Estimation of cost efficiencies from mergers: Application to U.S. radio *

Przemysław Jeziorski †

March 11, 2014

Abstract

This article develops means to estimate fixed-cost efficiencies from mergers. The estimates might be used to assess the total welfare impact of retrospective and counterfactual mergers. The procedure uses a structural model in which companies play a dynamic game with endogenous mergers and product-repositioning decisions. This formulation corrects for a sample selection of more profitable mergers and captures follow-up mergers as well as post-merger product repositionings. The basic idea behind the estimator is to treat mergers as endogenous, thereby allowing for a comparison between mergers observed in the data and counterfactual ones, based on simulated long-run gains for different levels of cost efficiencies. The framework is applied to estimate cost efficiencies after the 1996 deregulation of U.S. radio. I find that average yearly cost savings from mergers within the 1996-2006 time period amount to about $1.2 billion per year (equally split across economies of scale and within-format cost synergies).

*I would like to thank: Lanier Benkard and Peter Reiss, Ronald Goettler, Gautam Gowrisankaran, Joseph Harrington, Elena Krasnokutskaya, Ilya Segal, Alan Sorensen, Andrew Sweeting, Benjamin Van Roy, Miguel Villas-Boas, and Ali Yurukoglu, as well as participants at numerous seminars.

†Haas School of Business, UC Berkeley.

1 Introduction

Economic theory argues that horizontal mergers[1] can affect the performance of markets through increases in market power and supply-side efficiencies. When market power and supply-side efficiencies coexist, the net impact of mergers on welfare is ambiguous (see Williamson (1968)), so an antitrust regulator should empirically evaluate cost savings, in addition to measuring the decrease in competition. If the empirical estimates of cost savings are available, the regulator may compute the retrospective impact of past mergers on total surplus, as well as assess the total welfare impact of counterfactual mergers. However, the empirical literature on cost efficiencies of mergers is scarce because the reliable cost data are rarely available, and as a consequence, the natural estimator of cost efficiencies, which compares cost before and after the merger, is usually infeasible. This study provides an alternative method to assess cost efficiencies of mergers that is based on revealed preferences of firms, and is applicable when little or no cost data are obtainable. This method utilizes estimates of extra revenues generated by mergers, and provides the level of cost efficiencies that rationalizes the merger decisions in the data. In practice, I use a dynamic model with endogenous mergers to generate a set of inequalities bounding the level of cost synergies. On the one hand, when the model predicts a merger but the data do not show one, I infer that the presumed cost efficiencies are too large. On the other hand, when the model predicts no merger, but the data indicate one, I infer the presumed cost efficiencies are too small.

Implementing the proposed cost estimator requires robust long-run predictions of gains from mergers, which are obtained using a dynamic model with endogenous mergers and product characteristics. In contrast, previous empirical work analyzes mergers in a static framework and treats market structure as given (see Nevo (2000), Pinkse and Slade (2004), and Ivaldi and Verboven (2005)). Such static models are useful in addressing the short-run impacts of mergers but do not account for resulting long-run changes in the market structure. Benkard, Bodoh-Creed, and Lazarev (2008) evaluate a longer-run effect of a merger on market structure but still treat it as an exogenous one-time event. The proposed dynamic framework builds on the above methods accounting for dynamic processes such as self-selection of mergers, follow-up mergers leading to merger waves, and post-merger product repositioning.

[1] In this article, I use the terms merger and acquisition interchangeably to mean any change of ownership of a part of or a whole company.

Modeling and estimating models with endogenous mergers pose econometric and computational challenges. To evaluate a potential merger, both acquirer and acquiree must take into account the ownership structure and characteristics of all active products. Because the number of such variables is usually large, one has to deal with the curse of dimensionality, which increases data requirements and poses computational challenges. In this article, I overcome these issues by using a data set on thousands of mergers within one industry, and by applying recent advancements in the estimation of dynamic games (see Bajari, Benkard, and Levin (2007)). Moreover, modeling of mergers in a dynamic framework introduces several conceptual issues including simultaneous merger bids for a single product and multi-product bids by a single acquirer. This study addresses the former issue by modeling players' moves as sequential with bigger owners moving first, and the latter issue by approximating multi-product mergers with a series of highly correlated product-by-product acquisitions. The degree of correlation is estimated from the data and reflects the amount of common information used across the decisions.

I subsequently apply the model to analyze ownership consolidation in the U.S. radio industry. The Telecommunications Act of 1996 increased local-market radio station ownership caps, triggering an unprecedented merger wave that eliminated many small and independent radio owners. From 1996 to 2006, the average Herfindahl-Hirschman Index (HHI) in local radio markets grew from 0.18 to 0.26, the average number of owners in the market dropped from 16.6 to 12.4, and the average number of stations owned grew from 1.6 to 2.3. Such dramatic changes to the market structure have raised concerns about anti-competitive aspects of the deregulation (Leeper (1999), Drushel (1998), Klein (1997)). After estimating the model, I find that the main incentives to merge in the radio industry come from the cost side. Total cost-side savings amount to \$1.2 billion per year, constituting about 6% of total industry revenue. Such cost efficiencies are higher than the anti-competitive effects of these mergers, as identified by Jeziorski (2013). I can disaggregate these cost efficiencies further into economies of scale and within-format cost synergies. The economies of scale bring roughly 50% of total cost synergies, amounting to \$0.6b per year. The fact that consolidation leads to substantial cost-side efficiencies allows us to conclude that the Telecom Act enabled radio to better compete against other media, such as TV or the Internet.

The closest article to this work is Stahl (2010), who analyzes cross-market cost efficiencies from common ownership in the TV industry in the absence of market power incentives to merge.

Another article by O'Gorman and Smith (2008) uses a static oligopoly model to estimate the cost curve in radio, and find the fixed cost savings, when the company owns two stations, are bounded between 20% and 50% of per-station costs. (I estimate this number to be 60%.) One advantage of the approach presented in this article is that it is dynamic and controls for endogenous repositioning, which allows me to separate within-format cost synergies from economies of scale. Beyond estimating cost synergies, this study incorporates the impact of ownership concentration on product variety (see Berry and Waldfogel (2001)) by utilizing a joint model of repositioning and merger decisions. The estimation approach in this study builds on the empirical literature on demand and cost curve estimation (see Rosse (1967) and Rosse (1970)) by accounting explicitly for the demand- and supply-side incentives to merge. Finally, this study contributes to the static literature on determinants of mergers, such as Akkus and Hortacsu (2007) and Park (2011), by directly acknowledging that mergers are dynamic decisions. My model shares some similarities with Gowrisankaran (1999) and Sweeting (2011). The former contains numerical analysis of endogenous mergers with homogeneous products, while the latter evaluates the impact of music fees on endogenous product repositioning, without modeling mergers.

This article is organized as follows. Section 2 contains industry and data descriptions. Section 3 presents the model. Section 4 describes the estimations procedure. Section 5 presents the results. Section 6 concludes.

2 Industry and data description

Radio is an important medium in the United States, reaching about 94% of Americans aged 12 years and older. Moreover, the average consumer listens to about 20 hours of radio per week, and between the hours of 6am and 6pm, more people use radio than TV or print media (see A.Richter (2006)). Approximately 13,000 commercial radio stations broadcast in about 350 strictly defined Arbitron markets nationwide. Each station is characterized by a format that summarizes the type of programming; that is, the format includes information about type of music, the number of news and talk shows, as well as information about being inactive (DARK format). Specialized consulting companies assign and monitor the formats. The data on formats are released quarterly and reflect possible product repositioning in the form of format switching. Before 1996, this industry had

# of active stations	45+		30-44		15-29		0-14	
	Old cap	New cap	Old cap	New cap	Old cap	New cap	Old cap	New cap
	4	8	4	7	4	6	3	5

Table 1: Change in local ownership caps introduced by the 1996 Telecom Act.

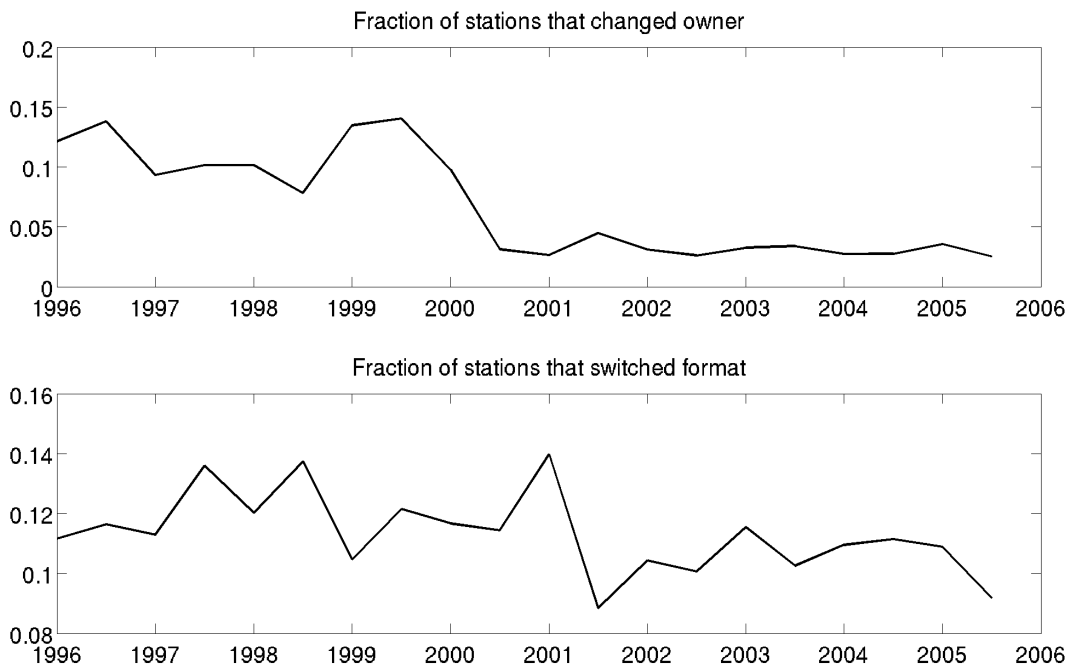

Figure 1: Dynamics of station acquisition and format switching. Source: BIA Inc.

ownership limitations, both nationally and locally, preventing big corporations from entering the market and thereby sustaining a large degree of family-based ownership. This situation changed with the Telecom Act of 1996 which, among other things, raised the ownership caps in local markets (see Table 1). This overhaul of the ownership restrictions triggered an unprecedented merger and product-repositioning wave that completely reshaped the radio industry. In the first week after the Act was passed, radio station owners closed nearly $700 million in merger deals (see Bednarski (2002)). Figure 1 contains the average percentage of stations that switched owners and formats. Between 1996 and 2000, more than 10% of stations switched owners, annually. After 2000, the number dropped to less than 4%. Greater ownership concentration in the 1996-2000 period was also associated with more format switching. The percentage of stations that switched formats peaked in 1998 and 2001 at 13%. In effect, the Herfindahl-Hirschman Index (HHI) in the listenership market grew from 0.18 in 1996 to about 0.3.

According to the Radio Advertising Bureau, radio industry revenues grew from $11 billion in 1994 to nearly $20 billion in 2000. In 2001, advertising budgets were cut across all media, which resulted in an 8% decline in radio revenue. However, since 2001, the industry has posted steady single-digit yearly increases in revenue and a steady, over 80%, listener share. Previous studies have examined the impact this revenue expansion on listener and advertiser surplus. In particular, Jeziorski (2013) finds that the consolidation of ownership harmed advertisers, causing deadweight loss and yearly $223 million decrease in advertiser surplus, but benefited listeners, raising listener welfare by 0.2%.[2] However, relatively little research has examined the cost implications on the Act.

The 1998 and 2004 Occupational Outlook Handbooks by U.S. Department of Labor acknowledge cost efficiencies from consolidation and state that "a network can run eight radio stations from one office, producing news programming at one station and then using the programming for broadcast from other stations, thus eliminating the need for multiple news staffs. Similarly, technical workers, upper level management, and marketing and ad sales workers are pooled to work for several stations simultaneously." Specifically, the U.S. Bureau of Labor Statistics quotes the 20,000 decline in total employment in broadcasting on a 3-digit level from 1996 to 2002. Also, between 1996 and 2006, the industry displaced more than 10,000 announcers, about half of which were in radio; see an on-line appendix for more detailed description of an employment dynamics. Beyond these aggregate numbers there are numerous case studies that indicate cost efficiencies. For example, in 2009, CBS Radio Chicago combined local ad sales force units, quoting efficiencies of selling the ads together.[3] Another example is the high valuation of the price of Citadel, acquired by one of dominant players Cumulus,[4] which was estimated at 3.2 times its yearly revenue, which amounts 27 times the cash flows, using a 12% median net industry margin.

