho$ and a deviation from the mean, $\vartheta_{ijt} + \varepsilon_{ijt}$. Following the literature, I assume that ε_{ijt} is drawn from a type I extreme value distribution with location parameter 0 and scale parameter 1. Plugging (2), (3) and (4) into (1) yields the probability that a household with taste $\boldsymbol{\varsigma}_i$ chooses newspaper j:

$$\tilde{\Psi}_j^{(1)}\left(\boldsymbol{\delta}_{ct}, \boldsymbol{x}_{ct}, \boldsymbol{\varsigma}_i; \boldsymbol{\sigma}\right) + \sum_{j' \neq j, 0} \left[\tilde{\Psi}_{j,j'}^{(2)}\left(\boldsymbol{\delta}_{ct}, \boldsymbol{x}_{ct}, \boldsymbol{\varsigma}_i; \boldsymbol{\sigma}, \kappa\right) - \tilde{\Psi}_j^{(3)}\left(\boldsymbol{\delta}_{ct}, \boldsymbol{x}_{ct}, \boldsymbol{\varsigma}_i; \boldsymbol{\sigma}, \kappa\right)\right],$$

where $\boldsymbol{\varsigma}_i = (\varsigma_{1i}, ..., \varsigma_{Ki})$, $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_K)$, $\boldsymbol{\delta}_{ct} = (\delta_{jct}, j = 1, ..., J_{ct})$, $\boldsymbol{x}_{ct} = (\boldsymbol{x}_{jt}, j = 1, ..., J_{ct})$ and

$$\begin{aligned}
\tilde{\Psi}_j^{(1)}\left(\boldsymbol{\delta}_{ct}, \boldsymbol{x}_{ct}, \boldsymbol{\varsigma}_i; \boldsymbol{\sigma}\right) &= \frac{\exp\left(\delta_{jct} + \vartheta_{ijt}\right)}{1 + \sum_{h=1}^{J_{ct}} \exp\left(\delta_{hct} + \vartheta_{iht}\right)}, \\
\tilde{\Psi}_{j,j'}^{(2)}\left(\boldsymbol{\delta}_{ct}, \boldsymbol{x}_{ct}, \boldsymbol{\varsigma}_i; \boldsymbol{\sigma}, \kappa\right) &= \frac{\exp\left(\delta_{jct} + \vartheta_{ijt}\right)}{\exp(\kappa) + \sum_{h \neq j'} \exp\left(\delta_{hct} + \vartheta_{iht}\right)}, \\
\tilde{\Psi}_j^{(3)}\left(\boldsymbol{\delta}_{ct}, \boldsymbol{x}_{ct}, \boldsymbol{\varsigma}_i; \boldsymbol{\sigma}, \kappa\right) &= \frac{\exp\left(\delta_{jct} + \vartheta_{ijt}\right)}{\exp(\kappa) + \sum_{h=1}^{J_{ct}} \exp\left(\delta_{hct} + \vartheta_{iht}\right)}.
\end{aligned} \tag{5}$$

The first summand that the probability of newspaper j is chosen as the first newspaper. $\tilde{\Psi}_{j,j'}^{(2)}$ and $\tilde{\Psi}_{j}^{(3)}$ are probabilities of it being chosen as the second newspaper when household i faces a constrained choice set with newspaper j' excluded and when household i faces an unconstrained choice set, respectively. The difference of these two is the probability that newspaper j is chosen as the second newspaper when j' is the first best.

County market penetration[11] of newspaper j is the aggregation of households' newspaper choices in the county:

$$\tilde{s}_{jct}\left(\boldsymbol{\delta}_{ct}, \boldsymbol{x}_{ct}; \boldsymbol{\sigma}, \kappa\right) = \int \tilde{\Psi}_{j}^{(1)} d\Phi\left(\varsigma_i\right) + \sum_{j' \neq j, 0} \int \left(\tilde{\Psi}_{j,j'}^{(2)} - \tilde{\Psi}_{j}^{(3)}\right) d\Phi\left(\varsigma_i\right), \tag{6}$$

If there is only one newspaper in county c, $\tilde{s}_{jc}\left(\boldsymbol{\delta}_{ct}, \boldsymbol{x}_{ct}; \boldsymbol{\sigma}\right) = \int \tilde{\Psi}_{j}^{(1)} d\Phi\left(\varsigma_i\right)$ as in a single discrete choice model. Let the last element of the county demographics vector, D_{Lct}, be the number of households in county c in year t. The demand for newspaper j, i.e. the total circulation of newspaper j, is then the sum of the circulation in all counties covered by newspaper j (denoted by $\mathcal{C}_{jt}$):

$$\tilde{q}_{jt}\left(\boldsymbol{\delta}_{ct}, \boldsymbol{x}_{ct}; \boldsymbol{\sigma}, \kappa\right) = \sum_{c:\ c \in \mathcal{C}_{jt}} D_{Lct} \tilde{s}_{jct}\left(\boldsymbol{\delta}_{ct}, \boldsymbol{x}_{ct}; \boldsymbol{\sigma}, \kappa\right). \tag{7}$$

I have assumed that readers only care about news hole and do not care about advertising. This assumption is necessary to keep the model tractable. The demand for advertising depends on the total circulation, in other words, on readers' decisions. If readers' decisions were, in turn, to depend on advertisers' decisions, solving for an equilibrium of subscription price and advertising rate would become a fixed-point problem. Moreover, it is not clear how to separately identify the effects of news hole and advertising on newspaper demand. Any exogenous variation that leads to changes in the news hole also affects the advertising demand through influencing circulations. Similarly, because advertising affects circulation and thus affects publishers' optimal choice of the news hole, any exogenous change in advertising will also alter news hole.[12]

I now derive the first estimation equation that will be taken to the data. Following Berry (1994), I do not use the demand equation directly. Instead, I use the relative mean utility in equation (4) to avoid a nonlinear endogeneity problem.[13] Berry (1994) and BLP show that in a single discrete choice model, under certain regularity conditions on the density of households' unobservable tastes, there exists a unique vector of mean utility levels, $\boldsymbol{\delta}_{ct}$, such that $s_{jct} = \tilde{s}_{jct}\left(\boldsymbol{\delta}_{ct}, \boldsymbol{x}_{ct}; \boldsymbol{\sigma}\right)$.[14] Theorem

[11]This is typically called "market share" in a single discrete choice model. But in a multiple discrete choice model, the sum of "market shares" can be larger than 1. "Market penetration" is therefore a better term.

[12]Rysman (2004) allows consumers to value advertising in his study of the network effects in the Yellow Pages market. But for one thing, Yellow Pages directories are free. Publishers choose advertising rates only and do not have a two-dimensional interdependent price decision. For another, there is no analogue of news hole in Yellow Pages.

[13]Berry (1994) notices that product prices (here product quality as well) are correlated with the taste shock ξ_{jct}, which is nonlinear the market penetration function, $\tilde{s}$. This therefore leads to a nonlinear endogeneity problem.

[14]That is to say, $\boldsymbol{\delta}_{ct}$ is uniquely determined by the data for given $\boldsymbol{\sigma}$. One can therefore treat it as if it were observable. Note that the unobservable taste shock ξ_{jct} is linear in δ_{jct}. This becomes a conventional linear endogeneity problem.

1 below states that this invertibility result can be extended to the current multiple discrete choice model. Furthermore, the contraction mapping defined in BLP is still viable, leading to a simple algorithm to solve for $\boldsymbol{\delta}_{ct}$.

Theorem 1 *For any* $(\boldsymbol{s}, \boldsymbol{x}) \in R^J \times R^{KJ}, \boldsymbol{\sigma} \in R^K, \kappa \in R^+$ *and distribution functions* $P_{\boldsymbol{\varsigma}}(.; \boldsymbol{\sigma})$, *define operator* $F : R^J \to R^J$ *pointwise as* $F_j(\boldsymbol{\delta}) = \delta_j + \ln s_j - \ln \tilde{s}_j(\boldsymbol{\delta}, \boldsymbol{x}; P_{\boldsymbol{\varsigma}}, \boldsymbol{\sigma}, \kappa)$, *where*

$$\tilde{s}_j(\boldsymbol{\delta}, \boldsymbol{x}; P_{\boldsymbol{\varsigma}}, \boldsymbol{\sigma}, \kappa) = \int \tilde{\Psi}_j^{(1)} dP_{\boldsymbol{\varsigma}}(\boldsymbol{\varsigma}; \boldsymbol{\sigma}) + \sum_{j' \neq j, 0} \int \int \left(\tilde{\Psi}_{j,j'}^{(2)} - \tilde{\Psi}_j^{(3)} \right) dP_{\boldsymbol{\varsigma}}(\boldsymbol{\varsigma}; \boldsymbol{\sigma}),$$

and $\tilde{\Psi}_j^{(1)}(\boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\varsigma}; \boldsymbol{\sigma})$, $\tilde{\Psi}_{j,j'}^{(2)}(\boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\varsigma}; \boldsymbol{\sigma}, \kappa)$ *and* $\tilde{\Psi}_j^{(3)}(\boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\varsigma}; \boldsymbol{\sigma}, \kappa)$ *are defined in (5). If (1)* $0 < s_j < 1$ *for* $\forall j = 1, ..., J$ *and (2)* $\sum_{j=1}^{J} s_j < 2$, *then the operator* F *has a unique fixed point.*

The proof of the theorem can be found in Appendix A. The first assumption means that there is always some household choosing newspaper j and some household not choosing it. The second assumption means that there is always some household with fewer than two newspapers. Under these two conditions, the solution to $s_{jct} = \tilde{s}_{jct}(\boldsymbol{\delta}_{ct}, \boldsymbol{x}_{ct}; \boldsymbol{\sigma}, \kappa)$ is unique. Denote this solution by $\boldsymbol{\delta}_{ct}(\boldsymbol{s}_{ct}; \boldsymbol{\sigma}, \kappa)$. Plugging it into the expression for the relative mean utility level in equation (4) gives for the true value of $(\alpha, \boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\varphi}, \rho, \boldsymbol{\sigma}, \kappa)$:

$$\delta_{jct}(\boldsymbol{s}_{ct}; \boldsymbol{\sigma}, \kappa) = p_{jt}\alpha + \boldsymbol{x}_{jt}\boldsymbol{\beta} + \boldsymbol{y}_{jct}\boldsymbol{\psi} + \boldsymbol{D}_{ct}\boldsymbol{\varphi} - (t - t_0)\rho + \xi_{jct}, \forall jct. \qquad [S](8)$$

This is the first estimation equation. For the reader's ease, I label all estimation equations with brackets. This equation is labeled as [S](8) because it is derived from the market penetration function $\tilde{s}_{jct}$. In the remainder of this section, subscript t is suppressed for ease of exposition and only restored in the estimation equations.

2.2 Demand for Advertising

Following Rysman (2004), I assume that a representative advertiser has the following maximization problem:

$$\max_{\{a_j\}} \sum_j \left(\eta_j' q_j^{\lambda_1'} A_j^{\lambda_2'} a_j^{\lambda_3'} - r_j a_j \right), 0 < \lambda_3' < 1, \eta_j' > 0, \qquad (9)$$

where a_j is the advertising space that the advertiser purchases in newspaper j, and r_j and q_j are the advertising rate and the total circulation of newspaper j, respectively. High circulation is expected to increase advertising effectiveness. A_j is the total advertising space in newspaper j. It affects the visibility of a specific advertisement. When λ_2' is negative, there exist negative externalities in advertising. η_j' captures the demographics of newspaper j's circulation area. It also influences the effectiveness of advertising in newspaper j.

The assumed additive separability of the profit function over newspapers implies that an advertisement in one newspaper is neither a substitute nor a complement to advertisements in other newspapers in terms of generating profit. Hence, the advertiser will keep advertising in each newspaper until the marginal profit from that advertising is 0. In other words, the advertiser maximizes its profit by independently choosing the advertisement level a_j for each newspaper j.

The solution to the advertiser's problem is

$$a_j = \left(\lambda_3' \eta_j'\right)^{\frac{1}{1-\lambda_3'}} q_j^{\frac{\lambda_1'}{1-\lambda_3'}} A_j^{\frac{\lambda_2'}{1-\lambda_3'}} r_j^{\frac{1}{\lambda_3'-1}}.$$

Aggregation (setting $a_j = A_j$) yields

$$A_j = \left(\lambda_3' \eta_j'\right)^{\frac{1}{1-\lambda_2'-\lambda_3'}} q_j^{\frac{\lambda_1'}{1-\lambda_2'-\lambda_3'}} r_j^{\frac{1}{\lambda_2'+\lambda_3'-1}}.$$

This can be rewritten as follows with $\lambda_1 = \frac{\lambda_1'}{1-\lambda_2'-\lambda_3'}$, $\lambda_2 = \frac{1}{\lambda_2'+\lambda_3'-1}$ and $\eta_j = \log\left[\left(\lambda_3' \eta_j'\right)^{\frac{1}{1-\lambda_2'-\lambda_3'}}\right]$:

$$\tilde{a}\left(r_j, q_j, \eta_j; \boldsymbol{\lambda}\right) = e^{\eta_j} q_j^{\lambda_1} r_j^{\lambda_2}. \tag{10}$$

As mentioned, η_j captures the demographics of newspaper j's circulation area. Specifically, I operationalize η_j as follows. Let $\boldsymbol{D}_c \boldsymbol{\phi}$ be a linear combination of observable demographics of county c. Then η_j is defined as the circulation-weighted sum of these county indices over the counties covered by newspaper j: $\eta_j = \sum_{c:\, c \in \mathcal{C}_j} \frac{q_{jc}}{q_j} \boldsymbol{D}_c \boldsymbol{\phi}$, and $\boldsymbol{\phi}$ is a vector of parameters to be estimated.

Let ι_{jt} be an i.i.d. and mean zero measurement error for display advertising linage, then the second estimation equation is

$$\log a_{jt} = \sum_{c:\, c \in \mathcal{C}_{jt}} \frac{q_{jct}}{q_{jt}} \boldsymbol{D}_{ct} \boldsymbol{\phi} + \lambda_1 \log q_{jt} + \lambda_2 \log r_{jt} + \iota_{jt}, \forall jt. \qquad \text{[ADV](11)}$$

2.3 Supply

This section presents an oligopoly model of the U.S. daily newspaper industry. I first show in Table 1 that the U.S. daily newspaper industry is indeed dominated by oligopoly markets due to the partial overlapping of circulation areas. To measure the extent of the overlap of a newspaper pair,

Table 1: Newspaper Coverage Overlapping

criterion	25%	20%	15%	10%
number of newspaper/year pairs	6109	6273	6692	7400

I compute the percentage of circulation in the common area as a fraction of the total circulation for each member of the pair. For example, for 6109 newspaper/year pairs in the data, the overlapping percentage is above 25% for both members. Therefore, Table 1 indicates that a daily newspaper market is typically an oligopoly market.

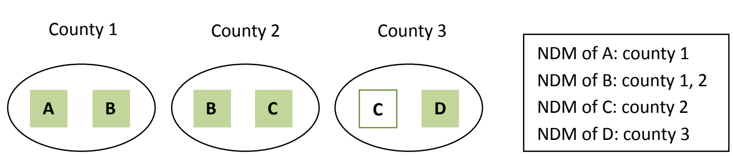
The term "market" is typically used to describe either a set of competing firms or a set of available products. This implies that a market is a geographic area that satisfies two criteria: (1) all consumers in the area face the same choice set and (2) the suppliers of these choices in the area compete with each other and with no one else. In the daily newspaper industry, however, the partially overlapping circulation areas of newspapers imply that no geographic area satisfies both criteria. For this reason, in the remainder of the paper, I use the term "choice set" to describe readers' options on the demand side and "the set of players" to describe the supply side. The latter term is justified because the supply side is modeled as a complete information two-stage game. I now specify the three basic elements of this game: the set of players, the timing and information, and the payoffs.

2.3.1 The set of players

When newspapers A and B compete in county 1 and newspapers B and C compete in county 2, the three newspapers A, B, C are all in one game because A and B, as well as B and C, are direct competitors, and hence A and C are competitor's competitors. Therefore, due to the partial overlapping of newspaper coverage, all newspapers in the U.S. are potentially in one game. To limit the number of players in a game, two assumptions are made. First, a newspaper competes only with the newspapers in its Newspaper Designated Market. The Newspaper Designated Market (NDM) is a set of counties that a newspaper reports (to the Audit Bureau of Circulations, a nonprofit circulation-auditing organization, and advertisers) as the market it serves. It is a predetermined subset of the counties where a newspaper circulates. The second assumption is that the behavior of the three national newspapers Wall Street Journal, New York Times and USA Today is taken as given in the model.

Figure 1 illustrates the definition of a set of players. The formal definition follows. In the

Figure 1: An Example of a Set of Players



example, if a newspaper circulates in a county, it is in the oval representing this county. If this county is also in its NDM, the newspaper is shaded. For example, newspaper C circulates in county 2 and 3. But its NDM consists of county 2 only. Therefore, according to the first assumption, it only competes with B in county 2, and does not compete with newspaper D in county 3. Because A and B are direct competitors in county 1, B and C are direct competitors in county 2, and these three newspapers do not have economic interaction with other newspapers, the set of players is given by the publishers of newspaper A, B and C.

Formally, two newspapers j and j' are defined as *interacting directly* if there exists at least one county that is in the NDMs of both newspapers. Two newspapers j and j' are defined as *interacting* if either j and j' interact directly or there exist a set of newspapers $\{h_n\}_1^N$ such that j interacts with h_1 directly, h_n interacts with h_{n+1} directly for $n = 1, ..., N-1$, and h_N interacts with j' directly. The set of players in a game is defined as the owners of the set of newspapers such that every newspaper interacts with some other newspaper in this set and none of the newspapers interacts with newspapers outside this set. In other words, a set of players is defined as the publishers of the closure of the *interacting* relation.[15] In the rest of the paper, I refer to a newspaper in the closure as a player newspaper and its publisher as a player publisher.

Some more notation is necessary for the remaining description of a typical game. For a given game, let $\mathcal{M}$ be the set of player publishers in the game with m being a typical element, and $\mathcal{J}_m$ be the set of player newspapers owned by m in this game. $\mathcal{J} = \cup_{m \in \mathcal{M}} \mathcal{J}_m$ represents the set of player newspapers in the game.

Under the two assumptions above, there might be newspapers that circulate in the NDMs of the player newspapers in a game but are not player newspapers in the game. They are called "non-players" in this game. For example, the three national newspapers are non-players. Since non-players in a game are assumed not to compete with the player newspapers, their choices of quality and prices do not depend on those of the players.[16] In other words, their quality and prices are taken as given in the game.

2.3.2 Timing and Information

The timing of the game is illustrated in Figure 2. The set of newspapers that each newspaper publisher owns, the NDM for each newspaper and the county demographics are predetermined before the start of the game. The exogenous newspaper characteristics, $\boldsymbol{y}$, are predetermined as well. All aspects of a non-player newspaper are taken as exogenous in the model. They are assumed to be realized before the start of the game.

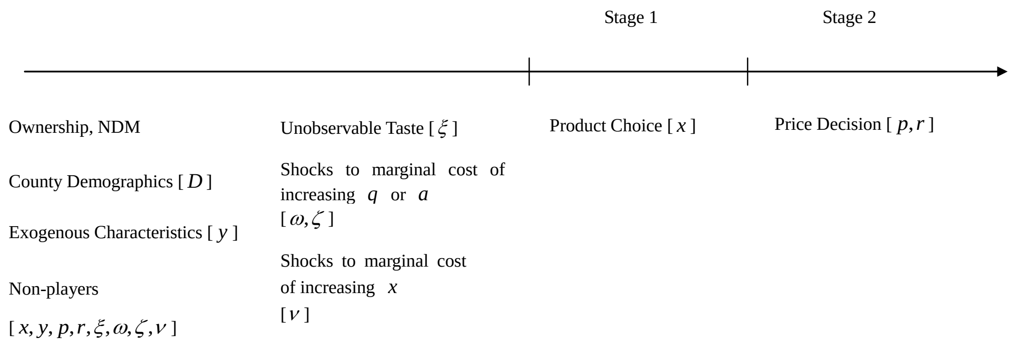
At the beginning of the game, the shocks are revealed: the newspaper/county specific taste (ξ_{jc})[17] and the marginal cost shocks $(\nu_{kj}, \omega_j, \zeta_j)$ to be specified below. The information is public

[15] This does not mean that all newspapers owned by a newspaper publisher are in one game. In fact, a newspaper publisher can be a player in different games. But since there is no economic interaction between two player newspapers in different games, one can without loss of generality label one newspaper publisher in two different games as two different newspaper publishers.

[16] But non-players quality and prices affect the players' decisions as they influence the newspaper demand.

[17] It is actually a newspaper/county/year specific shock. The subscript t is omitted here and also in $(\nu_{kj}, \omega_j, \zeta_j)$.

Figure 2: Timing of the Game



to all players in the game. Given this information, all players simultaneously choose the quality characteristics of their newspapers in the first stage. In the second stage, all players observe the newspaper characteristics and choose prices — newspaper subscription prices and advertising rates — simultaneously.

2.3.3 Payoffs

The profit of a newspaper publisher comes from both circulation profit and advertising profit. The advertising demand described in Section 2.2 is really demand for display advertising, which is printed on the newspapers' pages along with the news. In fact, there exists another type of advertisement: preprints, which are inserted into each copy of a newspaper and distributed along with it. This is essentially a delivery service provided by newspapers. I do not observe the advertising rate for preprint. Therefore, the preprint profit is not derived from a demand model. Instead, I assume that it is a simple quadratic function of circulation:

$$\mu_1 q_j + \frac{1}{2}\mu_2 q_j^2.$$

I now specify the cost structure. The demand for newspapers described in Section 2.1 and the demand for display advertising in Section 2.2 are both for annual demand: annual subscribers and annual advertising linage. Correspondingly, the costs modeled below are annual costs.

The cost of a newspaper consists of two parts: variable cost (variable with production) and fixed cost (fixed with respect to production). One variable cost is the cost of printing and delivery. It varies with circulation, q_j, and its marginal depends on publication frequency and the number of pages. I assume this marginal cost, $mc_j^{(q)}$, to be constant to circulation:[18]

$$mc_j^{(q)} = \gamma_1 + \gamma_2 f_j + \gamma_3 n_j f_j + \omega_j,$$

[18]In a more general specification, I do not find significant evidence of economies of scale.

where f_j is the publication frequency measured by the number of issues per year, n_j is the average number of pages per issue, and ω_j is a shock to the marginal cost. The annual number of pages, $n_j f_j$, is the sum of annual news hole (x_{1j}) and display advertising linage (a_j). Hence, the marginal cost can now be expressed in terms of characteristics, advertising linage, and the cost shock:[19]

$$\widetilde{mc}^{(q)}\left(f_j, x_{1j}, a_j, \omega_j; \boldsymbol{\gamma}\right) = \gamma_1 + \gamma_2 f_j + \gamma_3\left(x_{1j} + a_j\right) + \omega_j. \tag{12}$$

Note that no other newspaper characteristics besides news hole (x_{1j}) and frequency (f_j) affect the marginal cost. That is because the cost of increasing some characteristics of a newspaper, such as the number of reporters, is independent of circulation.

Another variable cost is the advertising sales cost. It is assumed to be

$$\widetilde{mc}^{(a)}\left(\zeta_j; \bar{\zeta}, \lambda_2\right) = \left(1 + 1/\lambda_2\right)\left(\bar{\zeta} + \zeta_j\right), \tag{13}$$

where λ_2 is the price elasticity of display advertising demand and ζ_j is a mean-zero exogenous random variable.[20]

Finally, the fixed cost (fixed with respect to circulation and advertising sales) consists of the cost of choosing a certain combination of newspaper quality characteristics. I assume that the marginal cost of increasing the k^{th} endogenous characteristic x_{kj} is

$$\widetilde{mc}^{(x)}\left(x_{kj}, \nu_j; \boldsymbol{\tau}_k\right) = \tau_{k0} + \tau_{k1} x_{kj} + \nu_{kj}, \tag{14}$$

where ν_{kj} is the shock to the marginal cost of increasing the characteristic. The fixed cost $\widetilde{fc}\left(\boldsymbol{x}_j, \boldsymbol{\nu}_j; \boldsymbol{\tau}\right)$ is then the sum of the integrals $\left(\sum_{k=1}^{K}\left(\tau_0 + \frac{1}{2}\tau_k x_{kj} + \nu_{kj}\right) x_{kj}\right)$ plus a constant.

Let $\theta = \left(\alpha, \boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\varphi}, \rho, \boldsymbol{\sigma}, \kappa, \boldsymbol{\phi}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma}, \bar{\zeta}\right)$ be the collection of parameters that are relevant for the second-stage decision. Denote the variable profit from newspaper j by $\tilde{\pi}_j^{\mathrm{II}}\left(\boldsymbol{p}, \boldsymbol{r}; \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{D}, \boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\zeta}; \theta\right)$, where $\boldsymbol{p}$ is a vector of subscription prices for all newspapers, player or non-player, in the game, and $\left(\boldsymbol{r}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{D}, \boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\zeta}\right)$ are analogously defined as vectors of attributes of all newspapers in the game. Variable profit is the difference between revenue and variable cost:

$$\tilde{\pi}_j^{\mathrm{II}}\left(\boldsymbol{p}, \boldsymbol{r}; \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{D}, \boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\zeta}; \theta\right) = \left(p_j \tilde{q}_j - \widetilde{mc}_j^{(q)} \tilde{q}_j\right) + \left(r_j \tilde{a}_j - \widetilde{mc}_j^{(a)} \tilde{a}_j\right) + \left(\mu_1 \tilde{q}_j + \frac{1}{2}\mu_2 \tilde{q}_j^2\right). \tag{15}$$

This is the profit function that is relevant for the decision in the second stage, where publishers observe the product choices, county demographics and the exogenous shocks, $\left(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{D}, \boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\zeta}\right)$, and choose the optimal prices $\left(\boldsymbol{p}, \boldsymbol{r}\right)$. If $\tilde{p}_j^*\left(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{D}, \boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\zeta}; \theta\right)$ and $\tilde{r}_j^*\left(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{D}, \boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\zeta}; \theta\right)$ are equilibrium prices, the overall profit of newspaper j can be expressed as

$$\tilde{\pi}_j^{\mathrm{I}}\left(\boldsymbol{x}; \boldsymbol{y}, \boldsymbol{D}, \boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\zeta}, \boldsymbol{\nu}_j; \theta, \boldsymbol{\tau}\right) = \tilde{\pi}_j^{\mathrm{II}}\left(\tilde{\boldsymbol{p}}^*, \tilde{\boldsymbol{r}}^*; \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{D}, \boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\zeta}; \theta\right) - \widetilde{fc}\left(\boldsymbol{x}_j, \boldsymbol{\nu}_j; \boldsymbol{\tau}\right). \tag{16}$$

[19] There is a slight abuse of notation here. The annual display advertising linage, a_j, is measured by column inches. According to *Editor and Publisher International Year Book*, a typical U.S. daily newspaper page has 6 columns with 21-inch depth. Therefore, in fact, $n_j f_j = x_{1j} + \frac{a_j}{126}$.

[20] λ_2 is added so that the optimal display advertising rate condition [RFOC](17) is simple.

2.3.4 Necessary Equilibrium Conditions

I now derive the optimality conditions for prices, advertising rates and quality characteristics.[21] Similar to Rosse (1967), these optimality conditions will be used for identifying the cost structure of newspaper production.

A newspaper publisher has a 2-dimensional price decision: it must select the subscription price and the display advertising rate for each newspaper it owns. Taking the derivative of the second stage profit function $\tilde{\pi}_j^{\mathrm{II}}$ in (15) with respect to advertising rate r_j yields the optimal display advertising rate as a function of circulation:

$$r_{jt} = \bar{\zeta} + \frac{\gamma_3}{1 + 1/\lambda_2} q_{jt} + \zeta_{jt}. \qquad \text{[RFOC](17)}$$

Similarly, combining [RFOC](17) and the first order condition with respect to subscription price gives

$$\boldsymbol{q}_m + \frac{\partial \tilde{\boldsymbol{q}}_m'}{\partial \boldsymbol{p}_m}\left(\boldsymbol{p}_m - \boldsymbol{mc}_m^{(q)}\right) + \frac{\partial \tilde{\boldsymbol{q}}_m'}{\partial \boldsymbol{p}_m}\left(\mu_1 + \mu_2 \boldsymbol{q}_m\right) - \frac{1}{\lambda_2}\frac{\partial \tilde{\boldsymbol{a}}_m'}{\partial \boldsymbol{p}_m}\boldsymbol{r}_m = 0, \qquad (18)$$

where $\boldsymbol{q}_m = (q_j, j \in \mathcal{J}_m)$ is a vector of circulations of publisher m's newspapers, $\left(\boldsymbol{p}_m, \boldsymbol{r}_m, \boldsymbol{a}_m, \boldsymbol{mc}_m^{(q)}\right)$ are analogously defined as the attributes of the newspapers owned by publisher m, and $\frac{\partial \tilde{\boldsymbol{q}}_m'}{\partial \boldsymbol{p}_m}$ is the transpose of the Jacobian matrix of $\tilde{\boldsymbol{q}}_m$.[22] The only difference between (18) and the standard first order condition for a multiple product firm involves the last two terms, which are the marginal effect on total advertising profit (preprint profit in the first term and display advertising in the second term) when there is an increase in the newspaper subscription price. Note that λ_2 is negative.

The first order condition with respect to price (18) holds for all publishers $m \in \mathcal{M}$. Inverting $\frac{\partial \tilde{\boldsymbol{q}}_m'}{\partial \boldsymbol{p}_m}$ in (18) gives the estimation equation [PFOC](19):

$$\begin{aligned} p_{jt} = & -\left[\left(\frac{\partial \tilde{\boldsymbol{q}}_m'}{\partial \boldsymbol{p}_m}\right)^{-1}\left(\boldsymbol{q}_m - \frac{1}{\lambda_2}\frac{\partial \tilde{\boldsymbol{a}}_m'}{\partial \boldsymbol{p}_m} r_m\right)\right]_{jt} \\ & + \gamma_1 + \gamma_2 f_{jt} + \gamma_3 n_{jt} f_{jt} - (\mu_1 + \mu_2 q_{jt}) + \omega_{jt}, \forall jt. \end{aligned} \qquad \text{[PFOC](19)}$$

When choosing their newspaper quality characteristics in the first stage, publishers take into account the impact of their product choice on the equilibrium price in the second stage. The formulation of the optimality condition for the product characteristics therefore requires knowledge of this impact of product choice on the equilibrium price. I take an approach different from the

[21] I assume that a pure-strategy Nash equilibrium exists. Finding a set of sufficient conditions for the existence of a Nash equilibrium of this two-stage game is beyond the scope of this paper.

[22] I follow the standard notation to denote the Jacobian of a function, $g(x) : R^n \to R^m$, as $\left(\frac{\partial g}{\partial x'}\right)_{n \times m}$ to emphasize the correspondence between the columns of the derivative and those in x'.

literature[23] by noticing that knowledge of the gradient of the equilibrium price function at the data points is sufficient to formulate the optimality conditions for the observed product characteristics. Also, the gradient at the observations can be easily computed by taking total derivatives of the first order conditions with respect to the newspaper price and the display advertising rate. Therefore, it is not necessary to compute the equilibrium price for each possible quality choice to obtain the gradient.[24]

Formally, the necessary optimality condition for the characteristics is that $\frac{\partial \sum_{h \in \mathcal{J}_m} \tilde{\pi}_h^{\mathrm{I}}}{\partial x_{kj}} = 0$ for all newspapers $j \in \mathcal{J}_m$ and all endogenous quality measures $k = 1, ..., K$. Each summand is given by

$$\frac{\partial \tilde{\pi}_h^{\mathrm{I}}}{\partial x_{kj}} = \frac{\partial \tilde{\pi}_h^{\mathrm{II}}}{\partial x_{kj}} + \sum_{j' \in \mathcal{J}} \frac{\partial \tilde{\pi}_h^{\mathrm{II}}}{\partial p_{j'}} \frac{\partial \tilde{p}_{j'}^*}{\partial x_{kj}} + \frac{\partial \tilde{\pi}_h^{\mathrm{II}}}{\partial r_h} \frac{\partial \tilde{r}_h^*}{\partial x_{kj}} - mc_{kj}^{(x)} 1\left(h = j\right), \forall h, j \in \mathcal{J}_m, \forall k.$$

The first term is the direct impact of increasing characteristic x_{kj} of newspaper j on the variable profit of newspaper h owned by the same publisher. A change in the characteristics of newspaper j also has an impact on the equilibrium subscription prices and advertising rates for all newspapers in a game, which is captured in the second and the third term in the above expression, respectively. Since in the model, the variable profit of newspaper h ($\tilde{\pi}_h^{\mathrm{II}}$) does not depend on the advertising rates of other newspapers, the indirect effect of characteristics x_{kj} on $\tilde{\pi}_h^{\mathrm{II}}$ in the third term is only through affecting the equilibrium advertising rate of newspaper h. This explains the difference between the second and the third term. Finally, the last term is the marginal cost of increasing the characteristic x_{kj}.