Another source of cost efficiencies was the restructuring of news production. According to a 2003 survey by the Radio Television Digital News Association, more than 95% of radio news

[2]Quantifying consumer surplus in dollar terms is difficult because the radio programming is provided free of charge.

[3]source: `http://blogs.suntimes.com/media/2009/08/cbs_radio_chicago_merges_local.html`

[4]Reported by the New York Times, "In Pandoras Valuation, a Few Sour Notes," published June 5, 2011. The full text is available from the author.

departments handle the news for more than one station with an average news department producing news for three local stations. Despite the fact that newsrooms have to serve more local stations, their average full-time staff is shrinking. Namely, from 2003 to 2010, an average news room shrank from 3.5 to 2.5 full-time staff. In addition to shrinking numbers, 80% of news directors report having other responsibilities: 18% report being talk show hosts, 16% are program directors, and 11.2% do some announcing at local stations, including sports and weather.

The above evidence points to at least three types of cost efficiencies, namely, (i) within-market economies of scale, (ii) within-format cost efficiencies, and (iii) cross-market synergies. In this paper, I am able to separately identify the first two, and leave the third one for future research.

2.1 Data

To analyze the supply-side effects of the consolidation in radio, I compiled a data set on stations in the 68 markets. I dropped the overlapping markets in a way following Sweeting (2011), that is, those "where more than 6% of listening was to stations based in other markets in order to avoid modeling cross-market interactions." I also dropped markets that do not have data on advertising prices. Furthermore, I drop 20 markets for which I could not compute a static equilibrium in a reasonable amount of time for some states along the simulation path.[5] For the purpose of the estimation, the markets are further categorized by population size; that is, markets with population more than 2.5m (13 markets), 1m-2.5m (22 markets), 0.5m-1m (20 markets), and less than 0.5m (13 markets).

BIA Financial Network Inc. provides a database of merger transactions for all stations in the U.S. radio market. From these transactions, one can infer the ownership of each station. BIA also supplies formats of stations between 1996 and 2006. Additionally, I use the estimates of station quality that are obtained using the procedure described in the online appendix. The data contain

[5]These computational problems might be due to existence issues or convergence to local minima in the best response function. The issue is related to the two-sidedness of the market and the fact that I impose non-negativity restriction on ad quantity, which is sometimes binding if the owner has multiple stations. I find that including some of the 20 markets for which I was able to simulate the value function does not change the results in a meaningful way. However, because of long computation time, these extra markets cannot be included in the bootstrap; thus I do not use them in the final estimation.

the formats and market shares measured every six months and revenues measured yearly. To obtain a half-year sample, I assume the revenues are spread evenly across both half-years.

For the purpose of this paper, I interpret a merger activity observed in a BIA data set in a particular way. I treat mergers involving many markets as independent market-by-market deals. Mergers involving many stations within one market are broken down as a series of highly correlated individual decisions. As a result, the model allows an owner to acquire only a part of another company (e.g., in case a full merger violates the ownership cap). Moreover, divestitures are treated as regular sales and I do not include entry of new owners. Omitting entry is equivalent to assuming players' beliefs, when merging, are consistent with no entry of new owners. I handle entry of radio stations through repositioning from an inactive state.

Because many stations were not purchased with cash, I observe an acquisition price for about 40% of the deals. Part of the remaining 60% either spanned across multiple stations and the individual prices were not specified, or payment was made in ways other than cash (station swaps, other equity, or debt transfers). The data set contains some information about these transactions; nevertheless, extracting exact station prices from it is difficult. As a result, I use only a subset of acquisition deals to estimate a pricing equation.

In the case of radio, BIA Inc. tracked 6,685 station acquisitions in 297 markets. Recorded acquisitions do not include any transactions that were not finalized or for which the buyer or seller was missing. The data set contains 454 transactions that span across more than one market, and only 21 transactions that span across more than 10 markets. These numbers seem relatively small compared to 6,685 total transactions, and suggest that ignoring cross-market optimization might not be an issue. However, because most of the cross-market mergers are big, the transactions that span more than one market compose 48% of total transactions if weighted by 1998 revenues. On the other hand, if one takes transactions that span more than 10 markets, the number drops to 18%, suggesting the cross-market transactions, while important, primarily matter locally and might be an issue for similar or overlapping markets. To partially address this problem, I select non-overlapping markets.

3 Model

This section presents the dynamic oligopoly model of a radio industry in the spirit of Ericson and Pakes (1995). I model the industry as a dynamic game in which the players are companies holding portfolios of differentiated products (radio stations). The modeling effort emphasizes the actions that change the portfolio of owned stations, specifically, repositioning and acquisitions.

3.1 Industry basics

The radio industry is composed of geographical markets based on stations' overlapping signal contours. Suppose that M markets exit and the payoff-relevant market characteristics at time t for market m are fully characterized by a set of demographic covariates $d^{mt} \in \mathcal{D}$ (demand shifters). In each market m, up to K_m operating firms and up to J_m active stations are present. (To simplify the exposition, I omit the market subscripts in the rest of the paper.) The set of stations is equivalent to a set of available broadcast frequencies. The set of available frequencies rarely changes over time, and is fixed in the remainder of the paper. Each frequency has an assigned owner and might contain active or inactive radio station. Both types of stations can be traded (trades of stations are equivalent to trading frequencies), and the owner can decide to activate an inactive frequency and vice versa.

I assume each product $j \in J$ is characterized by a triple $s_j^t = (f_j^t, \xi_j^t, o_j^t)$, where $o_j^t \in K$ is the owner, $f_j^t \in F$ is a type of broadcast content, called format, and $\xi_j^t \in \Xi$ is a continuous measure of programming quality unobservable to the econometrician. The state of the industry at the beginning of each period is a pair $(s^t, d^t) \in \mathcal{S} \times \mathcal{D}$, where $s^t = \{s_1^t, \ldots, s_J^t\}$.

The variable ξ_j^t contains information about the unobserved quality of programming as well as information on the strength and quality of signal. As mentioned in the data section, the latter tends to be constant over time; therefore, ξ_j^t is likely to be time-persistent. In particular, I model this time correlation as an AR(1) process. Formally,

$$\xi_j^t = \rho \xi_j^{t-1} + \zeta_j^t, \tag{3.1}$$

where ζ_j^t are mean zero independently identically distributed random variables, with an exception that ζ_j^t may have different variance for stations that switch formats or that switch to/from DARK format. Additionally, because radio owners' decisions are unlikely to affect demographic trends I assume d^t to be exogenous and Markov.

I do not make any distributional assumptions on ζ_j^t; however, I do not allow ξ_j^t to be endogenous. For this reason, the impact of mergers on station quality is beyond the scope of this paper. Statistical and economic significance of this assumption is testable; tests are conducted in Section 5.

3.2 Static payoffs and costs

Conditional on the state of the industry (s^t, d^t), each firm k gets a one-shot variable profits $\pi_k(s^t, d^t)$. Additionally, a firm has to pay a per-period fixed cost $F_k^C(s^t)$ to maintain station portfolio $\{j : o_j^t = k\}$. Estimating the properties of F_k^C, in particular, potential cost efficiencies of owning multiple stations, is a central question of this paper. In general, a functional form of a payoff function can be fairly nonrestrictive, and Doraszelski and Satterthwaite (2010) list assumptions that ensure the existence of a pure strategy equilibrium of the dynamic game.

Variable profits of the firm in the radio market have the following general form:

$$\pi_k(s^t, d^t) = \left(\sum_{\{j:o_j^t=k\}} p_j(s^t, d^t, \bar{q}^t) r_j(s^t, d^t, \bar{q}^t) - \mathrm{MC}_j \right) \bar{q}_j^t.$$

$p_j(\cdot)$ is a price per fraction of listenership in a given market (advertising inverse demand) of one ad slot, $r_j(\cdot)$ is a listenership market share (demand for programming), and $\bar{q}^t = (\bar{q}_1^t, \ldots, \bar{q}_J^t)$ is an equilibrium vector of advertising quantities. Term MC_j represents a marginal cost of selling advertising given by $\mathrm{MC}_j = \theta_1^A[\theta^C + \eta_j]$. Term $\theta_1^A\theta^C$ is a mean market-level marginal cost, and $\theta_1^A\eta_j$ is a firm-specific shock. Note that in addition to a marginal cost, selling advertising requires incurring fixed cost, because most of the agents work on incentive contracts with a lower bound on wages. Also, as mentioned in Section 2, running an ad sales department is likely to impose some fixed costs.

The station market share is computed using a logit model with random coefficients, following Berry, Levinsohn, and Pakes (1995). Let $\iota_j = (0, \ldots, 1, \ldots, 0)$, where 1 is placed in a position that indicates the format of station j. Denote the amount of broadcasted advertising minutes in station

j as q_j. For a given consumer i, the utility from listening to a station j is given by

$$u_{ij} = \theta^L_{1i}\iota_j - \theta^L_{2i}q_j + \theta^L_3\mathrm{FM}_j + \xi_j + \epsilon_{ji},$$

where θ^L_{1i} is a set of format fixed effects, θ_{2i} is a disutility of advertising, θ^L_3 is an AM/FM fixed effect, and ϵ_{ji} are idiosyncratic taste shocks that are independently and identically distributed as type-1 extreme value. I assume the random coefficients can be decomposed as

$$\theta^L_{1i} = \theta^L_1 + \Pi D_i + \nu_{1i}, \quad D_i \sim F_m(D_i|d), \quad \nu_{1i} \sim N(0, \Sigma_1)$$

and

$$\theta^L_{2i} = \theta^L_2 + \nu_{2i}, \quad \nu_{2i} \sim N(0, \Sigma_2),$$

where Σ_1 is a diagonal matrix, $F_m(D_i|d)$ is an empirical distribution of demographic characteristics, ν_i is an unobserved taste shock, and Π is the matrix representing the correlation between demographic characteristics and format preferences. I assume draws for ν_i are uncorrelated across time and markets. The market share of the station j is given by the aggregate probability of choosing station j; that is,

$$r_j(q|s,d) = \mathrm{Prob}\big(\{(\nu_i, D_i, \epsilon_{ij}) : u_{ij} \geq u_{ij'}, \text{for } j' = 1, \dots, J\}\big|q, s, d\big). \tag{3.2}$$

The radio station owners are likely to have market power over advertisers. Moreover, because of heavy ad targeting, the stations with different formats are not perfect substitutes. The simplest model that captures these features is a linear inverse demand for advertising. Denote the total quantity of advertising supplied in format f as Q_f, formally, $Q_f = \sum_{j:f_j=f} q_j$. Then the advertising prices per fraction of the market are given by

$$p_j(q) = \theta^A_1 \left(1 - \theta^A_2 \sum_{f'=1}^{F} \omega^m_{f_j f'} Q_{f'}\right), \tag{3.3}$$

where θ^A_1 is a scaling factor for the value of advertising, θ^A_2 is a market-power indicator, and $\omega_{ff'} \in \Omega$ are weights indicating competition closeness between formats f and f'.

Given the advertising-quantity choices of competing owners, each radio station owner k chooses q_j jointly for all stations that he owns to maximize his variable profits; formally,

$$\max_{\{q_j:o^t_j=k\}} \sum_{\{j:o^t_j=k\}} r_j(q|s,d)p_j(q)q_j - \mathrm{MC}_j q_j. \tag{3.4}$$

The market is assumed to be in a Nash equilibrium.[6] Such structure on variable profits is intended to capture payoff interactions between formats present in the data. In particular, a firm can either specialize in a particular format and extract local monopolistic rents, or spatially differentiate to cover the largest possible audience.

The details of the dynamic model are contained in the rest of this section.

3.3 Acquisitions and repositioning

Firms can undertake two types of actions: station acquisitions[7] and station repositioning. Each acquisition of product j by a player k is followed by a cash transfer P_{kj}^t from the buyer to the seller. For the reasons described in the data section, the merger execution costs are likely to be small, and for simplicity, I assume them to be zero. If this assumption is violated, my estimates of fixed-cost efficiencies would contain fixed costs of merging; thus the fixed-cost efficiencies could be underestimated. By contrast, repositioning costs are likely to be substantial because they involve staff turnover as well as extra marketing effort. I assume each repositioning action from f_j^t to f_j^{t+1} incurs a cost $F^S(f_j^t, f_j^{t+1})$.

Firms can potentially acquire any subset of competitors' stations, as well as choose characteristics of owned and newly acquired stations. One option would be to write down the simultaneous-move game and allow for set-valued actions. However, because of certain features of the radio industry described in the remainder of this section, a sequential-move game might be more realistic. Additionally, the sequential formulation offers conceptual and computational simplicity, which is a key to the feasibility of the estimation. Below I describe the timing of the acquisitions- and repositioning-stage game and follow up with a discussion of particular assumptions.