In this expression, $\left(\frac{\partial \tilde{\pi}_h^{\mathrm{II}}}{\partial p_{j'}}, \frac{\partial \tilde{\pi}_h^{\mathrm{II}}}{\partial r_h}, \frac{\partial \tilde{\pi}_h^{\mathrm{II}}}{\partial x_{kj}}\right)$ can be easily computed by taking derivatives of the variable profit function (15). The key is therefore to compute the gradients of the two equilibrium functions $\frac{\partial p_{j'}^*}{\partial x_{kj}}$ and $\frac{\partial \tilde{r}_h^*}{\partial x_{kj}}$. Since the equilibrium price and advertising rate satisfy the first order conditions [RFOC](17) and [PFOC](19), total differentiation of these two equations yields the gradients $\frac{\partial p_{j'}^*}{\partial x_{kj}}$ and $\frac{\partial \tilde{r}_h^*}{\partial x_{kj}}$. See Appendix B for the details. Plugging $\frac{\partial p_{j'}^*}{\partial x_{kj}}$ and $\frac{\partial \tilde{r}_h^*}{\partial x_{kj}}$ into the expression of $\frac{\partial \tilde{\pi}_j^{\mathrm{I}}}{\partial x_{kj}}$ gives

[23] A common solution in the literature is to compute the equilibrium of the whole game, i.e. to solve for the equilibrium product characteristics. Therefore, a typical estimation procedure involves a three-level nested algorithm: in the inner loop, the pricing equilibrium is solved for given product characteristics and model parameters; in the middle loop, the product choice equilibrium is solved for given model parameters; and in the outer loop, parameters are searched to minimize some estimation criterion function. The computational burden of such a nested fixed point problem is nontrivial. As a result, researchers typically use this method to study an industry with a simple market structure, such as the monopoly markets in the cable industry in Crawford and Shum (2006), or an industry where the possible choices for product characteristics are discrete and finite, such as the choice for ice cream flavors in Draganska, Mazzeo and Seim (2007).

[24] This approach, however, does require that the profit function has to be differentiable in characteristics. Also, the first order conditions of prices contain the first order partial derivatives of the profit function. Total differentiation of these conditions therefore involves the second order partial derivatives of the profit function. This requires that the model captures even the second order derivative of newspaper publishers' profit structure accurately. Note that the algorithm in the literature described in footnote 23 also requires that the model capture the true profit function accurately so that the equilibrium price function can be accurate.

the optimality conditions of characteristics, the set of estimation equations [XFOC](20):

$$\sum_{h\in\mathcal{J}_{mt}}\left(\frac{\partial\tilde{\pi}^{\mathrm{II}}_{ht}}{\partial x_{kjt}}+\sum_{j'\in\mathcal{J}_{g(jt)}}\frac{\partial\tilde{\pi}^{\mathrm{II}}_{ht}}{\partial p_{j't}}\frac{\partial p^*_{j't}}{\partial x_{kjt}}+\frac{\partial\tilde{\pi}^{\mathrm{II}}_{ht}}{\partial r_{ht}}\frac{\partial\tilde{r}^*_{ht}}{\partial x_{kjt}}\right)=\tau_0+\tau_k x_{kjt}+\nu_{kjt},\forall k,jt.^{25}\qquad\text{[XFOC](20)}$$

3 Data

For this study, I have compiled a new data from various sources. See Appendix C for a detailed explanation of the data sources and the variable definitions. It covers all daily newspapers in the United States from 1997 to 2005. Specifically, the data set contains information on quantities and prices on both sides of the market. On the readers' side, I observe county circulation and annual subscription price (q_{jct}, p_{jt}). On the advertisers' side, I observe annual display advertising linage and display advertising rate (a_{jt}, r_{jt}).

The data set also contains information on newspaper characteristics. A newspaper is described by the following attributes: (1) news hole (x_{1jt}), (2) the number of opinion section staff (x_{2jt}), (3) the number of reporters (x_{3jt}), (4) frequency of publication (f_{jt}), and (5) edition (morning or evening newspaper) (y_{2jt}). Data on all dimensions of the attributes except news hole is available. News hole is the space of a newspaper devoted to news, in other words, pages net of advertising space. As explained in the discussion leading to (12), news hole x_{1jt} can be replaced by $n_j f_j - \tilde{a}\left(r_j, q_j, \eta_j; \boldsymbol{\lambda}\right)$ in the estimation. The latter depends on observable variables and model parameters.

Data on all variables except advertising linage (a_{jt}), annual subscription price (p_{jt}) and pages per issue (n_{jt}) are available for all newspapers in the data period. Display advertising linage data is available for 485 newspaper/years between 1999 and 2005. Therefore, information on this subset of newspapers is used to identify the advertising demand parameters in [ADV](11). Missing data on price or pages per issue, however, leads to deletion of observations: all newspapers in the game of a newspaper with information on price or pages missing are deleted from the sample.[26] There are 1387 newspaper/years with missing data on price or pages, which lead to the deletion of 6551 newspaper/years, with 6316 newspaper/years remaining. Summary statistics for the main variables are provided in Table 2 and Table 3.

Since information on NDM is available only for a few newspapers, I infer NDM from data on county circulations. Specifically, for each newspaper/year jt, I sort the counties covered in

[25] $\mathcal{J}_{g(jt)}$ is the set of all player newspapers in the game that jt belongs to. It is exactly $\mathcal{J}$ defined in Section 2.3.1, where the model is described for a typical game and therefore the subscript $g(j)$ is unnecessary.

[26] This is because, for example, when the number of pages per issue of newspaper j is not observable, information on its news hole is not available, i.e. its characteristics are unobservable. Hence, for any newspaper j' in j's game, the optimality condition for characteristics conditional on j''s opponents choice, including j's choice, is not well-defined. Therefore, j's game are deleted.

[27] These observations are at the newspaper/county/year level.

[28] These observations are at the newspaper/year level.

Table 2: Summary Statistics of Player Newspapers in Sample

	mean	median	std	min	max	observations
market penetration (%)	19.13	11.77	18.62	0.3	97.08	23877[27]
county distance (100km)	0.71	0.47	0.81	0	6.64	
total circulation	22,729	9,849	43,847	1,132	783,212	6316[28]
price of newspapers ($)	101.47	97.15	33.75	15	365.31	
price of display advertising ($/column inch)	26.58	13.31	45.19	3.27	748.70	
frequency (issues/52 weeks)	310.70	312	53.87	208	364	
pages (pages/issue)	28.93	23.71	20.79	8	254.57	
opinion staff	2.11	1	2.92	0	20	
reporters	22.28	4	43.04	0	218.67	

Table 3: Summary Statistics of the Demographic Characteristics of Counties in Sample

	mean	median	std	min	max	observations
high education % of pop over 25	17.11	15.22	7.26	5.64	60.48	9357
median income ($1,000)	34.25	32.85	7.31	16.36	80.12	
median age	36.52	36.70	3.82	20.70	54.30	
urbanization (%)	49.82	50.96	26.51	0	1	
Households	36687	15588	85687	710	3282266	

descending order of county circulation and define NDM as the set of counties that covers 85 percent of total circulation: $\mathcal{C}_{jt} = \left\{(c_1, ..., c_H) \text{ s.t. } \sum_{h=1}^{H} q_{jc_ht} \geq 0.85 q_{jt} \text{ and } \sum_{h=1}^{H-1} q_{jc_ht} < 0.85 q_{jt}\right\}$, where q_{jt} is the total circulation.[29] With this "definition" of NDM, there are 3994 games in the sample.

4 Estimation

Five sets of model implications are taken to the data to estimate the model parameters. The model implications derived from newspaper demand [S](8) are used to identify readers' utility functions and those of advertising demand [ADV](11) are used to identify the dependence of advertising demand on the newspaper's circulation, its advertising rate and the demographics of its market. Optimality of the observed prices (see [RFOC](17) and [PFOC](19)) is used to identify the variable cost parameters, and optimality of the observed characteristics (see [XFOC](20)) is used to estimate the parameters related to the cost of choosing the characteristics.

The unobservable error terms in the above model implications are solved as functions of the data and the parameters, and then plugged into a set of moment conditions defined by instruments. A Generalized Method of Moments (GMM) estimator is formed based on these moment conditions. The estimation results are presented in Section 5. I now explain the instruments used in the study

[29]This is the criterion suggested by the Audit Bureau of Circulation. For the newspapers whose NDM information is observable, this information is consistent with the NDM inferred from county circulation data.

and how the model parameters are identified.

4.1 Instruments

In the model, newspaper publishers know the unobservable (to econometricians) newspaper-county specific taste ξ_{jct} and the unobservable cost shocks $\left(\zeta_{jt}, \omega_{jt}, \boldsymbol{\nu}_{jt}\right)$ before they choose the characteristics, the subscription prices and the advertising rates of their newspapers. These choices are therefore likely to be correlated with the unobservables. Instrumental variables are used to deal with this endogeneity.

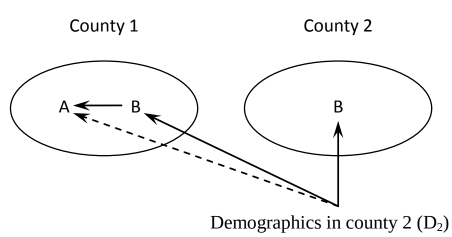
Specifically, I use the following as instruments for the newspaper quality and the price of newspaper j: the demographics in its own NDM, denoted by $\boldsymbol{D}_{jt}^{(1)}$, and the demographics in the NDMs of its competitors, denoted by $\boldsymbol{D}_{jt}^{(2)}$.

To see why $\boldsymbol{D}_{jt}^{(1)}$ is a valid instrument, first note that consistent with the timing of the model, all unobservable shocks are assumed to be revealed after the NDM of each newspaper is determined and are therefore uncorrelated with $\boldsymbol{D}_{jt}^{(1)}$. This timing assumption is plausible because location decisions are typically of a longer horizon than both quality and price decisions. But it excludes endogenous NDMs, i.e. endogenous entry/location choices. Secondly, $\boldsymbol{D}_{jt}^{(1)}$ is correlated with newspaper quality characteristics and prices. This is because county demographics affect the demand for newspapers as well as the advertising demand, which in turn, influence newspaper publishers' product quality choice and price decisions. Therefore, demographics $\boldsymbol{D}_{jt}^{(1)}$ can be used as instruments.

However, $\boldsymbol{D}_{jt}^{(1)}$ is not enough for identification. Even though it does not appear in some estimation equations such as [RFOC](17), it enters other estimation equations such as the mean utility equation [S](8). The demographics in the NDMs of newspaper j's competitors, $\boldsymbol{D}_{jt}^{(2)}$, on the other hand, is excluded from all estimation equations.

To see why $\boldsymbol{D}_{jt}^{(2)}$ can be used as an excluded instrument, first note that for the same reason as $\boldsymbol{D}_{jt}^{(1)}$, $\boldsymbol{D}_{jt}^{(2)}$ is uncorrelated with all unobservables. The intuition for why county demographics in $\boldsymbol{D}_{jt}^{(2)}$ are valid instruments for newspaper prices and qualities is illustrated in Figure 3. The

Figure 3: $D_{jt}^{(2)}$ as Instruments



demographics in county 2 (D_2) influence the demand for newspaper B as well as its advertising demand, and thus determine the prices and the attributes of this newspaper. Because newspapers A and B are direct competitors, B's decision on product characteristics and prices affects A's decision. Therefore, the demographics in county 2, D_2, indirectly affect newspaper A's product choice and price decisions. For example, a local newspaper in a small county close to a large city with a metropolitan newspaper might want to position itself as an inexpensive and low-quality newspaper. Thus, the demographics of the NDMs of a newspaper's competitors are good instruments for this newspaper's prices and quality characteristics.[30]

I now show the "quality" of the instruments. First of all, Table 4 reports the correlation of demographics in neighboring counties, specifically, the correlation between the educational level, for example, of a county in the NDM of a newspaper and the mean of the educational levels of other counties in this newspaper's NDM. The table shows that the demographics of neighboring counties are not highly correlated, i.e. the included instrument and the excluded instrument are not highly correlated.

Table 4: Correlation of Demographics in Neighboring Counties

	educational level	median income	median age	urbanization
correlation	0.1725	0.2388	0.1179	0.1369

Secondly, Table 5 shows the results from the first-stage regression corresponding to the mean utility equation ([S](8)). Specifically, column (1) reports the regression of the newspaper price on the included and excluded instruments. Columns (2) and (3) report the regression of the number of opinion section staff and the number of reporters on the instruments.

Note that among the demographic measures, only the number of households in a county varies across years. This is because the county-level demographics data comes from Census and a yearly data is not available. The exogenous sources of variation that lead to changes in prices and newspaper characteristics over time include the variation in market structure, — such as ownership (which newspapers belong to a newspaper company) and the overlap of NDM (which newspapers have overlap in which counties) — and the time trend.

4.2 Identification

The parameters to be estimated are (i) the parameters in the newspaper demand function, $(\alpha, \boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\varphi}, \rho, \boldsymbol{\sigma}, \kappa)$; (ii) the parameters in the display advertising demand function $(\boldsymbol{\phi}, \boldsymbol{\lambda})$; (iii) the cost parameters $\left(\boldsymbol{\gamma}, \bar{\zeta}, \boldsymbol{\tau}\right)$; and (vi) the parameters in the preprint profit function μ.

[30]This intuition behind this choice of instrument is similar to that in BLP, who uses the characteristics of competitors' products as instruments. These instruments are valid ecause firms consider what kind of products are available in the market when making a price decision and the product characteristics are assumed to be exogenous in BLP. Similarly, in this paper, where the product characteristics are endogenous, firms make a decision on product characteristics and prices based on what kind of consumers they serve, i.e. demographics. The demographics of competitors' NDM therefore can be used as excluded instruments.

Table 5: First-stage Regression Results

			(1) Price	(2) Opinion	(3) Reporter
included instruments	const		2174.22**	-19.52**	-188.24**
			(215.69)	(3.44)	(4.66)
	time		-1.11**	0.01**	0.09**
			(0.11)	(0.002)	(0.002)
	log(households in the NDM) ($y_{jt}^{(1)}$)		10.13**	0.45**	0.63**
			(0.36)	(0.006)	(0.008)
	morning edition ($y_{jt}^{(2)}$)		13.13**	0.06**	0.17**
			(0.68)	(0.01)	(0.01)
	county distance ($y_{jc}^{(3)}$)		40.59**	0.96**	1.50**
			(9.48)	(0.15)	(0.20)
	demographics (D_c)	education	28.88**	0.79**	0.97**
			(5.46)	(0.09)	(0.12)
		median income	-0.14	-0.31**	-0.27**
			(5.70)	(0.09)	(0.12)
		median age	42.81**	1.30**	1.25**
			(8.39)	(0.13)	(0.18)
		urbanization	2.03	0.07**	0.06*
			(1.48)	(0.02)	(0.03)
excluded instruments	mean of demographics over counties in the NDM of j except county c	education	52.43**	1.41**	1.61**
			(9.70)	(0.15)	(0.21)
		median income	-13.05	1.02**	1.71**
			(8.44)	(0.13)	(0.18)
		median age	39.38**	-1.35**	-1.78**
			(5.69)	(0.09)	(0.12)
		urbanization	-3.44	0.13**	-0.01
			(2.12)	(0.03)	(0.05)
	mean of demographics over counties in the NDMs of j's competitors but not in j's NDM	education	-47.19**	-2.12**	-1.14**
			(10.22)	(0.16)	(0.22)
		median income	69.31**	0.82**	0.51**
			(7.71)	(0.12)	(0.17)
		median age	-8.59	0.01	-0.33**
			(5.64)	(0.09)	(0.12)
		urbanization	-19.35**	-0.02	-0.28**
			(2.63)	(0.04)	(0.06)

** indicates 95% level of significance and * indicates 90% level of significance.

The identification of $(\alpha, \boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\varphi}, \boldsymbol{\sigma})$ in the first set of parameters in [S](8) is similar to the identification of analogous parameters in BLP. In BLP, product characteristics are considered exogenous. They are therefore used as the exogenous source of change in prices and in the choice set facing consumers to identify demand price effects and substitution patterns (which are parameterized by the above parameters). In this paper, product characteristics are endogenized. As explained in Section 4.1, I therefore use a different exogenous variation to identify the effects of product characteristics and prices: the variation of county demographics.