(A0) Acquisition stage starts. Owners receive a right to acquire according to a sequence specified by a permutation $\sigma(s^t, d^t)$ of the active owners' index $\{1, \dots, K\}$. The sequence σ is common knowledge.

(A1) Owner k receives a right to move and becomes a potential buyer. The buyer observes a vector

[6] To simplify the computation of the equilibrium, when simulating the value function I ignore the random shocks to the marginal cost of advertising.

[7] I do not deal directly with entry by acquisition. Under the assumption that cross-market cost synergies exist, entry of new owners through full acquisition of an existing owner is equivalent to relabeling of the name of the owner.

of stochastic one-time additive payoff shocks ϕ_k^t to integrating any competing radio station into the portfolio. The shocks $\phi_{jk}^t \in \mathbb{R}$ to acquiring a particular station j are revealed to k sequentially according to a permutation σ^A of the indexes of stations owned by competitors. The acquisition process proceeds as follows:

(i) Upon observing the shock $\phi_{kj}^t \in \mathbb{R}$ to payoff from acquiring station j, the buyer makes a take-it-or-leave-it offer to the seller $k' = o_j^t$.

(ii) The seller k' decides to accept or reject the offer. The acquisition decision is implemented and revealed to all players.

(iii) The buyer observes the next shock $\phi_{kj'}^t$ and the game proceeds to (i) until all shocks are revealed.

(A2) The next owner receives the right to acquire, and the game moves to (A1). If no more owners are present, the game proceeds to the repositioning stage (R0).

(R0) The repositioning stage starts. Owners receive a right to reposition according to a sequence specified by a permutation $\sigma(s^t, d^t)$

(R1) Owner k receives a right to reposition. He observes a vector of stochastic one-time additive payoff shocks ψ_k^t to repositioning any station to any format. The shocks $\psi_{jk}^t \in \mathbb{R}^F$ to repositioning a particular station j are revealed to k sequentially according to a permutation σ^R of the indexes of stations owned by k. The repositioning process proceeds as follows:

(i) Upon observing the shock $\psi_{kj}^t \in \mathbb{R}^F$ to repositioning of station j to any format, the owner makes a repositioning decision. The decision is implemented and revealed to all players.

(ii) The buyer observes the next shock $\psi_{kj'}^t$ and the game proceeds to (i) until all shocks are revealed.

(P) The new state (s^{t+1}, d^{t+1}) is drawn. Stage payoffs $\pi(s^{t+1}, d^{t+1})$ are realized. The game proceeds to the next period.

Below I motivate each step of the game.

The game is divided into two substages: acquisition and repositioning. It is designed to capture the fact that intentions to merge have to be submitted to the Federal Communication Commission. Thus we can safely presume that firms' decisions about mergers become public fairly quickly and get implemented with a delay. In these circumstances, firms should be able to condition repositioning decisions on the intended mergers. [8]

[8]In case the station was acquired and repositioned in the same period, the sequential formulation always assigns a repositioning action to the new owner.

Step (A1) prescribes endogenous sequence of moves $\sigma(s^t, d^t)$ as function of the industry state, which is intended to rank players by size and generate a Stackelberg type of competition. Specifically, during the estimation, σ ranks players by the last period's listenership size. Ranking by size is motivated by the fact that merger decisions usually involve some analytical and legal work, which is likely to be burdensome for smaller and less experienced players. In particular, big owners are more likely to pick up "low-hanging fruit" acquisitions first. The sequential formulation is also necessary to reduce the number of potential equilibria in the stage game [9] and to solve the problem of multiple companies trying to acquire the same product in the same period. Because the specific ranking of moves might change the results, I try alternative forms of σ and report the results in the online appendix.

Steps (A1) and (R1) prescribe action-specific payoff shocks. The shocks introduce unobserved heterogeneity in payoffs, which rationalize why observationally equivalent companies take different merger and repositioning actions. To lower the data requirements for estimating the model I assume ϕ_j^t and ψ_j^t are independently and identically distributed across time, players, and markets. Note that the model controls for some unobserved heterogeneity by pulling persistent unobserved station quality ξ_j^t into the state. An extension introducing correlation in ϕ_j^t and ψ_j^t is theoretically possible but is not implemented.

The sequential formulation enables me to handle large dimensionality of an action space while maintaining interdependence between multiple decisions by the same player. Controlling for this interdependence is necessary because acquisition and repositioning actions are correlated, for example, an acquisition is positively correlated with acquiring more stations in the same period, and repositioning is negatively correlated with repositioning another station into to the same format. By allowing future decisions to depend on past decisions, I effectively approximate decisions that span multiple stations with a series of highly correlated sequential station-by-station decisions. During estimation, I assume σ^A and σ^R rank stations by ξ_j^t; however, I verify that the results are robust to this choice by examining a random ranking as well (see online appendix).

Assuming buyers make take-it-or-leave-it offers effectively awards them most of the bargaining power. This assumption is consistent with the reality of the radio industry and removes a selection

[9] The game with sequential moves has the unique subgame perfect equilibrium in a stage game. However, it does not guarantee uniqueness in the dynamic game.

problem when estimating acquisition-pricing equations. In particular, radio industry contains small number of large buyers and a large number of potential acquisition targets; thus, large buyers can walk away and make an offer to another seller. Finally, note that the take-it-or-leave-it assumption does not imply that the large players internalize all the gains from mergers, instead the gains from merging are split in a way reminiscent of Rubinstein (1982).

In the next section, I describe the strategies in this game.

3.4 Strategies

I restrict my attention to Markov strategies that are a quadruples g_k consisting of an acquisition strategy g_k^A, a pricing strategy g_k^P, a merger bid-acceptance strategy g_k^B, and a repositioning strategy g_k^R. I define these strategies below.

Let A_{kj}^t be the set of merger decisions implemented by players that moved earlier in the sequence σ^A in the current period, and merger actions already undertaken by player k in the current period. Let $\mathcal{A}$ be a class of all possible sets A_{kj}^t. The acquisition and pricing strategies are mappings from observables to actions

$$g_k^A : \mathcal{S} \times \mathcal{D} \times \mathcal{A} \times J \times \Phi \to \{0, 1\},$$

$$g_k^P : \mathcal{S} \times \mathcal{D} \times \mathcal{A} \times J \times \Phi \to \mathbb{R},$$

where J is the index of an acquisition target and Φ is the support of the payoff shock ϕ_{kj}^t.

The set of feasible strategies is a set of such functions that are measurable with respect to the information (σ-field) generated by a move sequence. Actions of player k can depend on a current state and shocks (s^t, d^t, ϕ_k^t) as well as a vector A_{kj}^t. For any (s, d, j, ϕ), take two actions sets A and A' in $\mathcal{A}$ such that all observable past actions in these action sets are the same. A feasible strategy g_k^A is restricted to prescribe the same action for these sets; that is, $g_k^A(s, d, A, j, \phi) = g_k^A(s, d, A', j, \phi)$. Moreover, one cannot acquire a station that one already owns, so $g_k^A(s, d, A, j, \phi) = 0$ for j such that $o_j^t \neq k$.

The bid acceptance strategy of player k is allowed to depend on observables as well; that is,

$$g_k^B : \mathcal{S} \times \mathcal{D} \times \mathcal{A} \times J \times \mathbb{R} \times K \to \{\text{Accept}, \text{Reject}\},$$

where $\mathbb{R}$ represents the offer and K represents the identity of bidder k'.

Similarly, let R^t_{kj} be a set of merger actions undertaken by all players in the current period, repositioning actions made by players that moved earlier in the sequence σ^R in the current period, and repositioning actions already undertaken by player k in the current period. Let $\mathcal{R}$ be a class of all possible R^t_{kj}. Define a repositioning strategy

$$g^R_k : \mathcal{S} \times \mathcal{D} \times \mathcal{R} \times J \times \Psi \to \{1, \ldots, F\},$$

where Ψ is the support of the shock. The strategies do not need to explicitly depend on acquisition prices, because they are a sunk cost. Also, the stations that k does not own cannot be repositioned by k, and similarly, as in the definition of an acquisition strategy, g^B_k has to be measurable with respect to the information σ-field generated by a move sequence.

3.5 Equilibrium

Let $\mathbf{g} = (g_1, \ldots, g_K)$ be a stationary Markov strategy profile. It can be shown that this profile and an initial condition (s^0, d^0) determine an essentially unique, controlled Markov process $\mathcal{P}$ over states (s^t, d^t), acquisition actions a^t, acquisition prices P^t, bid-acceptance decisions b^t, repositioning actions r^t, and payoff shocks (ψ^t, ϕ^t). This process is supplied with a natural filtration such that $\mathbf{g}$ is adapted to it.

Given the realizations of $(s^t, s^{t+1}, d^{t+1}, P^t, \psi^t, \phi^t)$, the per-period payoff for player k is given by the equation

$$\begin{aligned} \Pi_k(s^t, s^{t+1}, d^{t+1}, P^t, \psi^t, \phi^t) = \pi_k(s^{t+1}, d^{t+1}) - F^C_k(s^{t+1}) + \sum_{j: o^t_j \neq k, o^{t+1}_j = k} (\phi^t_{kj} - P^t_{kj}) + \\ + \sum_{j: o^t_j = k, o^{t+1}_j \neq k} P^t_{o^{t+1}_j j} + \sum_{j: o^{t+1}_j = k} \left[\psi^t_{kjf^{t+1}_j} - \mathbf{1}(f^{t+1}_j \neq f^t_j) F^S(f^t_j, f^{t+1}_j) \right]. \end{aligned} \tag{3.5}$$

The third term of the above equation represents outgoing cash flows resulting from acquisitions, the fourth term represents incoming cash flows from selling stations, and the last term represents cash flows from repositioning. Note that the per-period payoff Π_k is not the same as variable profits π_k.

Each owner is maximizing the expected discounted sum of profits, taking the strategies of opponents $\mathbf{g}_{-k}$ as given. The value function for player k is defined as

$$V_k(s, d | \mathbf{g}_k, \mathbf{g}_{-k}) = E_{\mathcal{P}(\mathbf{g},s,d)} \sum_{t=0}^{\infty} \beta^t \Pi_k(s^t, s^{t+1}, d^{t+1}, P^t, \psi^t, \phi^t). \tag{3.6}$$

I assume the markets are in a Markov Perfect Equilibrium; that is, firms choose a strategy profile $\mathbf{g}^*$ such that for all k,

$$V_k(s, d|\mathbf{g}_k^*, \mathbf{g}_{-k}^*) \geq V_k(s, d|\mathbf{g}_k, \mathbf{g}_{-k}^*) \quad \forall \mathbf{g}_k. \tag{3.7}$$

For simplicity, I restrict attention to symmetric equilibria.

The bargaining process of this game allows me to further restrict my attention to MPEs in which all merger offers are accepted in the equilibrium. The seller does not have private information; thus the buyer makes offers equal to the seller's continuation value conditional on rejecting the merger. For this reason, Markov Perfect Equilibrium prices P_{kj}^t depend only on payoff-relevant variables of seller k', that is, the state of the game before k makes an offer to j denoted by (s^t, d^t, A_{kj}^t). As described in the next section, this feature is convenient for estimation because the equilibrium pricing function can be pre-estimated in the first stage.[10] Moreover, the acceptance strategy does not have to be estimated.

3.6 Cross-market decisions

Section 2 presents the anecdotal and survey evidence that points to the existence cost efficiencies which are predominantly local. Also, because the majority of ad sales are local, the market power is not likely to cross the boundary of geographical markets. For this reason, the above model has only a limited amount of across-market correlation in merger decisions. In reality, the owners decide which stations to acquire in every market, taking into account structure and demographic trends within the current market. However, because demographic transitions that represent trends in radio listening and profitability, are correlated across markets and incorporate national trends, mergers across markets would also be correlated.

4 Estimation

The estimation of the dynamic model is preceded by a static estimation of the advertising game. The details of the static estimation can be found in the online appendix. This estimation provides (i) profit function $\pi_k(s, d)$ for any owner and industry configuration and (ii) unobserved quality ξ_j^t

[10] The way in which the model is estimated allows some departures from the take-it-or-leave-it assumption as long as the price is only the function of the payoff-relevant states.

for each radio station at each point in the data, along with parameters of equation (3.1) and the distribution of ζ_j^t. In the remainder of this section, I assume the variable profit function π, the quality ξ_j^t, and the distribution of ζ_j^t have already been recovered. However, when I compute final standard errors, I still account for the fact that they were pre-estimated.

The data used in the estimation of the dynamic model are a set $X = \{x^{tm} : 1 \leq m \leq M, 1 \leq t \leq T\}$. Each point x^{tm} describes the state of the industry at the beginning of the period: $s^{tm} = (f^{tm}, \xi^{tm}, o^{tm})$, profit shifters d^{tm}, and a set of acquisition prices P^{mt} for each acquisition deal in market m at time t. I presume prices are measured with a classical measurement error; if this assumption is not satisfied, the variance of prices could be overestimated. The data do not have to contain any direct information on the cost for the cost curve to be identified.

To facilitate the inference process, I assume the data are generated by a single MPE strategy profile $\mathbf{g}^*$. Bajari, Benkard, and Levin (2007) make the same assumption because it allows pooling markets during the estimation. Note that this assumption is weaker than the implicit assumption about equilibrium selection the majority of full-solution (nested fixed point) estimation schemes make (for discussion, see Bajari, Benkard, and Levin (2007) page 1332). Single MPE assumption does not presume any particular selection; it merely requires the selection to be the same across markets.

I conduct the estimation of the dynamic model in two steps. In the first step, I propose a flexible parametric estimator that recovers merger and repositions strategies. In the second step, I use an MPE assumption to construct inequalities that identify the structural parameters.

4.1 First step

I start by constructing three auxiliary data sets using a sequential structure of the acquisition and repositioning process. For each t and the data point (s^t, d^t), the econometrician can infer the predefined sequence of player moves $\sigma(s_t, d^t)$. The move sequence determines the merger and repositioning actions $(a^t_{\sigma_1}, \ldots, a^t_{\sigma_K}, r^t_{\sigma_1}, \ldots, r^t_{\sigma_K})$ that occur between time t and $t+1$. This information can be used to construct a data set of acquisition prices

$$X^P = \{(P^t_{kj}, s^{tm}, d^{tm}, j, A^t_{kj}) : k \in K^m, 1 \leq m \leq M, 1 \leq t \leq T, a^t_{kj} = 1\},$$

acquisition actions

$$X^A = \{(a_{kj}^t, s^{tm}, d^{tm}, j, A_{kj}^t) : k \in K_m, j \in \{j' : o_{kj'}^t \neq k\}, 1 \leq m \leq M, 1 \leq t \leq T\},$$

and repositioning actions

$$X^B = \{(r_j^t, s^{tm}, d^{tm}, j, R_{kj}^t) : k \in K_m, j \in \{j' : o_{kj'}^t = k\}, 1 \leq m \leq M, 1 \leq t \leq T\}.$$

I do not estimate the equilibrium strategies directly.[11] Instead, I estimate conditional choice probabilities (CPP) for mergers

$$\text{Prob}^A(a_{kj}|s^{tm}, d^{tm}, j, A_{kj}^t) \in \Delta(\{0, 1\})$$

and repositioning

$$\text{Prob}^R(r_j|s^{tm}, d^{tm}, j, R_{kj}^t) \in \Delta(\{1, \ldots, F\}).$$

In the equilibrium, CCPs depend on the distributions of unobservables ψ and ϕ and differences between choice-specific value functions for available actions. Because shocks are additive, the CCPs' dependence on the multi-dimensional state space can be described by a single-index function, which is the difference between the choice-specific value function of the relevant action and the choice-specific value function of the reference action. Because the exact form of this index function is unknown, one has to use a non-parametric or a semi-parametric estimator. The estimator in this paper is similar to a series estimator, which, in the small sample, amounts to using a flexible parametric function $\widehat{\text{Prob}}^A(a_{kj}|s^t, d^t, j, A_{kj}^t, \theta^{ACQ})$ and $\widehat{\text{Prob}}^R(r_j|s^t, d^t, j, R_{kj}^t, \theta^{REP})$, and maximizing the pseudo-likelihood function based on the distribution of ϕ and ψ. The asymptotics of such estimators (as the size of a data set and dimensionality of a pseudo-parameter vector goes to infinity) is well behaved according to Newey (1994). To operationalize this approach, I use a linear link function of several statistics Υ about the state space computed from the data (a similar approach can be found in Ellickson and Beresteanu (2005), Ryan and Tucker (2011), and Ryan (2012)).

Suppose the payoff shock to the acquisition of station j is $\phi_{jk}^t = \epsilon_{jk}^t - \bar{\epsilon}_{jk}^t$, where ϵ_{jk}^t and $\bar{\epsilon}_{jk}^t$ are two independent type-1 extreme value random variables. In such a case, I propose a logit

[11] The merger bid acceptance strategy does not need to be estimated, because all merger bids are accepted in the equilibrium.

approximation

$$\widehat{\text{Prob}}^{A}(a_{kj}|s^t, d^t, j, A^t_{kj}, \theta^{ACQ}) = \frac{\mathbf{1}_{a_{kj}=1} \exp\left\{\theta^{ACQ} \cdot \Upsilon^{ACQ}(s^t, d^t, j, A^t_{kj})\right\} + \mathbf{1}_{a_{kj}=0}}{1 + \exp\left\{\theta^{ACQ} \cdot \Upsilon^{ACQ}(s^t, d^t, j, A^t_{kj})\right\}}.$$

The approximation is not exact, because Υ^{ACQ} is parametric. In particular, Υ^{ACQ} contains a set of statistics about the state space as well as previous acquisitions by a company k made this period. Inclusion of past actions by player k (contained in A^t_{kj}) in Υ^{ACQ} generates correlation in station-by-station decisions that approximates joint mergers that span across multiple stations. The estimation depends on the order in which merger shocks are revealed (see section 3.3, game stage (A1)). In the baseline specification, I order stations by quality ξ^t_j and perform a robustness analysis in the online appendix. I take a similar approach when estimating the repositioning strategy. I assume ψ^t_{jk} is distributed as a type-1 extreme value that generates the following multinomial logit approximation:

$$\widehat{\text{Prob}}^{R}(r^t_j|s^t, d^t, j, R^t_{kj}, \theta^{R}) = \frac{\exp\left\{\theta^{R} \cdot \Upsilon^{R}(r^t_j, s^t, d^t, j, R^t_{kj})\right\}}{\sum_{r'=1}^{F} \exp\left\{\theta^{R} \cdot \Upsilon^{R}(r', s^t, d^t, j, R^t_{kj})\right\}}.$$

Note the above estimators allow for some selection of actions on serially correlated unobservables. Because s^t contains an unobserved heterogeneity summarized for each station by ξ^t_j, the strategies (acquisitions, prices, acceptance and repositioning) are allowed to be a function of ξ^t_j and ξ^t_{-j}. [12]

4.2 Second stage

In the second stage, I employ a specific parametrization of the fixed cost function that accounts for within-format cost synergies and within-market economies of scale. First I adjust the number of owned stations to account for distribution of the portfolio across formats. I denote this adjusted number of owned stations as N_k and compute it using the following formula:

$$N_k(s^t|\theta^{SYN}) = \sum_{f=1}^{F} \left[\mathbf{1}_{n^t_{kf}=1} + \mathbf{1}_{n^t_{kf}>1}\theta^{SYN}(n^t_{kf} - 1)\right], \tag{4.1}$$

where n_{kf} is the number of stations of format f that player k owns at time t. The key parameter is θ^{SYN}, which measures the contribution of an extra station in the format the player k already

[12] Note that becuase of the structure of the game described in Section 3 the acquisition price does not depend of ϕ and ψ; however, the acquistion price is a function of ξ^t.

owns. In an extreme case when $\theta^{SYN} = 1$, N_k is just the number of owned stations. In another extreme, when $\theta^{SYN} = 0$, N_k is the number of distinct formats player k owns, and the contribution of extra stations in the same format is zero. Thus, the adjusted number of owned stations, N_k, is a rational number somewhere between the number of owned stations and the number of owned formats.

I use the adjusted number of owned stations, N_k, as an input for the below cost function:

$$F_k^{C,m}(s^t|\theta^{FIX}, \theta^{SYN}, \theta^{SCALE}) = \theta_m^{FIX} \sum_{n=1}^{\lfloor N_k^t \rfloor} MFC(n) + (N_k^t - \lfloor N_k^t \rfloor) MFC(\lfloor N_k^t \rfloor + 1), \tag{4.2}$$

where $MFC(n)$ is the marginal fixed cost contribution of the n$^{\text{th}}$ station and $\lfloor \cdot \rfloor$ is a floor operator. Note that I use linear interpolation between $MFC(\lfloor N_k^t \rfloor)$ and $MFC(\lfloor N_k^t \rfloor + 1)$ if N_k^t is a fraction.

I allow for four different values of cost level θ^{FIX}, depending on market population size, that is, for markets with populations greater than 2.5M, 1M-2.5M, 0.5M-1M, and less than 0.5M. The dependence of θ^{FIX} on market size is motivated by the fact that the Occupational Outlook Handbook quotes much larger salaries in the broadcasting industry in larger markets. The parameter $MFC(1)$ is set to 1 so θ^{FIX} is the cost of operating one station. I set $MFC(2) = \theta_1^{SCALE}$ and $MFC(8) = \theta_2^{SCALE}$. I compute MFCs for intermediate values by linear interpolation. This formulation allows for increasing ($\theta_1^{SCALE} > \theta_2^{SCALE}$) as well as decreasing ($\theta_1^{SCALE} < \theta_2^{SCALE}$) marginal fixed cost of operating an extra station. Because it is possible that $\theta_1^{SCALE} < 1$, the model allows for decreasing, increasing, as well as non-monotonic average fixed cost.

I estimate two versions of the above parametrization: (i) Specification 1, which includes both economies of scale and within-format synergies, and (ii) Specification 2, which assumes away within-format synergies. I estimate two models for the following reasons. First, knowing whether the final conclusions depend on the particular format of within-format synergies might be of interest. Second, allowing for both types of efficiencies requires more variation in the data and can lower precision of the estimation. In such a case, estimates of a more stylized model with tighter confidence bounds could provide a better idea about the magnitude of parameters.

To allow for heterogeneity of repositioning costs across markets and to keep the number of estimated parameters small, I set the repositioning cost to be $F_m^S(f_j^t, f_j^{t+1}|\theta) = \theta_m^{REPCOST} = \theta^S \theta_m^{FIX}$. This assumption means the heterogeneity in fixed costs captures all the heterogeneity in repositioning cost across markets. Such formulation is a compromise that emphasizes estimating the aggregate level of fixed cost efficiencies over cross-market heterogeneity. This compromise is necessary because richer heterogeneity is not identified, given the available data, and produces large confidence bounds. I notethat it effectively forces scaling of cost efficiencies and repositioning cost according to the same number, which might skew the comparison of fixed and repositioning costs across markets. However, at the same time, the estimated scale parameter θ^S enables comparison of the overall level of cost efficiencies with the repositioning cost, which is the most relevant for getting credible estimates of the aggregate level of fixed cost efficiencies.

The standard deviation of unobserved profit from mergers θ^{ϕ}_m and switching θ^{ψ}_m is assumed to be proportional to the average observable per-period market revenue of the owner; that is, $\theta^{\phi}_m = \theta^{\phi}(1-\beta)A_m$. This formulation allows for intuitive interpretation of the parameter θ^{ϕ} as the standard deviation of a percentage of one-time costs/profits from mergers that is unobserved.

The value function V_k (defined in equation (3.6)) can be separated into four parts,

$$V^t_k = A^t_k + \theta^{\phi} B^t_k + \theta^{\psi} C^t_k + D^t_k,$$

where

$$A^t_k = E\sum_{r=t}^{\infty} \beta^{r-t}\pi_k(s^t, d^t) + \sum_{j:o^r_j=k,o^{r+1}_j\neq k} P^r_{o^{r+1}_j j} - \sum_{j:o^r_j\neq k,o^{r+1}_j=k} P^r_{kj}$$

is the expected stream of revenues,

$$B^t_k = E\sum_{r=t}^{\infty} \beta^{r-t} \sum_{j:o^r_j\neq k,o^{r+1}_j=k} \phi^r_{kj}$$

is the expected stream of acquisition payoff/cost shocks,

$$C^t_k = E\sum_{r=t}^{\infty} \beta^{r-t} \sum_{j:o^{r+1}_j=k} \psi^t_{kjf^{r+1}_j}$$

is the expected stream of repositioning payoff/cost shocks, and

$$D^t_k = E\sum_{r=t}^{\infty} \beta^{r-t} \left[F^C_k(s^r|\theta^{FIX}, \theta^{SCALE}) + \sum_{j:o^{r+1}_j=k} \mathbf{1}(f^{r+1}_j \neq f^r_j)\theta^{REPCOST} \right]$$

is the expected stream of fixed costs and repositioning costs.

Accounting for B^t_k in the simulation of profits from a merger takes care of some selection on unobservable payoff shocks. Similarly to the first-stage estimation, I make an approximation that merger bids are made sequentially according to ξ_j^t. Under this approximation, given a merger action a_{jk}^{tm}, the contribution of unobserved profits is equal to $\theta^{\phi}E[\phi_{jk}^{tm}|a_{jk}^{tm}]$. Because a company observes the payoff shock before making an acquisition, the mergers that occur are selected for a high value of ϕ_{jk}^{tm}. If ϕ has zero mean, $E[\phi_{jk}^{tm}|a_{jk}^{tm}=1]>0$. [13] Assuming $E[\phi_{jk}^{tm}|a_{jk}^{tm}=1]=E[\phi_{jk}^{tm}]=0$ would cause underestimation of profits from mergers and could result in overestimation of fixed-cost efficiencies. One can make the same point about the selection on unobservables when estimating repositioning products.