The parameter that describes the time trend in the outside choice, ρ, is identified by the overall change of newspaper circulation over time. Identification of the diminishing utility parameter, κ, comes from variation in the number of newspapers in a county. In counties with only one newspaper, diminishing utility does not play a role in determining market penetrations. Suppose all parameters were identified using the data from such counties only. Then, based on these estimates, market penetrations in counties with multiple newspapers could be computed assuming that each household chooses *at most one newspaper*. The difference between the observed data and these counterfactual market penetrations assuming a single choice is then explained by the choice of a second newspaper, the probability of which is determined by κ as well as the price and quality of the available newspapers.

The second set of parameters is in the advertising demand in [ADV](11). $\boldsymbol{\phi}$ is the vector of the parameters determining the dependence of display advertising on the demographics of a newspaper's market. It is identified by variation in county demographics. For example, suppose two newspapers have the same circulation and advertising rate, but their circulation areas have different income levels. Any difference in advertising linage then identifies the parameter ϕ corresponding to median income.

λ_1 and λ_2 are the display advertising demand elasticities with respect to circulation and advertising rate, respectively. However, λ_1 and λ_2 cannot be separately identified from information on advertising linage only. This is because the source of variation in circulation and advertising is identical.[31] In other words, any exogenous variation that changes circulation also changes the advertising rate. But the price elasticity of advertising demand λ_2 determines optimal advertising rates for publishers. Therefore, λ_2 can be identified using the optimality condition with respect to the advertising rate,[32] which then leads to the separate identification of λ_1 and λ_2.

It is common in the literature to use firms' price decisions to identify the cost structure, for

[31] According to the model, the advertising rate only depends on circulation and the unobservable shocks (to the advertising sales cost) as can be seen in [RFOC](17): $r_{jt} = \bar{\zeta} + \frac{\gamma_3}{1+1/\lambda_2} q_{jt} + \zeta_{jt}$.

[32] Specifically, exogenous variation that leads to changes in circulation identifies $\frac{\gamma_3}{1+1/\lambda_2}$ in [RFOC](17), where γ_3 is the marginal cost of printing one page. λ_2 and γ_3 are then separately identified with exogenous variations in county demographics that change circulation but not the number of pages of a newspaper. This can be seen from [PFOC](19). The aforementioned exogenous variations change the marginal effect of increasing price on advertising demand through changing circulation ($\frac{\partial \tilde{\boldsymbol{a}}'_m}{\partial \boldsymbol{p}_m}$), but do not change the number of pages printed in a year ($n_{jt} f_{jt}$).

example, Rosse (1967). The idea is as follows. With identification of the demand system, the marginal revenue is also identified. Then, the optimal choice of price reveals information on the marginal cost. This is how variable cost parameters $(\boldsymbol{\gamma}, \bar{\zeta})$ in [PFOC](19) and [RFOC](17) are identified. Similarly, after identification of the variable cost system, the marginal benefit of increasing quality is also identified. Then, the optimal choice of quality characteristics reveals the underlying cost of increasing them, parameterized by $\boldsymbol{\tau}$ in [XFOC](20). For example, suppose an exogenous shock in county demographics or a change in market structure increases the marginal benefit in the variable profit from enlarging the reporter group. Then, the observed change in the number of reporters identifies $\boldsymbol{\tau}_k$ for reporters, i.e. the parameter affecting the marginal cost of increasing reporters.

Since the marginal preprint profit can be considered a subsidy to the marginal cost of increasing circulation (see [PFOC](19)), its identification is similar to that of the marginal cost parameters. An exogenous shock to the market size of a newspaper, for example, increases its circulation. Variation in the marginal benefit of increasing circulation, the left hand side of [PFOC](19), which is determined by the identified demand system, then identifies μ_2. The parameter μ_1, however, cannot be separately identified from γ_1. Recall that marginal cost of increasing circulation is $mc_{jt}^{(q)} = \gamma_1 + \gamma_2 f_{jt} + \gamma_3 n_{jt} f_{jt} + \omega_{jt}$, and the marginal benefit in preprint profit is $\mu_1 + \mu_2 q_{jt}$. As a result, it is $\gamma_1 - \mu_1$ that is relevant for publishers' decisions, not the values of γ_1 and μ_1 separately. Hence, only the difference can be identified.

5 The Estimation Results

The estimation results are presented in Table 6. The endogenous newspaper characteristic vector, $\boldsymbol{x}_{jt}$, includes news hole, the number of staff for the opinion sections and the number of reporters.[33] The estimates of the mean taste (β) and the disutility from price (α) imply that a combination of doubling the news hole of a newspaper and increasing its annual subscription price by 8.5 dollars leaves the mean utility unchanged. Since the estimated reader heterogeneity ($\boldsymbol{\sigma}$) is small, this also means the demand for newspapers would not change much in such a scenario. Similarly, decreasing the number of opinion section staff by half is equivalent to increasing the subscription price by 140 dollars, and decreasing the number of reporters by half is tantamount to increasing price by 24.5 dollars. The market size of a newspaper is measured by the logarithm of the number of households in its NDM. The negative sign of ψ_1 indicates that readers value a newspaper with, for example, 10 reporters and covering a small region more than a newspaper that has 10 reporters and serves a large area. County demographics used in this paper include

[33] In the estimation, I replace x_{kjt} in the utility function by $\log(1 + x_{kjt})$, as this specification of newspaper characteristics explains the data better. In the cost function, I use x_{jt}.

Table 6: Estimation Results

		parameter	estimate	standard error
Utility	price ($100)	α	-0.560**	0.166
	log(1+newshole), mean	β_1	0.069	0.147
	log(1+opinion), mean	β_2	1.128**	0.331
	log(1+reporter), mean	β_3	0.198*	0.108
	log(1+newshole), std. dev.	σ_1	0.013	0.837
	log(1+opinion), std. dev.	σ_2	0.008	11.501
	log(1+reporter), std. dev.	σ_3	0.009	2.099
	log(households in the NDM)	ψ_1	-1.395**	0.307
	morning edition	ψ_2	0.161	0.122
	county distance (1000km)	ψ_3	-2.117	1.578
	constant	φ_1	6.616**	1.730
	education	φ_2	4.744**	1.240
	median income ($10000)	φ_3	-1.506*	0.889
	median age	φ_4	0.165**	0.037
	urbanization	φ_5	2.699**	0.726
	time	ρ	1.909**	0.431
	Diminishing Utility	κ	46.258**	14.343
Display Ad Demand	total circulation	λ_1	1.758**	0.005
	ad rate	λ_2	-1.015**	0.022
	constant	ϕ_1	-1.824**	0.521
	median income ($10000)	ϕ_2	0.029	1.224

			parameter	estimate	standard error
$mc^{(q)}$		const	$\gamma_1 - \mu_1$	-575.810**	74.856
		frequency	γ_2	1.656**	0.374
		1000 pages	γ_3	1.831	2.660
$mc^{(a)}$			$\bar{\zeta}$	3.963**	0.559
$mc^{(x)}$	opinion	constant	τ_{20}	1329509**	377660
		opinion	τ_{21}	113940**	26712
	reporter	constant	τ_{30}	194435*	116630
		reporter	τ_{31}	1430	1127
Preprint Profit		total circulation (1000)	μ_2	-0.0001**	0.00009

** indicates 95% level of significance.

* indicates 90% level of significance.

educational level, median income, median age and urbanization,[34] of which educational level, age and urbanization positively affect the demand for newspapers. The positive sign of ρ indicates that readers' utility from subscribing to a newspaper is decreasing over time. This is consistent with the advent of online news, which motivates the inclusion of the time trend in the model.

The parameter κ measures the diminished utility of subscribing to a second newspaper. In a single discrete choice model, this parameter is essentially set to be infinite so that consumers buy at most one product. The estimate of κ in this multiple discrete choice model implies that in the majority of the year/county pairs, the percentage of households with two newspapers is close to zero. In the 89 year/county pairs with a nontrivial percentage of households with two newspapers, on average 10% of the households with newspapers subscribe to two newspapers.

All parameters in the advertising demand function have the expected sign: an increase in circulation and a decrease in advertising rate raise advertising demand. The price elasticity of display advertising demand is close to -1. The elasticity with respect to circulation, however, is larger than 1. As will be explained in the next section, this has an important implication for how publishers adjust the quality and price of their newspapers after a market structure change.

I set the parameters in the marginal cost of increasing news hole (τ_{10}, τ_{11}) to zero, because specifications that do not restrict these parameters indicate that news hole does not affect the fixed cost (fixed with respect to circulation). This can be explained as follows. News hole consists of stories written by reporters and those bought from news agencies. The former can be increased by hiring more reporters. But this effect on fixed cost is already captured by the cost of reporters. The cost of the latter is de facto a variable cost, because news agencies typically set their rates based on the circulation of a newspaper instead of the number of stories that the newspaper buys.

6 Counterfactual Simulations

In this section, I use counterfactual simulations to study how a change in market structure affects the product choice and price decisions of newspaper publishers. The resulting welfare implications are also investigated. Section 6.1 discusses the welfare measures used: reader surplus, advertiser surplus and publisher surplus. Section 6.2 studies a merger of two direct competitors in the Minneapolis market that was blocked by the Department of Justice. This section also analyzes the effect of a merger of two newspapers in this market that do not compete directly, but share a common competitor. A welfare analysis of ownership consolidation in duopoly and triopoly markets, where the publisher of the largest newspaper buys the second largest, is presented in Section 6.3. The correlation between welfare effects and the underlying market structure is also studied. Throughout this section, I use "ownership consolidation" and "merger" interchangeably.

[34]See Appendix C for the definitions of these county demographics.

6.1 Welfare Measures

Reader Surplus

The compensating variation for household i is given by

$$CV_{ict} = \frac{V^0_{ict} - V^1_{ict}}{\alpha},$$

where α is the negative of the household's marginal value of income, and $V^0_{ict} - \alpha I_i$ and $V^1_{ict} - \alpha I_i$ are the expected maximum utility (with respect to the extreme value taste shocks) before and after a merger for household i with income I_i. Specifically,[35]

$$V^0_{ict} = \ln\left(\sum_{j=0}^{J_{ct}} e^{U^0_{ijct}}\right) + \sum_{j=1}^{J_{ct}} \ln\left(\sum_{h\neq 0,j} e^{U^0_{ihct}-\kappa} + 1\right) - (J-1)\ln\left(\sum_{h\neq 0} e^{U^0_{ihct}-\kappa} + 1\right),$$

where $U^0_{ijct} = u^0_{ijct} - \varepsilon_{ijct}$ is the utility before the merger net of the extreme value taste shock.[36] V^1_{ict} is analogously defined to V^0_{ict}, replacing U^0_{ijct} by U^1_{ijct} and u^0_{ijct} by u^1_{ijct}.

Three welfare measures are reported. (1) Change in the average per-household reader surplus in county c in year t is measured by $\overline{\Delta RS_{ct}} = E_{\boldsymbol{\zeta}_i}(CV_{ict})$. (2) Total welfare change is the sum of the welfare change in all the counties in a game: $\Delta RS = \sum_{ct} D_{Lct}\overline{\Delta RS_{ct}}$, where D_{Lct} is the number of households in county c in year t. (3) Change in average per-household reader surplus: $\overline{\Delta RS} = \frac{\Delta RS}{\sum_{ct} D_{Lct}}$.

Advertiser Surplus

Since I only observe the advertising linage for each newspaper, instead of each advertiser's individual behavior, only the price elasticity of the market demand for advertising is identified, i.e. $\lambda_2 = \frac{1}{\lambda'_2+\lambda'_3-1}$. But due to the negative externality of aggregate advertising on the effectiveness of individual advertising, the market demand does not correspond to an individual agent's willingness to pay. Thus, information on the market demand function is not enough to measure advertiser surplus.

This can be seen as follows. The representative advertiser's profit function is given by (9) in Section 2.2. Plugging the advertising demand function (10) into the advertiser's profit function gives the measure for advertiser surplus

$$AS = \left(\frac{1}{\lambda'_3} - 1\right) a_j r_j,$$

where $\frac{1}{\lambda'_3-1}$ is the representative advertiser's demand elasticity with respect to price (see (9)). Since the representative advertiser's price elasticity parameter λ'_3, and the externality parameter,

[35] The derivation of this expression follows directly from Small and Rosen (1981) for a single discrete choice model. The only difference is the sum of the second and third term, the expectation (with respect to the extreme value taste shocks) of the second highest utility.

[36] After subtracting the idiosyncratic taste term ε_{ijct}, U^0_{ijct} still depends on the household-specific taste for newspaper characteristics ($\boldsymbol{\zeta}_i$), hence the expectation operator.

λ_2', cannot be identified separately given only aggregate data, I report the percentage change in advertiser surplus, which is essentially the percentage change in display advertising revenue, $a_j r_j$.

Publisher Surplus

Publisher surplus is given by the profit function in (16).

6.2 Two Case Studies in the Minneapolis/St. Paul Metropolitan Area

Case 1. Ownership Consolidation of Direct Competitors

In 2006, the McClatchy Company purchased its much larger rival Knight Ridder Inc. After the acquisition of Knight Ridder, McClatchy owned two daily newspapers in the Minneapolis/St. Paul metropolitan area: the *Minneapolis Star Tribune* and the *St. Paul Pioneer Press*. Three months after the announcement of the transaction, the Department of Justice filed a complaint. Two months later, McClatchy sold the *St. Paul Pioneer Press* to the Hearst Corporation, which later sold it to MediaNews Group. Neither Knight Ridder nor MediaNews owns another newspaper in this market. Therefore, this series of transactions did not lead to a market structure change in the framework of this paper, as the publisher of the *St. Paul Pioneer Press* was simply relabeled.

In this section, I investigate what would have happened to newspaper quality, subscription prices as well as advertising rates and welfare if the ownership consolidation of the *Minneapolis Star Tribune* and the *St. Paul Pioneer Press* had been upheld. These two newspapers are in a game with three other newspapers: the *Faribault Daily News*, the *Stillwater Gazette* and the *St. Cloud Times*. The NDMs of all newspapers are illustrated in Figure 4. The Minneapolis-based *Star Tribune* and the St. Paul-based *Pioneer Press* (henceforth, *Star* and *Pioneer*) are direct competitors as their NDMs overlap in five counties. *Star* circulates in a larger area. Whereas the *Stillwater Gazette* competes with both *Star* and *Pioneer* in Washington County, the *Faribault Daily News* and the *St. Cloud Times* compete with *Star* only.

Findings

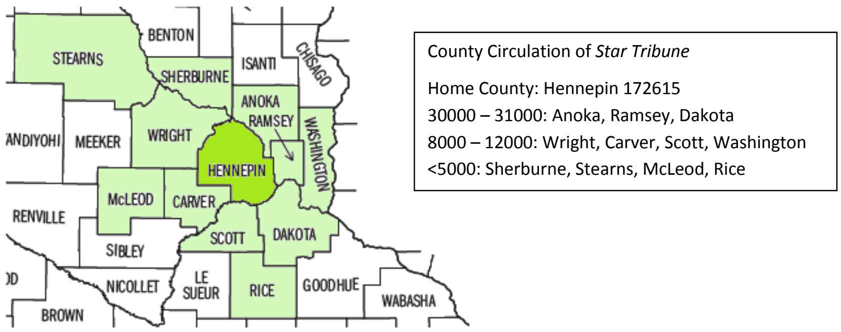
Table 7 and Table 8 present newspaper quality characteristics, subscription prices and advertising rates at the post-merger equilibrium when only prices are adjusted (Table 7) and when both quality and prices are endogenously chosen by publishers (Table 8).

From the table, we can see that: (1) Without quality adjustment, both *Star* and *Pioneer* increase their subscription prices. (2) With quality adjustment, both parties to the merger increase news hole, reduce the number of opinion-section staff and reporters, and decrease the overall newspaper quality.[37] (3) The adjustment of the smaller party to the merger (*Pioneer*) is much bigger than that of the larger party (*Star*) in both scenarios — with or without quality adjustment. (4) In

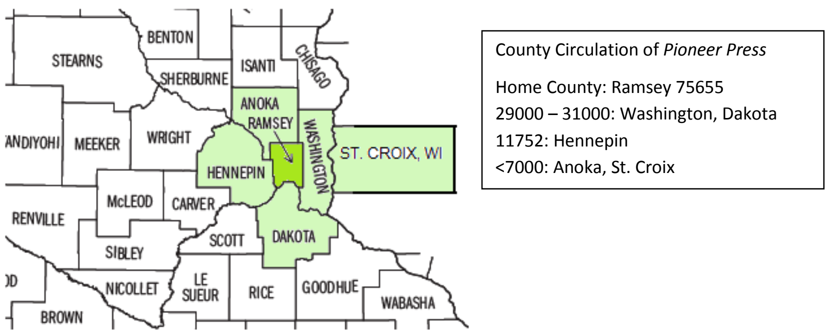
[37]Since the estimated reader heterogeneity is small, overall quality of newspaper j can be defined by the mean utility from each characteristic: $\beta_1 \log(1 + x_{1j}) + \beta_2 \log(1 + x_{2j}) + \beta_3 \log(1 + x_{3j})$.