Note that only the last part of D_k^t depends on the parameters of interest $\theta^{FIX},\theta^{SYN},\theta^{SCALE}$, and $\theta^{REPCOST}$, and the value function is linear in θ^{ϕ} and θ^{ψ}. Therefore, to compute the value function for different parameter values, one does not need to re-simulate the industry path (s^t,d^t); moreover, one does not need to recompute any of A_k^t, B_k^t, C_k^t, saving a large amount of processing power and making the estimator feasible.

Let V^n be an equilibrium value function, where n indexes players, states, markets, and time periods. Consider W types[14] of suboptimal strategies $\mathbf{g}_w^n$. For each $w\le W$, compute a suboptimal value function $\tilde{V}_w^n(\mathbf{g}_w^n,\mathbf{g}^{-n})$, where $\mathbf{g}^{-n}$ is an equilibrium strategy for the competitors of a player prescribed by an index n.

Consider a sample of size N of indexes (one could sample states, players, time periods, or markets). Following equation (3.7), I define a minimum distance estimator

$$(\hat{\theta}^{FIX},\hat{\theta}^{SYN},\hat{\theta}^{SCALE},\hat{\theta}^{REPCOST},\hat{\theta}^{\phi},\hat{\theta}^{\psi})=\operatorname{argmin}\frac{1}{N\times W}\sum_{n,w}\Omega_w\left(\max\{\tilde{V}_w^n-V^n,0\}\right)^2, \quad (4.3)$$

where Ω_w are positive weights. If enough variation in revenue shifters is available so that the above minimum is unique, one obtains the point identification. Then, according to the results in BBL, the estimator is consistent and asymptotically normal. The suboptimal value function $\tilde{V}_k$ is computed under four suboptimal strategies:

1. "More mergers": Increase the probability of merger by 50% (not by 50 percentage points)

[13] For the case of the extreme value distribution of shocks, $E[\phi|a=1]$ could be reduced to $-\log(p)-\frac{1-p}{p}\log(1-p)$, where p is the probability of acquisition.

[14] I introduce multiple types of sampled off-equilibrium policies to stress the argument validating identification of the model (different types identify lower and upper bounds on cost efficiencies). The original Bajari, Benkard, and Levin (2007) setup has only one type; however, my estimator is a special case of BBL and is consistent even if W is smaller than the number of parameters.

until first suboptimal merger happens. [15]

2. “Fewer mergers”: Prevent a first merger.

3. “More format switching”: Switch the first station into the random format.

4. “Fewer format switching”: Prevent any switches in the first year.

These four strategies produce four counterfactual value functions V_w^n. The condition that the value function V cannot be negative is included as a fifth set of restrictions. Because the nominal deviations for larger markets are higher, I weigh each deviation from equilibrium by an inverse of the counterfactual revenues, $(\tilde{A}_k^{tm})^{-1}$. These weights do not depend on parameters of the dynamic model and, in practice, prevent the outliers from dominating the results. Moreover, to ensure that neither of the restrictions is dominating others, I set to one the contribution of each inequality restriction at a starting point (all parameters are zero). The value function is averaged across 200 forward simulations of 80 half-year periods. The last period is assumed to persist forever. Equilibrium conditions are imposed for starting states of owners with the largest listenership shares during Spring of 1997, 1999, and 2001 in each of 68 markets. Consequently, the procedure uses $1,020$ inequalities.

Four main parts of the model need to be identified: (i) level of the fixed cost, (ii) fixed-cost synergies, (iii) repositioning cost, and (iv) variances of payoff shocks. The level of the fixed cost is identified by the combination of three assumptions: (i) repositioning cost to/from DARK is the same as repositioning to any other format, and operating a DARK station is free, (ii) operating one station is profitable in any market (upper bound on the fixed cost), i.e., $V > 0$, and (iii) the fixed cost of one station has to be large enough to generate efficiency rationalizing mergers (lower bound on the fixed cost). I find that in practice (ii) and (iii) are more important than (i) (see online appendix). Fixed cost synergies are identified as a residual from the merger prediction, and similarly repositioning cost is identified as a residual from the repositioning prediction. The variances of payoff and repositioning cost are identified from the variance in the observed actions conditional on the state. The normalization of the variance is not necessary because I observe revenues in dollars.

[15] A suboptimal merger is the one that happens only because of the perturbation; that is, my uniform number generator draws a number larger than the optimal CCP but smaller than the perturbed CCP.

4.3 FCC ownership caps

FCC ownership caps are an important feature of the radio market. They summarize the majority of the antitrust regulations the radio owner faces. In this paper, I ignore all other antitrust issues. I impose the ownership caps in both the first and second stages of the estimation. In the first stage, I assign probability zero to the mergers that are infeasible to execute, and because I control for the percentage of stations owned (effectively, an ownership cap), I allow for different propensities to merge when close and far away from the cap. In the second stage, when simulating the value function, I do not allow for any decisions that violate the ownership caps. Note that imposing the caps is important when calculating the optimal as well as suboptimal value function. If the caps were not imposed on the suboptimal path, this path might be infeasible and could violate the equilibrium inequalities even at the true parameters. Also, note that because the caps cannot be violated, the model will never collapse to the monopoly. Moreover, because the caps are relatively strict they give incentives not to merge early and keep an option value of merging later, which makes "More mergers" strategy yield suboptimal value function.

5 Results

This subsection describes the results of the estimation. I divide the exposition into three parts. First, I conduct a brief discussion of the static payoff function estimates. I present only a subset of these estimates because of space constraints and to avoid repeating the discussion contained in Jeziorski (2013). [16]

Second, I present the first-stage estimates: acquisition pricing, acquisition strategy, and format-switching strategy. The transition of ξ_j^t, prescribed by equation (3.1), as well as a distribution of ζ_j^t (non-parametric), is estimated jointly with a static payoff function. I find $\hat{\rho} = 0.56$ with 0.09 standard deviation. During the simulation, I draw from an empirical distribution of ζ_j^t, controlling for different variance in each market. Moreover, I control for heteroscedasticity of ζ_j^t by allowing for different distributions conditional on switching format, or switching to/from DARK. The model assumes the evolution of unobserved station quality ξ_j^t is exogenous, which rules out a causal

[16] A complete discussion is contained in the online appendix, which is available on the author's website: `http://jeziorski.me`

	Average of not acquired stations	Average of acquired stations	Difference
Half-year forward, $\xi_j^{t+1} - \xi_j^t$	-0.016 (0.004)	0.006 (0.020)	-0.022 (0.020)
1-year forward, $\xi_j^{t+2} - \xi_j^t$	-0.032 (0.004)	-0.003 (0.015)	-0.030 (0.016)
2-year forward, $\xi_j^{t+4} - \xi_j^t$	-0.070 (0.005)	-0.041 (0.013)	-0.029 (0.014)
3-year forward, $\xi_j^{t+6} - \xi_j^t$	-0.113 (0.007)	-0.067 (0.012)	-0.046 (0.014)

Table 2: Estimated change in the unobserved quality for not-acquired stations and stations that were acquired at time t. It can be used to investigate whether acquired stations have different evolutions of mean quality. The change in quality is measured as $\xi_j^{t+\Delta t} - \xi_j^t$, where ξ_j^t is a quality of stations j at time t and Δt is time after the acquisition. To assess an economic impact of the difference, one can compare it to 1.3 standard deviations of ξ_j^t across stations.

effect of a merger on station quality. This assumption is important because the positive impact of mergers on quality could be an alternative hypothesis to cost synergies that would rationalize the merger wave. Using my data, this assumption could be verified without imposing supply side (neither static, nor dynamic), because quality ξ_j^t is a residual in the demand estimation. First, I compare the changes to the mean of ξ_j^t for stations that switched owners with those that did not. Table 2 presents the mean change in ξ_j^t for both types of stations. It tracks half-year and one-, two-, and three-year changes in ξ_j^t because the mergers could have a long-run effect on quality. I find that on average, mergers have a negligible impact on mean quality, and I cannot reject that this impact is zero in the first six months. The long-run effect is statistically significant (I can detect even small effects because I use 26,778 observations of $\xi_j^{t+\Delta t} - \xi_j^t$, and standard errors assume independence of these observations) but economically negligible, amounting to between 2% and 3% of the standard deviation of ξ_j^t, and 1% to 3% of the standard deviation of $\xi_j^{t+\Delta t} - \xi_j^t$. I also performed a market-by-market Kolmogorov-Smirnov test to investigate whether conditional distributions of $\xi_j^{t+1\text{ year}} - \xi_j^t$ are different when conditioning on acquisition or lack thereof at time t. I cannot reject that the distributions are the same in 68 of 88 markets at the 5% level and in 80 markets at the 1% level. Additionally, because mergers are frequently followed by repositioning, which generates larger variance of the innovation ζ_j^t, conducting a weaker test that allows for heteroscedasticity might be more relevant. To obtain such a test, I normalize the distributions to have unitary standard deviations, and redo the K-S test. In this case, I cannot reject that the normalized distributions are the same in 86 markets at the 5% level and all markets at the 1% level. I repeated the test for a two-year gap and get similar results.

In Section 5.5, I present the estimates of fixed cost and repositioning cost parameters, followed by counterfactuals. I perform a correction of standard errors for sequential estimation using a parametric bootstrap; namely, I draw first-stage parameters from a joint asymptotic normal distribution of profit function parameters, quality auto-correlation parameter ρ, and first-stage equilibrium strategy estimates. Note that all these parameter estimates are correlated because station quality ξ_j^t is a function of profit parameters and is an input to the estimation of the strategies. Thus the first stage, in fact, comprises two substages. To obtain correct asymptotic distribution, I cast profit function and strategy estimation as a sequential system GMM estimation. This correction is valid according to the results of Ai and Chen (2007) and Ackerberg, Chen, and Hahn (2012). For each draw of profit function parameters, I recompute the implied quality of each station ξ_j^t, which is an input to the second stage. Recomputing ξ_j^t for each bootstrap draw accounts for the estimation error in ξ_j^t. Subsequently, I re-simulate the value functions and reestimate the second stage. I correct standard errors for a second-stage simulation error by using independent draws for each bootstrap iteration. Standard errors are based on 70 parametric bootstrap draws. Each re-estimation of the second stage takes about eight hours using a 48 CPU (2GHz AMD Opterons) cluster running an optimized and parallelized C code. Full estimation procedure takes about three weeks.

5.1 Static payoff function

First column of the Table 3 contains estimates of demand parameters for radio programming. The estimate of the mean effect of advertising on listeners' utility is negative and statistically significant. This finding is consistent with the belief that radio listeners have a disutility for advertising. Regarding the mean effects of programming formats, the Contemporary Hit Radio format gives the most utility, whereas the News/Talk format gives the least. The second column of Table 3 contains variances of random effects for station formats. The higher a format's variance, the more persistent the tastes of that format's listeners. For example, in response to an increase in advertising, if the variance of the random effect for that format is high, listeners tend to switch to another station of the same format. The estimates also suggest tastes for the Alternative/Urban

	Mean Effects (θ_1^L)	Random Effects (Σ_1)	Demographics characteristics (Π) Age	Sex	Education	Income	Black	Spanish
Advertising	-1.11 (0.002)	0.03 (0.009)	-					
AM/FM	0.86 (0.000)	-	-					
AC	-2.43 (0.008)	0.04 (0.004)	-0.17 (0.001)	-0.34 (0.064)	0.60 (0.013)	-0.02 (0.003)	0.12 (0.012)	-1.01 (0.008)
Rock	-1.56 (0.140)	0.00 (0.020)	-0.65 (0.072)	0.40 (0.031)	0.86 (0.006)	-0.15 (0.045)	-1.36 (0.007)	-1.64 (0.003)
CHR	-0.18 (0.025)	0.01 (0.006)	-2.54 (0.015)	0.48 (0.080)	1.77 (0.006)	-0.29 (0.005)	1.95 (0.015)	0.46 (0.001)
Alternative Urban	-2.34 (0.026)	0.35 (0.008)	-0.82 (0.008)	1.35 (0.018)	0.58 (0.025)	-0.14 (0.002)	3.15 (0.005)	0.27 (0.027)
News/Talk	-4.68 (0.010)	0.02 (0.002)	0.33 (0.002)	1.23 (0.012)	0.24 (0.009)	0.09 (0.005)	-0.32 (0.001)	-1.65 (0.005)
Country	-2.30 (0.006)	0.01 (0.003)	0.06 (0.004)	-0.15 (0.022)	0.13 (0.004)	-0.13 (0.003)	-1.55 (0.009)	-1.72 (0.002)
Spanish	-1.62 (0.004)	0.01 (0.001)	-0.02 (0.013)	-0.91 (0.012)	-0.33 (0.018)	-1.14 (0.002)	-2.56 (0.004)	0.80 (0.003)
Other	-4.66 (0.004)	0.01 (0.002)	0.26 (0.373)	0.62 (0.003)	0.34 (0.006)	-0.03 (0.063)	0.50 (0.001)	0.24 (0.002)
ρ	0.57 (0.091)	-	-					

Table 3: Estimates of demand for radio programming.

format are the most persistent.

Last six columns of the Table 3 contain estimates of interactions between listener characteristics and format dummies. These values measure local market power caused by taste preferences for formats. Thus they determine incentives to switch formats as well as acquire closer to or further from the current portfolio. The majority of the parameters are consistent with intuition. For example, younger people are more willing to choose a CHR format, whereas older people go for News/Talk. The negative coefficients on the interaction of the Hispanic format with education and income suggest less educated Hispanic people with lower incomes are more willing to listen to Hispanic stations. For Blacks, I find a disutility for Country, Rock, and Hispanic, and a high utility for Urban. This finding is consistent with the fact that Urban radio stations play mostly rap, hip-hop, and soul music performed by Black artists.

In markets with less than 0.5m people, radio stations have considerable control over per-listener price because the slope of the inverse demand, θ_2^A is large, namely 1.34 (0.046). However, such control significantly drops in markets with populations of 0.5m to 2m people, where I find the slope of 0.35 (0.026). Radio stations essentially price takers in markets with more than 2m people, because I cannot reject that θ_2^A in these markets.

I use the numbers in the aforementioned tables to discuss the impact of mergers on the static payoff. I conjecture that in the markets where the advertising demand is steeper, the merger should have an impact on payoffs similar to that in the Cournot model. I choose three markets with different slopes as examples: Los Angeles, CA (pop. 13M), Knoxville, TA (pop. 737k), and Bismarck, ND (pop. 99k). I compute a static merger counterfactual for each possible acquisition by Clear Channel in 1997 and 1998 in these markets. In particular, I keep everything at 1997 values and enlarge Clear Channel's portfolio by one station. Then I compare the revenue of the relevant group of stations before and after the merger. I find that about 7%-9% of potential mergers in Los Angeles, 18.7% in Knoxville, and 20%-30% in Bismarck are not profitable. This finding suggests fixed-cost synergies are needed to support some of the potential mergers. Note that if one accounts for dynamic effects such as post-merger repositioning, even more mergers might be unprofitable in the long run.

5.2 First stage: Demographic dynamics

For the purposes of this article, I am interested in capturing only the first-order mid- and long-run trends that might affect format switching. When simulating the value function for each period, I record the share of different demographic groups in all the markets (groups can be found in Table 6). For periods before 2009, I compute these shares using CPS. For periods 2009 and after, I use national census projections of growth rates of appropriate demographic groups and forecast their shares in each market (for education and income groups, I simply compute the mean 1996-2006 shares and fix it for all years after 2006). I use these shares when computing the integral (3.2) and enter them as a series of independent binomial random variables.

5.3 First stage: Acquisition pricing

Table 4 shows the results of an OLS regression of acquisition prices on chosen statistics from the information set. The top part of the table contains market-level covariates. The listeners' population is a big driver of acquisition price because per-listener ad prices are largely dependent on

	Variable	OLS	Heckman 2nd stage	Heckman 1st stage
	Constant	12.19*** (2.17)	12.13*** (2.06)	8.09 (16.18)
Market characteristics	Population (M)	0.00*** (0.00)	0.00*** (0.00)	0.00 (0.00)
	Population 4M-	1.87*** (0.21)	1.86*** (0.23)	1.40 (2.32)
	Population 2.5M-4M	1.83*** (0.14)	1.90*** (0.16)	1.60 (2.21)
	Population 1M-2.5M	1.25*** (0.09)	1.25*** (0.11)	1.04 (1.46)
	Population 0.5M-1M	0.49*** (0.08)	0.48*** (0.09)	0.30 (0.67)
	% of format	−1.28*** (0.35)	−1.25*** (0.35)	−1.93*** (0.57)
	Avg. quality of format	0.04 (0.04)	0.04 (0.04)	−0.04 (0.12)
	Spanish/Hispanic	1.35*** (0.39)	1.34*** (0.39)	0.58 (2.10)
	Urban/Black	0.38 (0.51)	0.37 (0.49)	−0.59 (1.32)
	News/Young	1.09*** (0.26)	1.09*** (0.26)	0.58 (1.60)
	CHR/Young	0.07 (0.41)	0.04 (0.40)	−0.74 (0.83)
Station characteristics	Quality	1.27*** (0.43)	1.26*** (0.41)	0.47 (2.06)
	Quality^2	−0.05** (0.02)	−0.05** (0.02)	−0.09*** (0.01)
	Dark	−0.00 (0.21)	−0.03 (0.20)	−0.42 (0.36)
	Reporting	−5.87*** (2.17)	−5.80*** (2.06)	−9.84*** (1.76)
	AM	−1.34*** (0.07)	−1.36*** (0.07)	−1.50 (1.22)
Competition characteristics	Number of stations owned	0.16*** (0.03)	0.16*** (0.03)	0.09 (0.22)
	Avg. quality of format, owner	−0.05 (0.03)	−0.05 (0.03)	−0.11*** (0.01)
	Entering buyer	0.31*** (0.10)	0.32*** (0.12)	0.09 (0.56)
	Top 3 seller	0.42*** (0.08)	0.41*** (0.16)	0.10 (0.71)

Standard errors (corrected for sequential estimation for OLS) in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Determinants of acquisition price conditional on a merger. Number of observations: 1,449 for OLS and 3,123 for Hackman.

the size of the market. Dummies, as well as the coefficient on the population size, are positive and highly significant. The percentage of stations in the format of the acquired station has a highly significant negative impact. The more stations in the same format, the tougher the competition for listeners and advertisers, which drives down station profitability. The large value of this coefficient (a 1 percentage point increase translates into a 1.2 percentage point decrease in acquisition price) and its high significance suggest high switching cost. Limited evidence suggests that demographics affect acquisition price. For example, interactions between the percentage of the Hispanic population and the Hispanic-format dummy are positive and significant.

The second part of the table consists of station-level covariates. Station quality positively affects price; however, the effect is diminishing. This finding is consistent with the fact that in the assumed profit function, station quality has a diminishingly positive effect on revenues. On average, DARK stations are cheaper than their active counterparts, and FM stations are more expensive than similar AM stations. Additionally, I use a dummy variable to control for the fact that some stations do not meet Arbitron minimum reporting standards (less than 0.05% market share).

The last part of the table consists of buyer and seller characteristics. The price is positively affected by the number of stations already owned. This finding can be explained by either larger marginal market power and cost efficiencies of extra stations for larger buyers, or higher bargaining power of larger buyers. The coefficient on the dummy controlling for the size of the seller (top three in a move sequence) is positive. This observation suggests that, controlling for station covariates, higher-ranked sellers obtain higher prices. This finding might be explained by a greater amount of business stealing if buying from a bigger competitor, or by better dynamic outside options of larger sellers. At the same time, I find no direct effect of the ranking of the buyer on price (coefficient not included in the final regression); however, including the number of owned stations already accounts for some of this effect.

The last two columns of the table present a robustness check using the two-step Heckman selection model. The analysis is not aimed at fully correcting for selection, because exclusion restrictions are not available. However, even in the absence of the exclusion restriction, one can check whether the results are robust to selection based on a particular functional form of error correlation. For example, the selection might be driven by the size and profitability of the stations because prices for larger deals are unobserved. Because the selection equation includes unobserved station quality ξ_t and a reporting standards dummy, which are proxies for size and unobserved profitability, the two-step estimator should provide some information about selection bias. Indeed, the selection is highly driven by the reporting dummy and the square of ξ_j^t. Another driver is the percentage of stations in the format, which suggests the data set on prices under-represents popular formats. However, the second-stage estimates of the selection model are similar to OLS, which suggests the aforementioned drivers of selection might not be significantly affecting the price regression.

5.4 First stage: Acquisition strategy

I use a flexible parametric approximation of the acquisition strategy which contains 234 covariates. These covariates include controls for the size of buyer and seller, station characteristics, acquirer's and competitors' current portfolio, and dependence between multiple acquisition actions within period. To focus the discussion on the most important parameters I report estimates of the subset of covariates in the article and report the remaining covariates in the online appendix. To estimate the acquisition strategy I run a joint Maximum Likelihood Estimation (MLE); however, the results are grouped into multiple tables to improve exposition.

First six columns of Table 5 show controls for buyer and seller size in the form of dummies on buyer's and seller's ranking by listenership last period. I find higher-ranked buyers are more likely to acquire new stations as a result of either an increasing amount of market power or cost efficiencies. The structural estimation in the second step can disentangle these two stories. Additionally, I find companies are less likely to purchase stations from higher-ranked sellers, which is consistent with higher-ranked sellers quoting on average higher prices for similar stations.

In the second section of Table 5 I present the impact of chosen station characteristics on the propensity to acquire this station. I find that smaller stations are acquired more often because the listenership ranking of the target has a negative coefficient. Moreover, companies are less likely to purchase AM stations and more likely to purchase stations that do not meet Arbitron reporting standards. Given the fact that entry in the market is limited and the price of such stations is much lower, purchasing these stations is a relatively inexpensive way to enter or to introduce new stations. I find no statistical impact of the target's quality ξ on the propensity to acquire. This

Buyer and seller size	
Acquirer has the largest listenership	1.26*** (0.12)
Acquirer has the second largest listenership	0.82*** (0.11)
Acquirer has the third largest listenership	0.44*** (0.10)
Seller has the largest listenership	−0.34*** (0.06)
Seller has the second largest listenership	−0.07 (0.06)
Seller has the third largest listenership	−0.08 (0.06)
Chosen station characteristics	
Station listenership ranking	−0.04*** (0.00)
AM	−0.18*** (0.05)
Below Arbitron reporting standard	0.44*** (0.14)
Station quality ξ	0.03 (0.04)
Avg. quality in the format of the acquisition	
Acquirer	0.19*** (0.02)
Two largest competitors	−0.00 (0.02)
Others	0.06*** (0.02)
Executed acquisitions this period	
One station acquired	−5.32*** (0.11)
Two stations acquired	−2.05*** (0.11)
Three stations acquired	−1.02*** (0.10)
Four stations acquired	−0.67*** (0.11)
Controls for acquirer's portfolio	YES
Controls for portfolios of the competitors	YES
Controls for demographic composition of the market	YES

Standard errors (corrected for sequential estimation) in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 5: Chosen estimates of the acquisition strategy covariates. Number of mergers in the data: 3,123. Overall number of opportunities to acquire (data size): 732,041.

finding is consistent with the fact, as described in the previous section, that higher-quality stations bring in more revenue but also cost more to acquire.

Third section of Table 5 presents interactions between the quality of the potential acquiree with the quality of already owned and competing stations. I find that the average quality of the owned stations in the format of a potential acquiree increases the propensity to merge. This result is consistent with demand-side quality synergies. I demonstrate a negligible impact of the quality of competitors on the propensity to acquire.

As I explain in section 3.3, I approximate the multi-station acquisitions with a series of highly correlated decisions. I summarize the correlation structure between the decisions by allowing conditional probabilities of acquiring additional stations to depend on the number of previous acquisitions in the same period. Additionally, to control for the characteristics of previously acquired stations, I update the current portfolio covariates with all previously executed acquisitions. In the fourth section of Table 5, I present the coefficient on dummies indicating the number of past acquisitions. I observe economies of scale in acquiring multiple stations at the same time, which is most likely caused by the fact that they are acquired from the same seller.

The estimation of the acquisition strategy contains flexible controls for the station portfolios of the acquirer and the competitors. Namely, I fully interact the format of the potential acquiree with the percentages of the number of stations in each format owned by the acquirer and its top competitors. Full interaction matrix is reported in the online appendix, however, I summarize the main results below. The coefficients for the interaction between target format and percentage of owned stations in the same format are negative and highly significant, which suggests that owners avoid formats they already own possibly because of cannibalization and high switching costs. In general, the percentage of stations owned is negatively related to further acquisitions, which suggests that the closer the owner is to the ownership cap, the fewer incentives it has to acquire an extra station. This relationship is consistent with an intuition that owners want to keep an option value to acquire in the future, which can be useful in case of changes in market demographics or quality of competitors. Table 6 contains interactions between the format of a potential acquiree and the percentage of different demographic groups within the market. I find demographics are not a big driver of acquisitions. A notable exception is the percentage of low income listeners, which is correlated with acquisition in almost any format.