Figure 4: Newspaper Designated Market (NDM)

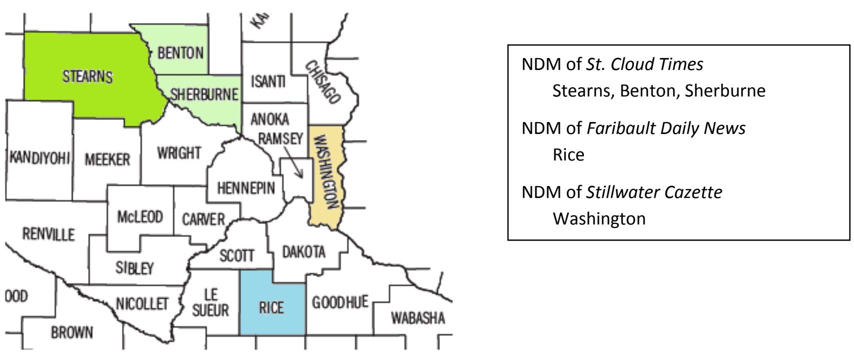
(a) NDM of Star Tribune



(b) NDM of Pioneer Press



(c) NDM of St. Cloud Times, Faribault Daily News, Stillwater Gazette



both scenarios, the *Stillwater Gazette*, which competes with the two parties to the merger, lowers its subscription price. Its quality increases when quality adjustment is allowed. (5) The magnitude of the price adjustment in the two scenarios are different: the price adjustment for the two parties to the merger is smaller when quality adjustment is allowed, while that for the *Stillwater Gazette* is larger. In other words, ignoring quality adjustment leads to an overestimation of the price adjustment for the newspapers involved in the merger and an underestimation for the newspaper

Table 7: Effect of Ownership Consolidation on Prices and Circulation without Quality Adjustment

	price ($/year)			ad rate ($/column inch)			circulation		
	before	after	change	before	after	change	before	after	change
Star Tribune	173	182	9	230.88	223.90	-6.98	317337	310288	-7049
Pioneer Press	172	204	32	153.08	135.31	-17.77	159864	141908	-17956
Stillwater Gazette	78	74	-4	11.13	11.47	0.34	3341	3679	338
Faribault Daily News	111	112	1	12.37	12.45	0.08	6384	6470	86
St. Cloud Times	150	150	0	44.15	44.24	0.09	24578	24667	89

Table 8: Effect of Ownership Consolidation on Quality, Prices and Circulation with Quality Adjustment

	news hole (pages/year)			opinion			reporter		
	before	after	change	before	after	change	before	after	change
Star Tribune	11639	11788	149	29.08	28.86	-0.22	110.92	110.09	-0.83
Pioneer Press	12794	14690	1896	19.92	18.84	-1.08	66.92	62.78	-4.14
Stillwater Gazette	2325	3620	1295	1	2.15	1.15	0	0	0
Faribault Daily News	7186	7178	-8	0	0	0	1	1.03	0.03
St. Cloud Times	14759	14511	-248	2.33	2.37	0.04	8	8.11	0.11
	price ($/year)			ad rate ($/column inch)			circulation		
Star Tribune	173	181	8	230.88	223.84	-7.04	317337	310224	-7113
Pioneer Press	172	198	26	153.08	131.12	-21.96	159864	137673	-22191
Stillwater Gazette	78	40	-38	11.13	15.13	4.00	3341	7385	4044
Faribault Daily News	111	112	1	12.37	12.47	0.10	6384	6487	103
St. Cloud Times	150	150	0	44.15	44.51	0.36	24578	24932	354

competing with the merged newspapers. (6) In both scenarios, the *Faribault Daily News* and the *St. Cloud Times* only adjust marginally after the merger.

Intuition Underlying the Findings

A detailed explanation of the economic interactions that drive these results can be found in Appendix D. Here, I provide some intuition underlying these observations.

(1) After the publisher of *Star*, McClatchy, purchases *Pioneer*, it internalizes the positive price cross-effect of these two newspapers: a higher price of *Star*, for example, leads to an increase in the market share of *Pioneer* and therefore raises its profit. This explains observation (1).

(2) Analogously, there also exist quality cross-effects. As will be shown in Appendix D, the cross-effect of news hole is positive and the cross-effects of opinion staff and reporters are negative at the pre-merger equilibrium. The sign of the cross-effect of news hole can be positive[38] because, unlike the other two characteristics, news hole also affects the marginal cost of increasing circulation. Specifically, increasing news hole leads to a higher marginal

[38] Appendix D shows that the sign of the cross-effect of news hole is not determinate. In this merger, it is positive at the original equilibrium.

cost $mc^{(q)}$ and hence it does not always yields an advantage over other newspapers in price competition.[39] This is consistent with observation (2).

(3) As explained in Appendix D, the estimates indicate that the advertising profit function is convex in circulation, implying that the marginal value of circulation is higher for larger newspapers. Therefore, a multi-newspaper publisher has an incentive to shift circulation from its small newspapers to large newspapers. Here, even though McClatchy decreases the quality of both newspapers, it adjusts the smaller newspaper by a bigger margin due to this incentive.

(4) An increase in the prices or a decrease in the quality of *Star* and *Pioneer*, the competitors of the *Stillwater Gazette*, leads to an increase in the latter's marginal benefit from increasing circulation. The *Stillwater Gazette* thus raises its circulation by decreasing its price and increasing its quality as in observation (4).

(5) As the publisher McClatchy internalizes the overall negative quality cross-effect between *Star* and *Pioneer*, it decreases the quality of the two newspapers involved in the merger. Also the positive price cross-effect is weakened. Therefore, price adjustments are smaller when quality adjustment is allowed. Similarly, as the quality of *Star* and *Pioneer* decreases, the marginal benefit for the *Stillwater Gazette* from decreasing its subscription price is even higher. Therefore, its price adjustment is larger when quality adjustment is allowed.

(6) Finally, observation (6) is explained by the NDMs of the newspapers involved (see Figure 4). The *Faribault Daily News* increases its price marginally because it only competes with *Star*, which does not change much after the merger. Similarly, because *Star* does not have a strong presence in the NDM of the *St. Cloud Times* , the *St. Cloud Times* almost does not adjust its price either.

Welfare Implications and Comparison to a Merger without Quality Adjustment

These adjustments in quality and subscription price influence the circulation of each newspaper and hence the optimal advertising rate. All these changes decrease the overall reader surplus by 7.94 million dollars, advertiser surplus by 5.59% and increase publisher surplus by 0.52 million dollars. So, the total welfare declines. Specifically, as Table 9 shows, households in all counties except Stearns County are worse-off. There the dominating newspaper *St. Cloud Times* increases its quality slightly. Across all counties, the average per-household reader surplus ($\overline{\Delta RS}$) declines by 6 dollars. Counties covered by the two merged parties are affected the worst. For example, readers' welfare falls by nearly 15 dollars in Ramsey County, which is the home county of *Pioneer* and close to Hennepin County, the home county of *Star*. The profits of all publishers increase. The profit of the originally smallest newspaper, the *Stillwater Gazette,* increases by a larger margin than do the profits of the two median-sized newspapers because the latter interact only marginally

[39] Quality cross-effects here take into account the impact of quality on the equilibrium price.

Table 9: Welfare Implications of the Ownership Consolidation of *Star* and *Pioneer* with Quality Adjustment

Welfare Implications	
ΔRS	-7.94 million
$\%\Delta AS$	-5.59%
ΔPS	0.52 million

Change in Publisher Surplus

newspapers	ΔPS
Star & Pioneer	374000
Stillwater Gazette	84460
Faribault Daily News	29500
St. Cloud Times	24110

Average Change in Reader Surplus per Household

county	$\overline{\Delta RS_{ct}}$	county	$\overline{\Delta RS_{ct}}$
Anoka	-4.36	Rice	-3.18
Benton	-0.70	Scott	-3.74
Carver	-3.25	Sherburne	-1.59
Dakota	-9.83	Stearns	0.43
Hennepin	-4.48	Washington	-5.44
McLeod	-2.02	Wright	-2.30
Ramsey	-14.58	St. Croix, WI	-9.10

with the two parties to the merger. *Star* and *Pioneer* lower their quality and leave space for the *Stillwater Gazette* to increase its quality. In fact, it even overtakes the *Faribault Daily News* in terms of circulation.

The welfare change *without* quality adjustment is -7.93 million for readers, -4.96% for advertisers and 0.91 million for publishers. Therefore, ignoring quality adjustment overestimates the gain in publisher surplus, and underestimates the loss in reader and advertiser surplus in this merger. The overestimation of the price adjustment noted in observation (5) can be consistent with this underestimation of the loss in readers' welfare. Even though the price adjustment is smaller with quality adjustment, the quality of the newspapers is lower as well. It is the overall utility from the newspapers that determines readers' welfare. Section 6.3 analyzes the relationship between the bias in estimating the welfare effect from ignoring quality adjustment and the underlying market structure. In particular, I show that the bias in the estimate for the change in reader surplus can be significantly larger than 10,000 dollars.

Case 2. Ownership Consolidation of Indirect Competitors

In the above ownership consolidation study, the two parties to the merger are direct competitors. This is usually the main focus in policy analyses. In fact, similar quality and price cross-effects exist even when the merged parties just share a common competitor. Therefore, an ownership consolidation of such two newspapers also affects the quality and prices of the newspapers involved. To illustrate this point and quantify the effect, I study a counterfactual merger of *Pioneer* and the *St. Cloud Times*, which do not compete directly the NDMs of these two newspapers in Figures 4(b) and 4(c) show.

The results are presented in Table 10. Again, the smaller party to the merger adjusts more than the larger party. As the *St. Cloud Times* increases news hole by about 2 pages per issue and reduces

Table 10: The Effect of the Ownership Consolidation of *Pioneer* and *St. Cloud Times* on Quality, Prices and Circulations with Quality Adjustment

	news hole (pages/year)			opinion			reporter		
	before	after	change	before	after	change	before	after	change
Star Tribune	11639	11638	-1	29.08	29	-0.08	110.92	110.92	0
Pioneer Press	12794	12802	8	19.92	19.91	-0.01	66.92	66.89	-0.03
Stillwater Gazette	2325	2327	2	1	1	0	0	0	0
Faribault Daily News	7186	7186	0	0	0	0	1	1	0
St. Cloud Times	14759	15512	753	2.33	2.23	-0.10	8	7.69	-0.31
	price ($/year)			ad rate ($/column inch)			circulation		
Star Tribune	173	173	0	230.88	230.90	0.02	317337	317362	25
Pioneer Press	172	171	-1	153.08	153.06	-0.02	159864	159847	-17
Stillwater Gazette	78	78	0	11.13	11.13	0	3341	3345	4
Faribault Daily News	111	111	0	12.37	12.37	0	6384	6484	100
St. Cloud Times	150	151	1	44.15	43.42	-0.73	24578	23839	-739

the number of opinion-section staff and reporters, its overall quality falls. Therefore, households in the counties served by it are worse off. The impact of such an ownership consolidation is much smaller than that of merging two direct competitors. The loss in readers' welfare, for example, is 18 cents on average and 3 dollars in the county that is worst affected. Overall, reader surplus decreases by 0.22 million, publisher surplus increases by 0.02 million and the change in advertiser surplus is negligible.

6.3 Welfare Analysis of Duopoly Mergers and Triopoly Mergers

In this section, I study the welfare implications of ownership consolidations in duopoly and triopoly markets. In a duopoly merger, the publisher of one newspaper buys the other and becomes a monopolist in the market. A triopoly merger is defined as one in which the publisher of the largest newspaper buying the second largest. The welfare effect of an ownership consolidation in a market depends on the details of the market structure. I will present the distribution of the welfare effects for such mergers in all duopoly and triopoly markets in the 2005 sample, and then examine how they vary with market characteristics.

Figures 5 and 6 show welfare changes after an ownership consolidation in 40 duopoly markets and 13 triopoly markets in the 2005 sample, the last year in the data set. The markets are sorted according to $\overline{\Delta RS}$, the change in average per-household reader surplus with quality adjustment. A dot in the upper-left graph of Figure 5, for example, represents $\overline{\Delta RS}$ in a market when quality adjustment is allowed. An asterisk represents $\overline{\Delta RS}$ without quality adjustment. The difference between an asterisk and a dot on the same vertical line therefore represents the bias in estimating $\overline{\Delta RS}$ when quality adjustment is ignored. The upper-right graph plots changes in total reader surplus (ΔRS). For example, in the market shown in the middle of the graph, the total reader

Figure 5: Welfare Implications of Duopoly Mergers

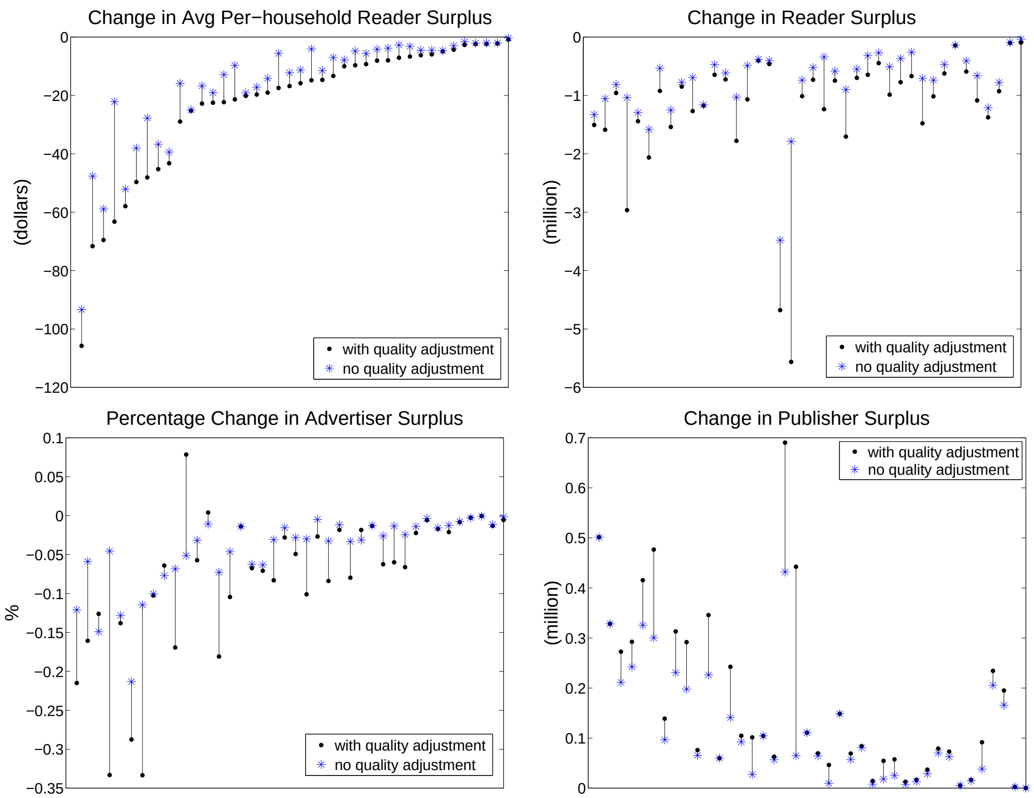


Figure 6: Welfare Implications of Triopoly Mergers

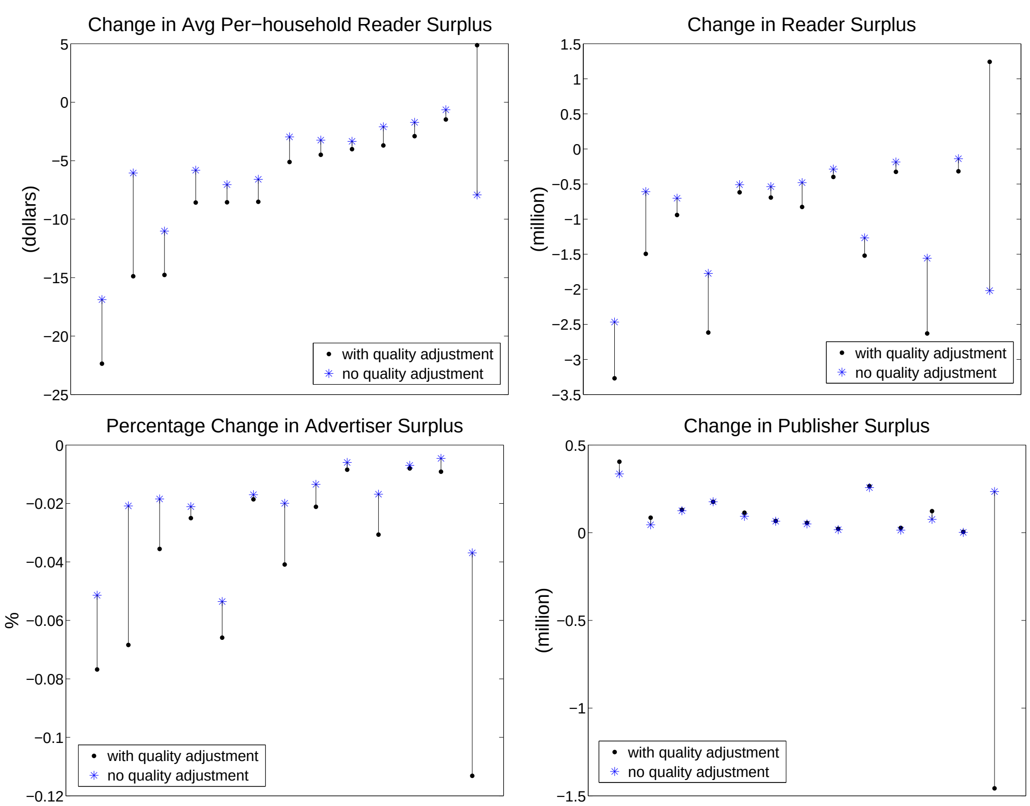


Table 11: Median Welfare Changes across Duopoly and Triopoly Markets

	$\overline{\Delta RS}$ (dollars)	ΔRS (millions)	$\%\Delta AS$ (%)	ΔPS (millions)
Duopoly with quality adjustment	-16.32	-0.97	-6.12%	0.10
Duopoly without quality adjustment	-10.52	-0.64	-3.04%	0.07
Triopoly with quality adjustment	-5.11	-0.83	-3.06%	0.09
Triopoly without quality adjustment	-5.82	-0.61	-1.85%	0.08