	AC	Rock	CHR	Urban Alt.	News Talk	Country	Spanish	Other
Age 12-24	2.08 (1.71)	1.84 (2.19)	1.59 (3.48)	8.01*** (3.06)	−0.00 (1.91)	1.78 (2.00)	2.52 (2.74)	−0.67 (1.64)
Age 25-49	3.19 (2.74)	4.19 (3.33)	4.59 (5.09)	4.72 (3.99)	−2.08 (2.74)	2.99 (3.03)	−1.93 (3.84)	−3.69 (2.39)
Some HS	0.02 (1.62)	−1.00 (1.85)	−1.79 (2.66)	−4.35** (2.21)	−2.63* (1.56)	−0.29 (1.73)	−0.95 (2.33)	−1.81 (1.32)
HS Grad.	0.75 (1.45)	−1.12 (1.74)	−0.41 (2.37)	−1.98 (2.12)	−0.80 (1.46)	0.51 (1.55)	1.00 (2.62)	−0.08 (1.26)
Some College	2.54 (1.61)	−0.17 (1.90)	0.63 (2.62)	−2.66 (2.39)	−1.84 (1.54)	1.56 (1.74)	0.01 (2.65)	0.65 (1.35)
Income 0-25k	2.64*** (0.94)	4.79*** (1.12)	5.42*** (1.63)	5.69*** (1.31)	2.70*** (0.92)	2.49** (1.03)	−2.25 (1.60)	2.10*** (0.80)
Income 25k-50k	1.24 (1.18)	1.30 (1.36)	2.55 (1.95)	3.76** (1.60)	1.98* (1.16)	1.75 (1.24)	−6.37*** (1.94)	2.25** (0.96)
Income 50k-75k	1.01 (1.50)	4.27** (1.72)	5.08** (2.53)	2.60 (2.10)	0.83 (1.49)	2.95* (1.60)	−5.80** (2.82)	1.35 (1.22)
Black	−0.30 (0.67)	−0.86 (0.77)	−0.19 (1.14)	0.30 (0.80)	−0.29 (0.66)	0.30 (0.72)	−0.92 (1.31)	0.17 (0.50)
Hispanic	−0.88 (0.65)	−1.45* (0.77)	−0.41 (1.06)	−0.87 (0.95)	−1.10* (0.63)	−0.22 (0.66)	−1.41* (0.77)	0.27 (0.53)

Standard errors (corrected for sequential estimation) in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6: Estimates of acquisition strategy: impact of an interaction between demographic composition in the market and acquisition target's format on acquisition decision. Demographic variables are measured as a fraction of the market population with a particular characteristic. Number of mergers in the data: 3,123. Overall number of opportunities to acquire (data size): 732,041.

To check for goodness of fit, in Table 7 I report the value of the average likelihood across acquisition data points. Average likelihood across all the data is reported in the "All" column as 0.98. Subsequently, I check the robustness of this probability to slicing the data into subsamples. I compute the average likelihood of observed merger decisions only for stations of a particular format. The numbers are stable and consistently high. Even though such an exercise is a within-sample robustness analysis, it suggests the model fits consistently well across heterogeneous subsamples. Because mergers are infrequent events, most of the actions result in "no merger." Therefore, we can reasonably expect most of the variation to be accommodated by an intercept and format dummies. In such a case, high values of the likelihood in column 1 of Table 7 might be misleading. To correct for this issue, I computed McFadden's pseudo R^2 measure, which compares the performance of the

Station format	All	AC	Rock	CHR	Urban Alt.	News Talk	Country	Spanish
Average likelihood of the merger decision observed in the data	0.978	0.976	0.977	0.979	0.980	0.980	0.982	0.979
Pseudo R^2	0.233	0.23	0.27	0.26	0.26	0.22	0.25	0.22
Pseudo R^2 with format dummies	-	0.23	0.27	0.26	0.26	0.22	0.25	0.21

Table 7: Goodness of fit of the merger strategy, measured as the likelihood (at the estimated parameter) of the merger decisions in the data separately for every acquisition target format. Third row uses a baseline model with just an intercept, and the last row uses a baseline model with format dummies.

full model with the baseline model, which has just an intercept. I make the adjustment using the formula[17] $1 - \frac{\ln L(\text{full model})}{\ln L(\text{baseline model})}$. I find the model has pseudo R^2 of 0.23. Considering that predicting mergers is generally difficult, an R^2 of 0.23 obtained by using just covariates in the local market can be reassuring. Furthermore, I repeat the exercise for the subsamples to ensure the model predicts well across different data slices. The R^2 for all formats holds well above 0.2, which suggests no over-fitting on the particular subsamples occurs. To check how much information format dummies contain, I compare the full model to the model with only format dummies. The last row of Table 7 contains the results. Because the new R^2 values are not different from the R^2, which uses only an intercept, I conclude the explanatory power of the model lies not in format dummies, but in other covariates. Similarly one can investigate if acquisition strategy is stationary, conditional on covariates. For example, the model's explanatory power cannot drop too much after the year 2000, when I observe a sharp decline in merger activity. I present at the performance of the model across different time cross sections in Table 8. The first column contains the average likelihood of observed merger decisions, and I find no sharp drop after year 2000. However, because most of the actions result in no merger, more meaningful insights are provided with pseudo R^2 measures. Again, I see no sharp decline in this measure of fit after the year 2000. Last column compares the performance of the full model with a non-stationary model consisting of only half-year time dummies. If the pure time-dummy model did as well as (R^2 of 0) or better than the stationary model (negative R^2), we could infer one should not use a stationary policy to simulate a long-run behavior. However, the stationary model always does better than time dummies. Moreover, the

[17]Although more than one accepted equivalent of R^2 exists for a logistic regression, Menard (2002) argues the above formula closely resembles the relevant OLS calculations.

Cross section	Average likelihood of the merger decision observed in the data	Pseudo R^2	Pseudo R^2 with time dummies
Spring 1997	0.967	0.17	0.16
Spring 1998	0.972	0.25	0.25
Spring 1999	0.971	0.27	0.27
Spring 2000	0.968	0.31	0.30
Spring 2001	0.973	0.25	0.25
Spring 2002	0.990	0.20	0.13
Spring 2003	0.988	0.19	0.14
Spring 2004	0.986	0.15	0.11
Spring 2005	0.988	0.20	0.15
Spring 2006	0.985	0.24	0.23

Table 8: Goodness of fit of the merger strategy measured by an average likelihood of the merger decision observed in the data, separately for each half year (only Spring reported to save space). Third column uses a baseline model with just an intercept, and the last column uses a baseline model with half-year time dummies.

R^2 is always greater than 0.1, with an average value of 0.23 before and 0.16 after the year 2000. The drop suggests the model is losing a bit of explanatory power after 2000, though not much. Comparing columns 2 and 3 might help determine the cause of the loss of precision. If the missing time trend caused the precision loss, one should observe a drop in values in column 3 and no drop in values in column 2. However, R^2s in columns 2 and 3 are lower after 2000. Moreover, these R^2s do not differ by much, which suggests that time dummies do not add much explanatory power. In other words, relative performance of the full model compared to the model with no time trend is roughly the same as compared to the model with a time trend. The drop of R^2 is likely to be due to the increased volatility of merger decisions, as opposed to time trend.

Another way to measure fit of the first stage is to allow for correlation across merger decisions in the alternative model and recomputing the pseudo R^2. I find that in this case, R^2 amounts to 0.08, which suggests that the model has predictive power for the timing and target of the first acquisition in a possible sequence.

Finally, to test whether adding more covariates makes a difference, I re-estimated the model with a richer first-stage specification allowing for an interaction between the market category and acquisition target format (27 extra parameters). This change does not affect the results (see online appendix).

Similar to the acquisition strategy, I estimated the format-switching strategy as one joint MLE run but grouped results into multiple tables. I report only a subset of parameters, but the estimation additionally contains a full set of past-future format dummies, and controls for the portfolio structure that are similar to the acquisition strategy. All unreported numbers are contained in the online appendix.

Table 9 presents the impact of station characteristics on the probability of staying in the current format. Each row represents a current format, and each column represents a station characteristic. First row, which is a diagonal of past-future format interaction matrix, captures the format persistence. Rock is the most persistent format and Dark is the least persistent one. I find AM stations are more likely to stay in their formats, with the exception of News/Talk and Other. The third row of the table presents the impact of the acquisition on staying in the current format. The highly significant negative numbers mean that the probability of format switching conditional on acquisition is much higher than the unconditional probability. The last three rows contain the impact of the average quality of other stations in the format on the propensity to stay in the current format. Owning better stations in the format decreases the probability of switching, although the fact that competitors own them increases the probability of switching out.

Table 10 presents the relationship between the current demographic composition of the market and format-switching decisions. One can observe many patterns that suggest that firms respond to the current state of population demographics, according to demographic tastes for formats. For example, a larger Hispanic population is related to the stations switching to a Hispanic format. One can observe a similar pattern for Blacks and the Urban format, as well as for older people and the News/Talk format. Those patterns largely reflect correlations between tastes for formats and demographics described in Jeziorski (2013).

5.5 Second stage: Fixed costs and switching costs

Table 11 shows the second stage estimates. First section of that table provides estimates of fixed-cost parameters for a model with within-format synergies (Specification 1) and a model without within-format synergies (Specification 2). The cost of operating one station θ^{FIX} is decreasing

	Stay in the current format					
	Fixed effect	If AM	If acquired	Avg. quality in format		
				Owner	Top2	Others
AC	6.71*** (0.51)	0.50** (0.21)	−0.76*** (0.16)	0.05* (0.03)	−0.01 (0.04)	−0.04 (0.03)
Rock	7.12*** (0.69)	0.99** (0.43)	−0.70*** (0.21)	0.16*** (0.05)	−0.12* (0.07)	−0.13*** (0.05)
CHR	6.67*** (0.67)	0.82 (0.55)	−0.85*** (0.26)	0.16*** (0.05)	−0.02 (0.10)	−0.06 (0.06)
Urban Alt.	6.34*** (0.56)	1.17*** (0.30)	−0.63*** (0.23)	0.20*** (0.05)	−0.03 (0.08)	−0.01 (0.04)
News Talk	5.68*** (0.54)	−1.53*** (0.25)	−1.22*** (0.18)	0.17*** (0.03)	−0.10** (0.04)	−0.02 (0.03)
Country	6.49*** (0.49)	0.07 (0.25)	−1.07*** (0.18)	0.07** (0.03)	−0.09 (0.06)	−0.04 (0.03)
Spanish	5.10*** (0.53)	−0.23 (0.29)	−1.74*** (0.23)	0.08** (0.04)	−0.06 (0.10)	−0.05 (0.04)
Other	6.69*** (0.44)	−0.42** (0.17)	−1.07*** (0.13)	0.01 (0.02)	0.01 (0.03)	0.01 (0.03)
Dark	-	−0.38 (0.35)	−0.38 (0.28)	-	-	-

Standard errors (corrected for sequential estimation) in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 9: Format-switching-strategy estimates: The reported parameters are (i) fixed effects for every combination of source and target format (a diagonal is reported in the second column of the above table and the complete switching matrix is reported in the online appendix), (ii) stay-in-format-fixed-effects interacted with AM, being acquired this period, and an average quality stations in the current format by owner. The MLE additionally contains: (iii) interaction between a target format and the fraction of stations owned in each format by the current owner and two largest competitors (reported in the online appendix) and (iv) interactions between market demographics and the target format (reported in Table 10). Number of data points: 49,212.

with the size of the market. The point estimates for θ^{FIX} produced by both cost specifications are similar, which suggests the inference of the general level of fixed cost is robust to some changes in the specification of the cost curve. Specification 1 produces larger confidence bounds than Specification 2, because it makes fewer assumptions about the source of cost efficiencies. Note that because inequalities are imposed only on large players, the level of fixed cost is representative of larger stations, which are the relevant group of stations for the merger counterfactual because