surplus drops by more than 5 million dollars after the merger when quality adjustment is allowed, and by around 2 million dollars without quality adjustment. Ignoring quality adjustment therefore underestimates the total readers' welfare loss by around 3 million dollars. The lower-left graph and the lower-right graph show percentage changes in advertiser surplus and changes in publisher surplus in millions, respectively. Finally, Figure 6 represents the same measures for the 13 triopoly markets.

The median changes in different welfare measures are presented in Table 11. Total welfare falls unambiguously in 38 duopoly markets.[40] And total welfare in all triopoly markets simulated falls after the merger. However, the presence of a competitor mitigates the welfare loss for readers and advertisers. This is because the merged parties downgrade the quality of their newspapers by a smaller margin than they would have done without a competitor and some even increase newspaper quality (in 6 markets). This mitigation in welfare loss is also partially due to the competitors sometimes increasing the quality. Figures 5 and 6 show that ignoring quality adjustment typically leads to an underestimation of the loss in reader surplus and the gain in publisher surplus. The median bias in estimating welfare is 4 dollars per household in triopoly mergers and 2 dollars in duopoly mergers.

To examine these welfare changes more closely, I now study the following: (1) variation of the change in average per-household reader surplus ($\overline{\Delta RS}$) across markets; (2) variation of the bias in $\overline{\Delta RS}$ when quality adjustment is ignored ($\Delta\overline{\Delta RS} = \big(\overline{\Delta RS}$ without quality adjustment$\big)$ - $\big(\overline{\Delta RS}$ with quality adjustment$\big)$); and (3) the difference between duopoly and triopoly markets. To understand how welfare effects vary across markets, I run two regressions of $\overline{\Delta RS}_m$ and $\Delta\overline{\Delta RS}_m$, where the subscript m represents a market, on some market characteristic variables. As explained before, market structure cannot be captured by simple indices. I therefore regress $\overline{\Delta RS}_m$ and $\Delta\overline{\Delta RS}_m$ on a triopoly dummy and endogenous variables, which are correlated with the underlying market characteristics that determine the welfare change. The regression therefore shows a correlation pattern rather than a causal effect.

[40]Reader surplus falls in all duopoly markets. As expected, publisher surplus increases uniformly. But the net change in the sum of reader surplus and publisher surplus is negative in all 40 duopoly markets simulated. Among 11 markets where the circulation of at least one newspaper increases, 2 markets witness an increase in advertiser surplus. Therefore, in 38 markets, the total welfare change is unambiguously negative.

The result of the first regression is presented below. Standard errors are in parentheses.

$$\overline{\Delta RS}_m = \underset{(7.77)}{25.44} - \underset{(0.16)}{0.99} pen_m - \underset{(0.05)}{0.28} overlap_m^{(1)} + \underset{(2.04)}{4.67} \log\left(\frac{q_{1m}}{q_{2m}}\right) + \underset{(3.74)}{5.04} tri_m + \underset{(0.12)}{0.25} tri_m * overlap_m^{(2)} + \varrho_m$$

The impact of ownership consolidation on readers' welfare depends on how much readers in a market value newspapers in general. Obviously, if households in a market do not like reading newspapers, then changes in newspaper quality do not affect their welfare much. The pre-merger newspaper penetration rate (pen_m), measured by the ratio of the total newspaper circulation to the number of households in market m, is used to capture this aspect of the market.[41] The negative sign in the estimation is as expected: readers' welfare loss ($-\overline{\Delta RS}_m$) increases when readers care about newspapers. An increase in the penetration rate by 1 percentage point leads to an increase in the average welfare loss per household of 99 cents.

Another market feature that affects $\overline{\Delta RS}_m$ is the importance of the merging parties' common circulation area to these two newspapers. This influences how strong the cross-effect is. Suppose that two newspapers only compete with each other in a county that is far away from their home counties. Then, this county might not play a large role in generating profit for these two newspapers because of readers' taste for local newspapers. When this is the case, a change in the quality of one newspaper does not affect the profit of the other newspaper much and thus the cross-effect is weak. Hence, the post-merger adjustment is small. This feature is captured by the pre-merger overlapping rate of the two largest newspapers in the market: $overlap_m^{(1)} = 100 \times \left(\sum_{c \in CTY_{1,2}} q_{1mc}\right)/q_{1m}$, where $CTY_{1,2}$ is the intersection of the NDMs of the two largest newspapers, and q_{1mc} and q_{1m} are county circulation (in county c) and total circulation of the largest newspaper in the market, respectively.[42] The above regression indicates a negative correlation between $\overline{\Delta RS}_m$ and $overlap_m^{(1)}$, meaning that the larger the overlapping is, the larger is the welfare loss for readers.

The third factor is the pre-merger asymmetry of the two parties to the merger in terms of circulation, measured by $\log\left(\frac{q_{1m}}{q_{2m}}\right)$. As explained in Section 6.2, due to the increasing marginal benefit of a higher circulation, the publisher of the merged parties will not adjust the quality and prices of the larger party by much. Since an adjustment in a larger newspaper has a bigger impact on readers' welfare than the same adjustment in a smaller newspaper, asymmetry matters, and specifically, the larger the asymmetry, the smaller the welfare loss for readers, as indicated by the positive sign in the above regression.

Finally, as explained before, the presence of a competitor mitigates the welfare loss for readers and advertisers, because the merged parties decrease the quality of their newspapers by a smaller

[41] pen_m also captures that for a given change in quality and prices of newspapers, the welfare change is decreasing in pre-merger utilities, i.e. the welfare loss is increasing in pre-merger utilities. This can be seen from the welfare measure in Section 6.1.

[42] The pre-merger overlapping rate can be also defined for the second largest newspaper in the market as $100 \times \left(\sum_{c \in CTY_{1,2}} q_{2mc}\right)/q_{2m}$, where q_{2mc} and q_{2m} are similarly defined for the second largest newspaper. It is not included in the regression, because it is 1 in the majority of the markets simulated.

margin when facing a competitor. Therefore, the triopoly dummy has a positive sign in the regression. The strength of the competition effect depends on how strong the cross-effect between the two merged parties and their competitor, which is again captured by the pre-merger overlapping rate: $overlap_m^{(2)} = 100 \times \left(\sum_{c \in CTY_{1,3}} q_{1mc} + \sum_{c \in CTY_{2,3}} q_{2mc}\right) / (q_{1m} + q_{2m})$, where $CTY_{1,3}$ is the intersection of the NDMs of the largest newspaper and the competitor, and $CTY_{2,3}$ is analogously defined for the second largest newspaper and the competitor. The bigger the overlap, the stronger the competition effect, and thus the smaller the welfare loss.

The second regression studies the bias in welfare effect when quality adjustment is ignored. The regression result is as follows:

$$\begin{aligned} \Delta\overline{\Delta RS}_m = & \underset{(5.23)}{2.71} \; \underset{(2.18)}{-4.86} tri_m + \underset{(0.11)}{0.30} pen_m - \underset{(0.11)}{0.23} s_{1m} + \varrho_m \end{aligned}$$

Again, the triopoly dummy and penetration rate matter. For example, the positive coefficient of pen_m means that the higher the penetration rate, the larger the bias in measuring welfare loss. Another factor that determines $\Delta\overline{\Delta RS}_m$ is the demand elasticity with respect to price. To understand this, denote the post-merger/with-quality-adjustment equilibrium by $\left(\boldsymbol{p}^1, \boldsymbol{x}^1\right)$ and that without quality adjustment by $\left(\boldsymbol{p}^2, \boldsymbol{x}^0\right)$, where $\boldsymbol{x}^0$ is a vector of the pre-merger quality of all newspapers in the market. Given that the estimated reader heterogeneity is small, what matters for readers' welfare is the mean utility component $p_j\alpha + \boldsymbol{x}_j\boldsymbol{\beta}$. I now explain how demand elasticity with respect to price affects $\left(p_j^2\alpha + \boldsymbol{x}_j^0\boldsymbol{\beta}\right) - \left(p_j^1\alpha + \boldsymbol{x}_j^1\boldsymbol{\beta}\right)$. When a publisher is prevented from setting its quality at $\boldsymbol{x}_j^1$ and has to stay at $\boldsymbol{x}_j^0$, it can increase price by $\left(\boldsymbol{x}_j^0 - \boldsymbol{x}_j^1\right)\boldsymbol{\beta}/(-\alpha)$, while keeping the mean utility and thus its circulation unchanged. But the publisher's goal is to maximize its profit, not to keep the circulation at a certain level. It will therefore continue to increase the price, and thus decrease the mean utility, until the marginal profit from doing so is 0. How much it will increase price beyond $\left(\boldsymbol{x}_j^0 - \boldsymbol{x}_j^1\right)\boldsymbol{\beta}/(-\alpha)$, i.e. the difference between $\left(p_j^2 - p_j^1\right)$ and $\left(\boldsymbol{x}_j^0 - \boldsymbol{x}_j^1\right)\beta/(-\alpha)$, depends on the price elasticity of demand. Therefore, how much readers' welfare will be affected depends on the price elasticity as well. A large elasticity leads to a small increase in price and hence a small decrease in the mean utility of the newspaper. The welfare effect of a change from $\left(\boldsymbol{p}^1, \boldsymbol{x}^1\right)$ to $\left(\boldsymbol{p}^2, \boldsymbol{x}^0\right)$ for readers is therefore small. Since the price elasticity in logit models depends positively on market shares when market shares are smaller than 1/2, I use s_{1m}, the pre-merger market penetration rate of the largest newspaper in its largest circulation county, to capture this factor. The sign in the regression result is consistent with the conjecture: a higher price elasticity of demand leads to smaller welfare changes induced by a change from $\left(\boldsymbol{p}^1, \boldsymbol{x}^1\right)$ to $\left(\boldsymbol{p}^2, \boldsymbol{x}^0\right)$, i.e. a smaller bias from ignoring quality adjustment.

7 Conclusion

In this paper, I set up a structural model and use counterfactual simulations to study the welfare implications of newspaper ownership consolidation, taking into account endogenous product choice as well as price choices. A large new data set is collected to estimate the model. Based on the estimates, I study a direct and an indirect merger in the Minneapolis market. I also quantify the welfare implications of ownership consolidations in all duopoly and triopoly markets in the 2005 sample. The distribution of the welfare effects across markets is used to study the correlation between the welfare effect of ownership consolidation in a market and the structure of the market. The main findings are as follows.

First, in the counterfactual ownership consolidation of the *Star Tribune* and the *St. Paul Pioneer Press* in the Minneapolis market, the publisher of these two newspapers decreases the overall quality and increases the prices of both newspapers. The adjustment of the *St. Paul Pioneer Press*, the smaller newspaper, is much larger than that of the bigger newspaper because advertising profit is convex in circulation and thus a multiple-newspaper publisher has an incentive to shift circulations to its larger newspaper.

Second, the simulation results show that the median loss in reader surplus is 16 dollars per household in duopoly mergers and 5 dollars in triopoly mergers. Readers' welfare loss resulting from ownership consolidation in a market is positively correlated with how much households in the market care about newspapers in general and how important the overlapping area of the two merged parties is to these two newspapers. It is negatively correlated with the asymmetry of newspaper size measured by pre-merger circulations. The existence of a competitor mitigates the loss in readers' welfare due to a competition effect; the larger the competition effect is, the smaller the welfare loss.

Third, ignoring quality adjustment typically leads to an underestimation of the loss in reader surplus and the gain in publisher surplus. In general, the bias in measuring the welfare effect of ownership consolidations is smaller in a triopoly merger and when the price elasticity of newspaper demand is higher. It is larger when households care more about reading newspapers.

Fourth, ownership consolidation has an impact on quality choice and thus welfare even when the newspapers involved in the merger do not compete directly. This welfare effect, however, is more than an order of magnitude smaller than the effect of ownership consolidation of direct competitors in the simulated market.

References

Bacon's Information Inc., *Bacon's Newspaper Directory*, 1998-2006 edition. Chicago, IL.

Berry, Steven (1994), "Estimating discrete-choice models of product differentiation." *RAND Journal of Economics*, 25, 242–262.

Berry, Steven and Philip Haile (2008), "Nonparametric identification of multinomial choice demand models with heterogeneous consumers." Working paper, Yale University.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995), "Automobile prices in market equilibrium." *Econometrica*, 63, 841–90.

Berry, Steven and Joel Waldfogel (2001), "Do mergers increase product variety? evidence from radio broadcasting." *The Quarterly Journal of Economics*, 116, 1009–1025.

Berry, Steven and Joel Waldfogel (2003), "Product quality and market size." NBER Working Papers 9675, National Bureau of Economic Research.

Crawford, Gregory and Matthew Shum (2006), "The welfare effects of endogenous quality choice: The case of cable television." mimeo, University of Arizona.

Dirks, Van Essen & Murray (2005), "Number of independent dailies steadily declines."

Draganska, Michaela, Michael Mazzeo, and Katja Seim (2007), "Beyond plain vanilla: Modeling joint product assortment and pricing decisions." Working paper.

Editor & Publisher, *Editor & Publisher International Year Book*, 1998-2006 edition. New York, NY.

Ferguson, James M (1983), "Daily newspaper advertising rates, local media cross-ownership, newspaper chains, and media competition." *Journal of Law & Economics*, 26, 635–54.

Gandhi, Amit, Luke Froeb, Steven Tschantz, and Gregory Werden (2008), "Post-Merger Product Repositioning." *The Journal of Industrial Economics*, 56, 49–67.

Genesove, David (1999), "The adoption of offset presses in the daily newspaper industry in the united states." NBER Working Papers 7076, National Bureau of Economic Research.

Gentzkow, Matthew (2007), "Valuing new goods in a model with complementarity: Online newspapers." *American Economic Review*, 97, 713–744.

Gentzkow, Matthew and Jesse Shapiro (2006), "What drives media slant? evidence from u.s. daily newspapers." NBER Working Papers 12707, National Bureau of Economic Research.

George, Lisa (2007), "What's fit to print: The effect of ownership concentration on product variety in daily newspaper markets." *Information Economics and Policy*, 19, 285–303.

Hendel, Igal (1999), "Estimating multiple-discrete choice models: An application to computerization returns." *The Review of Economic Studies*, 66, 423–446.

Mazzeo, Michael (2002), "Product choice and oligopoly market structure." *The RAND Journal of Economics*, 33, 221–242.

Nevo, Aviv (2000), "Mergers with differentiated products: The case of the ready-to-eat cereal industry." *RAND Journal of Economics*, 31, 395–421.

Nevo, Aviv, Daniel L. Rubinfeld, and Mark McCabe (2005), "Academic journal pricing and the demand of libraries." *American Economic Review*, 95, 447–452.

Rosse, James (1967), "Daily newspapers, monopolistic competition, and economies of scale." *The American Economic Review*, 57, 522–533.

Rosse, James N. (1970), "Estimating cost function parameters without using cost data: Illustrated methodology." *Econometrica*, 38, 256–275.

Rysman, Marc (2004), "Competition between networks: A study of the market for yellow pages." *Review of Economic Studies*, 71, 483–512.

Seim, Katja (2006), "An empirical model of firm entry with endogenous product-type choices." *RAND Journal of Economics*, 37, 619–640.

Small, Kenneth and Harvey Rosen (1981), "Applied welfare economics with discrete choice models." *Econometrica*, 49, 105–130.

SRDS, *SRDS Circulation*, 1998-2006 edition. Des Plaines, IL.

Appendix

A Proof of Theorem 1

In this appendix, I show that the invertibility result in BLP can be extended to a multiple discrete choice model. I only prove Theorem 1 for a multiple discrete choice model where the number of discrete choices is limited to at most two. An extension of the result to a model where consumers can choose up to $\bar{n} \leq J$ products, where J is the total number of products available in a choice set, is available upon request.

The proof is similar to that in BLP, where a key step is to show the existence of an upper bound for the fixed point of the mapping F. The main difference between the proof here and the BLP proof is at this step. In a multiple discrete choice model, the value of δ_j that solves $\sum_{h=1}^{J} s_h = \sum_{h=1}^{J} \tilde{s}_h (\boldsymbol{\delta}, \boldsymbol{x}; P_{\boldsymbol{\varsigma}}, \boldsymbol{\sigma}, \kappa)$ where $\delta_k = -\infty$ for $\forall k \neq j$ is no longer necessarily the upper bound of δ_j. This value does not even exist when the left hand side $\sum_{h=1}^{J} s_h > 1$.[43] Note that the supremum of the right hand side is 1:

$$\sup_{\boldsymbol{\delta}=(\delta_1,\ldots,\delta_J),\ \delta_k=-\infty \text{ for } k\neq j} \sum_{h=1}^{J} \tilde{s}_h (\boldsymbol{\delta}, \boldsymbol{x}; P_{\boldsymbol{\varsigma}}, \boldsymbol{\sigma}, \kappa) = 1.$$

The theorem is proved in two steps. All statements below are true for any given $(\boldsymbol{x}, P_{\boldsymbol{\varsigma}}, \boldsymbol{\sigma}, \kappa)$. Therefore, these arguments in $\tilde{s}_j$ are omitted for presentational simplicity.

Claim 1 *There exist $\underline{\delta}$ and $\bar{\delta}$ such that if F has a fixed point δ^*, δ^* must be in $[\underline{\delta}, \bar{\delta})^J$.*

Proof. Construction of the lower bound $\underline{\delta}$ is the same as in BLP. As will be shown in the proof of Claim 2, $F_j(\boldsymbol{\delta})$ is increasing in all dimensions of $\boldsymbol{\delta}$. Define $\underline{\delta}_j = \lim_{\boldsymbol{\delta}\to-\boldsymbol{\infty}^J} F_j(\boldsymbol{\delta}) = \int \exp(\boldsymbol{x}_j \boldsymbol{\varsigma})\, dP_{\boldsymbol{\varsigma}}(\boldsymbol{\varsigma}; \boldsymbol{\sigma})$. If $\boldsymbol{\delta}^*$ is a fixed point of F, $\delta_j^* = F_j(\boldsymbol{\delta}^*) \geq \underline{\delta} = \min_{j'} \left(\underline{\delta}_{j'}\right)$.

Note that $\boldsymbol{\delta}^*$ as a fixed point of F satisfies $\sum_{j=1}^{J} \tilde{s}_j(\boldsymbol{\delta}^*) = \sum_{j=1}^{J} s_j$. If two or more dimensions of $\boldsymbol{\delta}$ go to infinity, $\sum_{j=1}^{J} \tilde{s}_j(\boldsymbol{\delta})$ approaches 2. But $\sum_{j=1}^{J} s_j < 2$. So, there exists at most one j such that δ_j^* is unbounded. Given that all other $\delta_k, k \neq j$ are bounded, $\lim_{\delta_j\to\infty} \tilde{s}_j(\boldsymbol{\delta}) = 1$. So, δ_j^* has to be bounded as well to ensure $\tilde{s}_j(\boldsymbol{\delta}) = s_j < 1$. Therefore, all dimensions of $\boldsymbol{\delta}^*$ are bounded. Let the upper bound be $\bar{\delta}'$. Define $\bar{\delta} = \bar{\delta}' + 1$. □

Claim 2 *$F : [\underline{\delta}, \bar{\delta}]^J \to R^J$ has a unique fixed point.*

Proof. Define $\hat{F} : [\underline{\delta}, \bar{\delta}]^J \to R^J$ as $\hat{F}(\boldsymbol{\delta}) = \min\left(F(\boldsymbol{\delta}), \bar{\delta}\right)$. Since $\underline{\delta}$ is the lower bound of $F_j(\boldsymbol{\delta})$, $\hat{F}$ is a mapping from $[\underline{\delta}, \bar{\delta}]^J$ to itself. According to BLP, if (i) $\partial F_j(\boldsymbol{\delta}) / \partial \delta_h \geq 0$ for any h, j and (ii) $\sum_{h=1}^{J} \partial F_j(\boldsymbol{\delta}) / \partial \delta_h < 1$ for any j, then $\hat{F}$ is a contraction mapping.

[43] In a single discrete choice model, $\sum_{h=1}^{J} s_h < 1$, while in a multiple discrete choice model, the sum of market penetration for all products $\sum_{h=1}^{J} s_h$ can be larger than 1.

To show that these two conditions hold, first note that $0 < \tilde{\Psi}_j^{(1)}, \tilde{\Psi}_{j,j'}^{(2)}, \tilde{\Psi}_j^{(3)} < 1$, when $\boldsymbol{\delta}$ is in a bounded set, $\tilde{\Psi}_j^{(1)} \geq \tilde{\Psi}_j^{(3)}$ for $\forall j \geq 0$ and $\tilde{\Psi}_{j,j'}^{(2)} \geq \tilde{\Psi}_j^{(3)}$ for $\forall j, j' \neq j$.

(i) I first show that condition (i) holds when $h = j$. Note that

$$
\begin{aligned}
& \partial \tilde{s}_j / \partial \delta_j \\
= & \int \tilde{\Psi}_j^{(1)} \left(1 - \tilde{\Psi}_j^{(1)}\right) dP_{\boldsymbol{\varsigma}} (\boldsymbol{\varsigma}; \boldsymbol{\sigma}) + \sum_{j' \neq j} \int \left[\tilde{\Psi}_{j,j'}^{(2)} \left(1 - \tilde{\Psi}_{j,j'}^{(2)}\right) - \tilde{\Psi}_j^{(3)} \left(1 - \tilde{\Psi}_j^{(3)}\right)\right] dP_{\boldsymbol{\varsigma}} (\boldsymbol{\varsigma}; \boldsymbol{\sigma}).
\end{aligned}
$$

In the first summand, $\tilde{\Psi}_j^{(1)} \left(1 - \tilde{\Psi}_j^{(1)}\right) < \tilde{\Psi}_j^{(1)}$. In the second summand, $\left[\tilde{\Psi}_{j,j'}^{(2)} \left(1 - \tilde{\Psi}_{j,j'}^{(2)}\right) - \tilde{\Psi}_j^{(3)} \left(1 - \tilde{\Psi}_j^{(3)}\right)\right] \leq \left[\tilde{\Psi}_{j,j'}^{(2)} - \tilde{\Psi}_j^{(3)}\right]$. Therefore, $\partial \tilde{s}_j / \partial \delta_j < \tilde{s}_j$ and $\partial F_j (\boldsymbol{\delta}) / \partial \delta_j = 1 - (\partial \tilde{s}_j / \partial \delta_j) / \tilde{s}_j > 0$.

For $h \neq j$,

$$
\begin{aligned}
& \partial \tilde{s}_j / \partial \delta_h \\
= & -\int \tilde{\Psi}_j^{(1)} \tilde{\Psi}_h^{(1)} dP_{\boldsymbol{\varsigma}} (\boldsymbol{\varsigma}; \boldsymbol{\sigma}) + \int \tilde{\Psi}_j^{(3)} \tilde{\Psi}_h^{(3)} dP_{\boldsymbol{\varsigma}} (\boldsymbol{\varsigma}; \boldsymbol{\sigma}) \\
& + \sum_{j' \neq j,h} \int \left(-\tilde{\Psi}_{j,j'}^{(2)} \tilde{\Psi}_{h,j'}^{(2)} + \tilde{\Psi}_j^{(3)} \tilde{\Psi}_h^{(3)}\right) dP_{\boldsymbol{\varsigma}} (\boldsymbol{\varsigma}; \boldsymbol{\sigma}) \\
\leq & \sum_{j' \neq j,h} \int \left(-\tilde{\Psi}_{j,j'}^{(2)} \tilde{\Psi}_{h,j'}^{(2)} + \tilde{\Psi}_j^{(3)} \tilde{\Psi}_h^{(3)}\right) dP_{\boldsymbol{\varsigma}} (\boldsymbol{\varsigma}; \boldsymbol{\sigma}) \text{ since } \tilde{\Psi}_j^{(1)} \geq \tilde{\Psi}_j^{(3)} \text{ for } \forall j \\
\leq & \ 0 \text{ since } \tilde{\Psi}_{j,j'}^{(2)} \geq \tilde{\Psi}_j^{(3)} \text{ for } \forall j, j' \neq j.
\end{aligned}
$$

Therefore, $\partial F_j (\boldsymbol{\delta}) / \partial \delta_h = -(\partial \tilde{s}_j / \partial \delta_h) / \tilde{s}_j \geq 0$.

(ii) $\sum_{h=1}^J \partial F_j (\boldsymbol{\delta}) / \partial \delta_h = 1 - \sum_{h=1}^J (\partial \tilde{s}_j / \partial \delta_h) / \tilde{s}_j$ and $\sum_{h=1}^J \frac{\partial \tilde{s}_j (\boldsymbol{\delta})}{\partial \delta_h} = \frac{\partial \tilde{s}_j (\boldsymbol{\delta} + \Delta)}{\partial \Delta}|_{\Delta=0}$.

$$
\begin{aligned}
& \frac{\partial \tilde{s}_j (\boldsymbol{\delta} + \Delta)}{\partial \Delta}|_{\Delta=0} \\
= & \int \left(\tilde{\Psi}_j^{(1)}\right)^2 \frac{1}{e^{\delta_j + \boldsymbol{x}_j \boldsymbol{\varsigma}}} dP_{\boldsymbol{\varsigma}} (\boldsymbol{\varsigma}; \boldsymbol{\sigma}) + \sum_{j' \neq j} \int \left[\left(\tilde{\Psi}_{j,j'}^{(2)}\right)^2 - \left(\tilde{\Psi}_j^{(3)}\right)^2\right] \frac{e^{\kappa}}{e^{\delta_j + \boldsymbol{x}_j \boldsymbol{\varsigma}}} dP_{\boldsymbol{\varsigma}} (\boldsymbol{\varsigma}; \boldsymbol{\sigma}) \\
> & \ 0.
\end{aligned}
$$

Therefore, $\sum_{h=1}^J \partial F_j (\boldsymbol{\delta}) / \partial \delta_h < 1$.

According to BLP, (i) and (ii) implies that $\hat{F}$ is a contraction mapping from $[\underline{\delta}, \bar{\delta}]^J$ to $[\underline{\delta}, \bar{\delta}]^J$. Note that $\left([\underline{\delta}, \bar{\delta}]^J, \|\cdot\|\right)$ is a complete metric space. The contraction mapping $\hat{F}$ therefore has a unique fixed point. Denote it by $\boldsymbol{\delta}^*$. Hence, $\hat{F}(\boldsymbol{\delta}^*) = \min\left(F(\boldsymbol{\delta}^*), \bar{\delta}\right) = \boldsymbol{\delta}^*$. Claim 1 shows that if $\delta_j^* = \bar{\delta}$ for any j, there exists k such that $F_k(\boldsymbol{\delta}^*) \neq \delta_\kappa^*$. Moreover, the proof of Claim 1 implies that $F_k(\boldsymbol{\delta}^*) < \delta_\kappa^*$. Therefore, $\hat{F}_k(\boldsymbol{\delta}^*) = F_k(\boldsymbol{\delta}^*) < \delta_\kappa^*$, which is a contradiction to $\boldsymbol{\delta}^*$ being a fixed point of $\hat{F}$. So, the unique fixed point of $\hat{F}$ cannot be on the bound, and hence it is also the unique fixed point of $F(\delta)$ on $[\underline{\delta}, \bar{\delta}]^J$. $\square$

Claim 1 implies that the unique fixed point in claim 2 is also the unique fixed point of $F : R^J \to R^J$. This completes the proof of the theorem.

B Jacobian of the Equilibrium Price Functions

The jacobian of the equilibrium price function is obtained by total differentiation of the two optimal pricing conditions. To see the details, combine the optimality conditions (18) for all $m \in \mathcal{M}$ as follows.

$$\tilde{\boldsymbol{q}} + \left(\frac{\partial \tilde{\boldsymbol{q}}'}{\partial \boldsymbol{p}}\right)^{\#} \left(\boldsymbol{p} - \widetilde{\boldsymbol{mc}}^{(q)} + \mu_1 + \mu_2 \boldsymbol{q}\right) - \frac{1}{\lambda_2}\left(\frac{\partial \tilde{\boldsymbol{a}}'}{\partial \boldsymbol{p}}\right)^{\#} \boldsymbol{r} = 0,$$

where the (j,k)-element of $\left(\frac{\partial \tilde{\boldsymbol{q}}'}{\partial \boldsymbol{p}}\right)^{\#}$ is $\frac{\partial \tilde{q}_k}{\partial p_j}$, if newspaper j and k are owned by the same newspaper publisher and 0 otherwise. In other words, notation $(X)^{\#}$ represents the element-wise multiple of matrix X and a dummy matrix defined by ownership. When the optimal display advertising rate $\tilde{r}_j\left(q_j, \zeta_j\right)$ defined in [RFOC](17) is plugged in, the above equation can be rewritten as

$$\begin{aligned}
&\tilde{\boldsymbol{m}}\left(\boldsymbol{\delta}, \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\omega}\right) = \tilde{\boldsymbol{m}}_1\left(\boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p}, \boldsymbol{\xi}\right) + \tilde{\boldsymbol{m}}_2\left(\boldsymbol{\delta}, \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\omega}\right) = 0 \\
\text{where}\quad &\tilde{\boldsymbol{m}}_1\left(\boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p}, \boldsymbol{\xi}\right) = \left(\frac{\partial \tilde{\boldsymbol{q}}'}{\partial \boldsymbol{p}}\right)^{\#} \boldsymbol{p} \\
&\tilde{\boldsymbol{m}}_2\left(\boldsymbol{\delta}, \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\omega}\right) = \tilde{\boldsymbol{q}} - \left(\frac{\partial \tilde{\boldsymbol{q}}'}{\partial \boldsymbol{p}}\right)^{\#} \left(\widetilde{\boldsymbol{mc}}^{(q)} - \mu_1 - \mu_2 \boldsymbol{q}\right) - \frac{1}{\lambda_2}\left(\frac{\partial \tilde{\boldsymbol{a}}'}{\partial \boldsymbol{p}}\right)^{\#} \tilde{\boldsymbol{r}}.
\end{aligned}$$

Note that $\boldsymbol{p}$ enters the second term only through $\boldsymbol{\delta}$. Total differentiation with respect to the k^{th} dimension of the characteristics gives

$$\left[\alpha \frac{\partial \tilde{\boldsymbol{m}}_1}{\partial \boldsymbol{\delta}'} + \left(\frac{\partial \tilde{\boldsymbol{q}}'}{\partial \boldsymbol{p}}\right)^{\#}\right] \frac{\partial \tilde{\boldsymbol{p}}^*}{\partial \boldsymbol{x}_k'} + \left(\beta_k \frac{\partial \tilde{\boldsymbol{m}}_1}{\partial \boldsymbol{\delta}'} + \frac{\partial \tilde{\boldsymbol{m}}_1}{\partial \boldsymbol{x}_k'}\right) + \alpha \frac{\partial \tilde{\boldsymbol{m}}_2}{\partial \boldsymbol{\delta}'} \frac{\partial \tilde{\boldsymbol{p}}^*}{\partial \boldsymbol{x}_k'} + \left(\beta_k \frac{\partial \tilde{\boldsymbol{m}}_2}{\partial \boldsymbol{\delta}'} + \frac{\partial \tilde{\boldsymbol{m}}_2}{\partial \boldsymbol{x}_k'}\right) = 0$$

Therefore, the gradient of the equilibrium price function is

$$\frac{\partial \tilde{\boldsymbol{p}}^*}{\partial \boldsymbol{x}_k'} = -\left[\alpha \frac{\partial \tilde{\boldsymbol{m}}}{\partial \boldsymbol{\delta}'} + \left(\frac{\partial \tilde{\boldsymbol{q}}'}{\partial \boldsymbol{p}}\right)^{\#}\right]^{-1} \left[\beta_k \frac{\partial \tilde{\boldsymbol{m}}}{\partial \boldsymbol{\delta}'} + \frac{\partial \tilde{\boldsymbol{m}}}{\partial \boldsymbol{x}_k'}\right]. \tag{21}$$

This expression has an intuitive explanation. Suppose that the term $\left(\frac{\partial \tilde{\boldsymbol{q}}'}{\partial \boldsymbol{p}}\right)^{\#}$ did not exist and there were no reader heterogeneity, i.e. characteristics affected the system only through the mean utility level $\boldsymbol{\delta}$, and $\frac{\partial \tilde{\boldsymbol{m}}}{\partial \boldsymbol{x}_k'} = 0$. Then, the Jacobian $\frac{\partial \tilde{\boldsymbol{p}}^*}{\partial \boldsymbol{x}_k'}$ would be $-\frac{\beta_k}{\alpha} I$, where I is an identity matrix . This is because when all readers have the same taste, a combination of an increase in characteristic x_{kj} by Δ and an increase in price p_j by $\frac{\beta_k}{|\alpha|}\Delta$ has no impact on utility, hence circulation. Therefore, $\frac{\partial \tilde{p}_j^*}{\partial x_{kj}}$ would be $\frac{\beta_k}{|\alpha|}$ if the pricing strategy of j's publisher were to keep the circulation at a certain level. And for j's opponents, nothing has changed as this combined change of x_{kj} and p_j leaves a reader's utility from newspaper j unchanged. However, a rational price setter can do better. Her objective is to maximize profit instead of keeping up with a circulation level. After an increase in x_{kj} by Δ, the publisher can raise the price by more than $\frac{\beta_k}{|\alpha|}\Delta$. It can keep on increasing the price until the downward-sloping newspaper demand curve determines that any marginal increase in the price will decrease the profit. So, the newspaper publisher does take into account the slope of the demand function $\frac{\partial \tilde{q}_j}{\partial p_j}$. When the newspaper publisher has multiple newspapers, it also considers the cross-effect among its newspapers, hence, the term $\left(\frac{\partial \tilde{\boldsymbol{q}}'}{\partial \boldsymbol{p}}\right)^{\#}$ in the expression.

The Jacobian of the equilibrium advertising rate function is

$$\frac{\partial \tilde{r}_h^*}{\partial x_{kj}} = \frac{\gamma_3}{1 + 1/\lambda_2} \left(\sum_{j' \in \mathcal{J}} \frac{\partial \tilde{q}_h}{\partial p_{j'}} \frac{\partial \tilde{p}_{j'}^*}{\partial x_{kj}} + \frac{\partial \tilde{q}_h}{\partial x_{kj}} \right). \tag{22}$$

C Data Sources and Definition of Variables

Demand Data on county circulation for newspapers that are members of the Audit Bureau of Circulation (ABC) is from the *County Circulation Report* of ABC. ABC members account for about 2/3 of all daily newspapers in the U.S. For non-ABC members, county circulation figures are from newspapers' sworn postal statements available in *SRDS circulation.* Display advertising linage data is available for 485 newspaper/years between 1999 and 2005. The data comes from *TNS Media Intelligence.*

Prices Data on annual subscription prices and display advertising rates is from *Editor and Publisher International Year Book (E&P).* A few newspapers have multiple subscription prices. The local price is used. Display advertising rate is the open inch rate measured in dollars per column inch.[44]

Characteristics A newspaper is described by the following characteristics: (1) news hole, the space of a newspaper that is devoted to news, (2) the strength of the opinion-oriented section of a newspaper, (3) the number of staff-bylined stories in a newspaper, (4) frequency of publication, and (5) edition (morning or evening newspaper).

News hole is the difference between total pages and display advertisements. Data on average pages per issue is from *E&P*. It is defined as the weighted sum of average pages per issue for weekdays and that for Sunday with weights $(\frac{6}{7}, \frac{1}{7})$.

As for the second and third newspaper characteristics, I use the number of staff on opinion sections, such as columnists and editorial editors, and the number of reporters as proxies. Data on these variables is collected from *Bacon's Newspaper Directory.* Bacon's Directory provides information on the titles, for example "Business Reporter", and names of all managing and editorial staff for all daily newspapers in the U.S. For each newspaper, I collect the name of all reporters and assign a weight to each one of them. The weight is the inverse of the number of titles that this person has. I then sum up the weights to get "the number of reporters". For example, if a person is a reporter and has only one title, she is counted as 1. If she is a court reporter and a crime reporter, she is counted as 1 as well. But if she holds some managing job at the same time and has therefore another entry in the directory, she contributes to 2/3 in "the number of reporters". The number of columnists and editorial editors are similarly defined.

Data on frequency of publication and edition (morning or evening newspaper) is from *E&P*.

Another factor that influences a reader's utility from a newspaper is the distance of her county to the newspaper's head county. Information on the head county of a newspaper is gathered from *E&P*. The distance of two counties is computed based on the data of latitude and longitude of

[44] Therefore, price discrimination in both subscription prices and advertising rates is ignored, albeit for different reasons. I ignore the price discrimination in newspaper prices because most newspapers offer only one price. There is not enough data variation to identify the difference in demand or marginal cost across geographic areas. I ignore price discrimination in advertising rates because of data limitations.

county centers provided by the Census Bureau.

The data source and the description of the variables are summarized in Table 12.

Table 12: Data Description and Source

	Variable	Data Description	Data Source
Newspaper Demand	q_{jct}	County circulation	ABC, SRDS
Display advertising Demand	a_{jt}	Annual Display Advertising linage (column inch)	TNS
Price of Newspaper	p_{jt}	Annual Subscription Price (1997 $)	E&P
Price of Display Advertising	r_{jt}	Adverting Rate (1997 $/column inch)	E&P
Newspaper Characteristics	x_{2jt}[45]	Weighted sum of reporters and correspondents	Bacon
	x_{3jt}	Weighted sum of columnists and editorial editors	Bacon
	f_{jt}	Frequency of publication (issues/52 week)	E&P
	y_{2jt}	Edition (morning or evening)	E&P
	n_{jt}	Average pages per issue	E&P
County Distance	y_{3jct}	Distance between county c and the head county of newspaper j (100km)	E&P, Census
Owner		Publisher	Bacon
County Demographics	D_{2c}[46]	% of population over 25 with bachelor's degree or higher	Census
	D_{3c}	Median income (1997 $)	Census
	D_{4c}	Median age	Census
	D_{5c}	% of urban population	Census
	D_{6ct}	Number of households	ABC

ABC: County Circulation Report by Audit Bureau of Circulation

Bacon: Bacon's Newspaper Directory

E&P: Editor and Publisher International Year Book

SRDS: SRDS Circulation

TNS: TNS Media Intelligence

[45] Recall that news hole x_{1jt} is not observable. $x_{1jt} = n_j f_j - a_j$.

[46] Recall that $D_{1jt} = 1$.

D Explanation of the Results in Section 6.2

In this appendix, I provide a detailed explanation for observations (2), (3) and (4) in Section 6.2.

I first show that the advertising profit function is convex in circulation according to the estimates. Advertising profit is the sum of display advertising profit $\left(mc_j^{(a)} - r_j\right) e^{\eta_j} q_j^{\lambda_1} r_j^{\lambda_2}$ and the preprint profit $\mu_1 q_j + \frac{1}{2}\mu_2 q_j^2$. Even though the preprint function is concave in circulation ($\hat{\mu}_2 < 0$), its second order derivative is positive for all newspapers in the sample at the estimates. Note that the estimated elasticity of display advertising demand with respect to circulation is larger than 1 ($\hat{\lambda}_1 > 1$). Since the circulation profit $\left(p_j - mc_j^{(q)}\right) q_j$ is linear in q_j, the overall profit function is also convex in circulation.

This convexity has two implications. First, because the marginal advertising value of circulation is larger for larger newspapers, a multi-newspaper publisher has an incentive to shift the circulation from its smaller newspapers to larger newspapers. Second, a newspaper has an incentive to increase its circulation when a decrease in quality or an increase in the price of its competitors leads to an increase in its circulation.

Observation (4) is a direct result of the second implication. As *Star* and *Pioneer* reduce their overall quality, the *Stillwater Gazette*, a local newspaper competing with them in Washington County, has an incentive to increase its circulation. Therefore, it decreases its price and improves its quality to achieve that. For example, its incentive for decreasing its price can be seen from the first order condition with respect to price: $q_j + \left(p_j + \frac{\partial \pi_j^{(a)}}{\partial q_j}\right) \frac{\partial q_j}{\partial p_j} = 0$, where $\pi_j^{(a)}$ is the overall advertising profit function. When $\frac{\partial \pi_j^{(a)}}{\partial q_j}$ increases, it is optimal to decrease price p_j.

I now explain observations (2) and (3) on the two merged newspapers *Star* and *Pioneer*. Recall that *Star* is the larger party. Two forces affect McClatchy's decision on the quality and prices of these two newspapers. The first force is a quality cross-effect. For example, when *Star* increases its number of reporters, the circulation of *Pioneer* falls and hence the profit of *Pioneer*. The other force is the concern of leaving space in the old quality region for competitors to shift their newspapers there and compete for readers in order to attract advertisers as well.

To understand the quality cross-effect, I plot the profit functions of *Star* and *Pioneer* as well as the sum of these two profit functions around the original equilibrium (i.e. pre-merger equilibrium) in Figure 7 and Figure 8. For example, in the left graph of Figure 7, the x-axis represents log(1+news hole) of *Star*, and the y-axis is profit in million and (π_1, π_2) represents the profit from *Star* and *Pioneer*, respectively.[47] Profit is plotted as a function of *Star*'s news hole when the other dimensions of *Star*'s quality measures as well as the quality of other newspapers are fixed at the pre-merger equilibrium. In contrast, prices are allowed to fully adjust to a second-stage equilibrium. In

[47]For presentational convenience and because only the shapes of the profit curves are relevant for the arguments, I adjust the location of these curves.

the middle and the right graph, profits are plotted as the number of opinion section staff and reporters of *Star* vary. Figure 8 shows how the profits of *Star* (π_1) and *Pioneer* (π_2) change as the characteristics of *Pioneer* (in contrast to *Star* in Figure 7) vary.

Figure 7: Profit of Star (π_1) and Pioneer (π_2) as the Characteristics of *Star* Vary

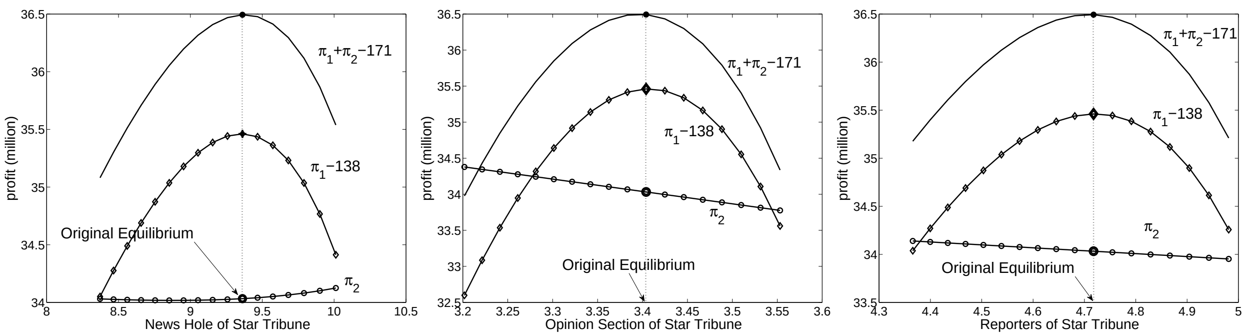
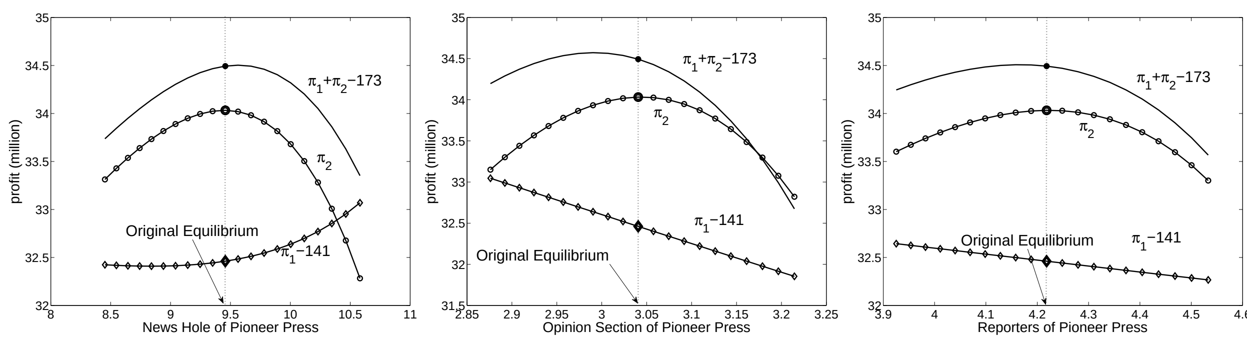


Figure 8: Profit of Star (π_1) and Pioneer (π_1) As the Characteristics of *Pioneer* Vary



The quality cross-effects are shown in the π_2 curve in Figure 7 and the π_1 curve in Figure 8: how the profit from *Pioneer* varies as the quality characteristics of *Star* change, and vice versa. The cross-effect curves are downward-sloping in the middle and right graphs of both figures, implying that as one newspaper increases its quality in terms of the number of staff for the opinion-oriented section or the number of reporters, the other newspaper's profit falls. However, the left graphs in both figures show that the sign of the cross-effect of news hole is not determinate. As mentioned in the intuition for observation (2) in Section 6.2, news hole also affects the marginal cost of increasing circulation. Increasing news hole therefore leads to a higher marginal cost $mc^{(q)}$ and hence does not always affect the other newspaper's profit adversely when prices are fully adjusted to an equilibrium. In particular, at the original equilibrium, the cross-effect with respect to news hole is positive. This is consistent with the observation that news hole of both newspapers increases after the merger while the other dimensions of newspaper characteristics decline.

The asymmetric incentive to adjust quality for *Star* and *Pioneer* can be seen from comparing the original equilibrium quality characteristics to the argumentum of the total profit function $\pi_1 + \pi_2$. After the merger, McClatchy internalizes the cross-effect of *Star* and *Pioneer* and maximizes the total profit, the maximum of which is close to the original equilibrium in Figure 7 but noticeably different in Figure 8. For example, the argumentum of $\pi_1 + \pi_2$ in the middle graph of Figure 8 is left to the original equilibrium point, indicating that McClatchy can increase its profit by decreasing the number of staff for the opinion-oriented section in *Pioneer* when other characteristics of *Pioneer* and the quality of all other newspapers are fixed. In contrast, McClatchy does not have such a strong incentive to adjust the quality of *Star*, the larger party, as indicated in Figure 7. This difference is due to the convexity of the advertising profit function in circulation. A marginal improvement in the quality of *Pioneer* leads to a reduction in *Star*'s circulation, which decreases the profit of *Star* by a larger margin than the marginal effect of an increase in *Star*'s quality on *Pioneer*'s profit. In other words, the cross-effect of *Star*'s quality on *Pioneer*'s profit is smaller than vice versa.[48]

I have shown that the adjustment under the quality cross-effect only is consistent with the full adjustment at the new equilibrium. This holds for both the direction of adjustment and the asymmetry in adjustment, which means that the cross-effect essentially determines the full equilibrium outcome. This explains observation (2) and (3).

[48] The same argument applies to the asymmetry of the price cross-effect, which explains the difference in the price adjustments in Table 7.