	AC	Rock	CHR	Urban Alt.	News Talk	Country	Spanish	Other
Age 12-24	−3.49 (3.34)	−3.36 (3.61)	0.01 (3.89)	−5.81 (3.70)	−5.13 (3.41)	−0.18 (3.53)	−5.33 (3.61)	−3.96 (3.23)
Age 25-49	−3.10 (4.53)	−1.58 (4.89)	0.71 (5.35)	−1.68 (4.95)	−6.03 (4.59)	−11.09** (4.79)	−6.19 (4.86)	−4.00 (4.33)
Some HS	7.05*** (2.57)	7.10** (2.77)	5.84** (2.98)	6.43** (2.78)	9.11*** (2.58)	6.73** (2.71)	8.09*** (2.77)	7.93*** (2.44)
HS Grad.	2.20 (2.51)	3.48 (2.68)	1.78 (2.82)	1.22 (2.71)	−2.60 (2.54)	0.89 (2.64)	−2.12 (2.83)	0.98 (2.42)
Some College	3.92 (2.75)	3.11 (2.93)	0.58 (3.12)	2.33 (2.97)	2.66 (2.76)	2.10 (2.89)	4.49 (3.02)	4.31 (2.64)
Income 0-25k	−4.14** (1.65)	−5.33*** (1.77)	−3.95** (1.88)	−4.19** (1.77)	−6.35*** (1.66)	−4.82*** (1.73)	−7.54*** (1.81)	−4.98*** (1.59)
Income 25k-50k	0.61 (2.04)	−0.08 (2.16)	2.78 (2.31)	−0.10 (2.20)	−0.15 (2.05)	−0.09 (2.13)	0.42 (2.22)	−0.07 (1.96)
Income 50k-75k	2.95 (2.65)	2.37 (2.80)	2.47 (3.01)	3.95 (2.84)	1.11 (2.66)	1.42 (2.77)	−0.81 (2.94)	1.28 (2.55)
Black	−1.18 (1.08)	−1.16 (1.17)	−0.29 (1.27)	2.14* (1.15)	−0.30 (1.09)	−0.58 (1.15)	−0.89 (1.21)	0.13 (1.03)
Hispanic	−0.51 (0.98)	−0.63 (1.07)	−0.62 (1.15)	1.08 (1.07)	0.10 (0.97)	−0.72 (1.02)	2.01** (1.00)	−0.37 (0.92)

Standard errors (corrected for sequential estimation) in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 10: Format-switching-strategy estimates: choice-specific parameters on current demographic composition in the local market. Demographic variables are measured as the fraction of the market population with a particular characteristic. Number of data points: 49,212.

they compose a vast majority of transactions. To investigate whether the general level of the fixed cost is reasonable, I provide a couple of examples. In large markets, such as Houston, the revenue of large stations amounts to \$20m-\$30m depending on the year and format. Multiplied by an average variable profit margin, this revenue translates to \$5.6m-\$8.5m in profits before fixed cost. One could do a similar calculation for other markets such as Oklahoma City, which is on the smaller side of the second market category, with a population of 1.2m. Larger stations in this market generate roughly \$0.84m and \$1.4m in profits before the fixed cost. This profit compares to \$2.4m of fixed cost without any synergies. Note the above calculations do not include unobserved sources of revenue captured in payoff shocks ψ_k^t and ϕ_k^t. However, low margins are still consistent with one-digit and sometimes negative median industry EBIT margins reported by the Review of the Radio Industry published by the Federal Communications Commission in 2001.

	Specification 1	Specification 2
Fixed cost		
Fixed cost θ_m^{FIX}, markets >2.5M pop.	10.44*** (2.25)	12.64*** (1.91)
Fixed cost θ_m^{FIX}, markets 1M-2.5M pop.	1.98 (1.21)	2.40** (0.98)
Fixed cost θ_m^{FIX}, markets 0.5M-1M pop.	1.16** (0.46)	1.47*** (0.34)
Fixed cost θ_m^{FIX}, markets <0.5M pop.	0.00 (0.02)	0.00 (0.12)
Fixed cost efficiencies		
2nd station scale economy θ_1^{SCALE}	0.44* (0.42)	0.04*** (0.28)
8th station scale economy θ_2^{SCALE}	1.00 (0.30)	1.00 (0.00)
Within-format synergy θ^{SYN}	.38** (.29)	-
Switching costs		
Switching cost $\theta_m^{REPCOST}$, markets >2.5M pop.	55.67** (23.96)	55.63*** (5.08)
Switching cost $\theta_m^{REPCOST}$, markets 1M-2.5M pop.	10.55 (14.70)	10.56*** (2.66)
Switching cost $\theta_m^{REPCOST}$, markets 0.5M-1M pop.	6.19 (5.42)	6.45*** (1.23)
Switching cost $\theta_m^{REPCOST}$, markets <0.5M pop.	0.00 (0.29)	0.00 (1.00)
Payoff shocks		
Merger payoff shock θ^{ϕ}	1.36*** (0.48)	1.37*** (0.45)
Format switching payoff shock θ^{ψ}	0.00** (0.00)	0.00 (0.01)

Table 11: Second stage estimates. First section contains baseline fixed cost of owning a single station. Second section presents the estimates of the cost curve. Fixed cost efficiencies are tested if statistically different from 1 (no cost efficiencies). Third section shows the estimates of the switching cost. Last section contains standard deviations of the action-specific payoff shocks.

Second section of the Table 11 presents the estimates of the cost-function parameters.[18] I find that extensive efficiencies of operating multiple stations are when operating few stations; however,

[18]In the estimation, I restricted the estimates of a marginal fixed cost of adding a station to being less than or equal to the cost of the first station θ^{FIX}. This restriction means that I prohibit diseconomies of scale when a company owns a large number of stations. The available data variation does not allow a test against diseconomies of scale on the margin when a company owns large stations. The procedure with an unrestricted parameter value produces a large value for a θ_2^{SCALE} with a large standard error. In effect, I can only test for economies of scale against constant returns to scale on the margin when a company owns a large number of stations.

as the portfolio grows, such savings vanish. This finding is consistent with the results of O'Gorman and Smith (2008), who find a similar relationship for the radio industry using a static model. At the same time, I find large within-format cost synergies; namely, operating an extra station in the already-owned format costs more than 60% less.

Third section of the Table 11 presents format-switching costs. These numbers are fairly large, consistent with but larger than the findings of Sweeting (2011). Such repositioning costs can justify some of the behavior found when analyzing the merger probabilities; namely, stations tend to stay away from purchasing the formats they already have. If the format-switching costs were low, purchasing stations close to one's portfolio to get rid of competition, and repositioning those stations to avoid cannibalization could be optimal if cannibalization is higher than within-format synergy. However, if the switching costs were high, purchasing stations farther away to avoid paying for repositioning might be optimal. The previous subsection and Sweeting (2010) present evidence of the latter type of behavior, reinforcing the finding of high switching-cost estimates.

The last section of the Table 11 contains the estimates of variances of payoff shocks. The standard deviation of an unobserved one-time merger revenue/cost distribution is estimated to be about 130% of an average radio owner's per-period revenue. Note that those payoffs represent an aggregate value of an expected stream of unobservables. Under the assumption that the station would be held forever, one could compute a rough per-year percentage to be (1-0.95)*130%=6.5%, which measures the extent of selection on unobservables during the merger process. I note that the standard deviation of the format-switching shock, which is a nuisance parameter, is likely to be underestimated which is a results of the particular choice of the inequalities. In order to investigate this issue further I perform robustness checks with alternative set of inequalities (see online appendix) and find that the results are qualitatively and quantitatively robust.

Table 12 presents the interpretation of cost efficiency parameters θ^{SCALE}. I computed average cost per station as a function of the portfolio and cumulative cost. I cannot reject the premise that the average cost per station is flat, but I can reject that it is equal to the cost of operating one station. Large cost efficiencies early on suggest the presence of a structural difference in efficiency between companies owning one station and companies owning multiple stations. Such a difference might be a result of the family companies usually owning one station and corporations owning multiple stations. Because wage schedules and management practices among these two

	Number of stations owned							
	1	2	3	4	5	6	7	8
Specification 1: Average cost	100.0% (-)	71.8% (18.0)	65.5% (21.7)	64.7% (21.9)	66.2% (21.0)	68.7% (19.8)	71.8% (18.6)	75.3% (17.9)
Specification 1: Cumulative cost	100.0% (-)	143.6% (36.0)	196.6% (65.0)	259.0% (87.6)	330.8% (105.0)	412.0% (118.5)	502.6% (130.4)	602.6% (143.6)
Specification 2: Average cost	100.0% (-)	52.0% (13.9)	41.3% (17.0)	40.0% (17.4)	42.4% (16.7)	46.6% (15.4)	52.0% (13.9)	57.9% (12.1)
Specification 2: Cumulative cost	100.0% (-)	104.0% (27.8)	124.0% (51.0)	159.9% (69.5)	211.9% (83.4)	279.8% (92.6)	363.7% (97.2)	463.5% (97.2)

Standard errors (full parametric bootstrap) in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 12: Second stage: Implied marginal operation cost of a last station and cumulative operation cost. To obtain yearly operation costs in millions of dollars, multiply by θ_F^m.

firms are likely to be different, one can also expect large fixed-cost differences. The action to consolidate ownership of two stations therefore can be interpreted as an action to incorporate and to commercialize.

Next, I calculate the economic significance of the second-stage estimates. I compare average, post-1996 Telecom Act, per-year cost savings for the whole country with the decreased advertiser surplus computed by Jeziorski (2013) (listener surplus increased after 1996). Because for many markets the pre-1996 ownership caps were binding, this calculation can be regarded as a simple counterfactual that evaluates the impact of deregulation on total surplus. According to Specification 1, mergers that occurred after 1996 provided an additional \$1,192m (with a standard error of \$618m) of fixed-cost savings (about 5% of total industry revenue). These savings outweigh the \$223m decrease in advertiser surplus by about \$1b per year (the difference is significant with a 1-tail 10% test). Therefore, one could conclude that despite the dead-weight loss from the drop in advertiser surplus, the post-1996 merger wave increased total surplus. The answer does not qualitatively change if we look at Specification 2, which implies \$987m (std. error \$382m) fixed-cost savings.

6 Conclusions

This article proposes an estimator of a production-cost curve that enables the identification of cost efficiencies from mergers. The estimation uses inequalities representing an equilibrium of a dynamic game with endogenous mergers and product-repositioning decisions.

The biggest advantage of this estimator is that it enables the identification of the cost curve just from merger decisions, without using cost data. Therefore, it provides a tool for policy makers to improve their merger assessments if reliable cost-side data are unavailable. It can also serve as a robustness check if the alternative cost-side estimates are accessible. The policy makers can use the estimates for retrospective merger analysis, as well as to compute cost savings from future mergers.

Because the proposed method is based on a fully dynamic framework, it provides more robust estimates of cost efficiencies than the static merger analysis. For example, the dynamic model allows correction for follow-up mergers and merger waves. Additionally, endogenizing product characteristics enables correction for post-merger product repositioning, which produces more robust estimates of within-format cost synergies.

The estimator belongs to a class of indirect estimators proposed by Hotz, Miller, Sanders, and Smith (1994) and Bajari, Benkard, and Levin (2007). Therefore, it shares all the benefits of those estimators, such as conceptual simplicity of implementation and computational feasibility, because it avoids the computation of an equilibrium. However, it also shares their downsides, such as a loss in efficiency.

I apply the method to analyze the cost-side benefits of a deregulation of the U.S. radio industry. I find the consolidation wave in that industry between 1996 and 2006 provided substantial cost synergies. The total cost savings from mergers after 1996 amount to about $1 billion, which outweighs the $223m loss of advertiser surplus caused by the increased market power. Such increase in total surplus provides an argument for the supporters of the deregulation bill, and serves as an example of how cost-curve estimation can provide additional insights supplementing traditional merger analysis.

References

ACKERBERG, D., X. CHEN, AND J. HAHN (2012): "A practical asymptotic variance estimator for two-step semiparametric estimators," *Review of Economics and Statistics*, 94(2), 481–498.

AGUIRREGABIRIA, V., AND P. MIRA (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75(1), 1–53.

AI, C., AND X. CHEN (2007): "Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables," *Journal of Econometrics*, 141(1), 5 – 43, Semiparametric methods in econometrics.

AKKUS, O., AND A. HORTACSU (2007): "The Determinants of Bank Mergers: A Revealed Preference Analysis," Discussion paper.

ARCIDIACONO, P., AND R. A. MILLER (2011): "Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity," *Econometrica*, 79(6), 1823–1867.

A.RICHTER, W. (2006): *Radio: Complete Guide to the Industry*. Peter Lang Publishing.

ASHLEY HERWEG, G. H. (1997): *Radios Niche Marketing Revolution FutureSell*. Focal Press.

BAJARI, P., C. L. BENKARD, AND J. LEVIN (2007): "Estimating dynamic models of imperfect competition," *Econometrica*, 75(5), 1331–1370.

BEDNARSKI, A. (2002): "From diversity to duplication: Mega-mergers and the failure of the marketplace model under the Telecommunications Act of 1996," *Fed. Comm. LJ*, 55, 273.

BENKARD, C. L., A. BODOH-CREED, AND J. LAZAREV (2008): "Simulating the Dynamic Effects of Horizontal Mergers: U.S. Airlines," Discussion paper, Stanford University.

BERNILE, G., E. LYANDRES, AND A. ZHDANOV (2012): "A theory of strategic mergers," *Review of Finance*, 16(2), 517–575.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–90.

BERRY, S. T., AND J. WALDFOGEL (1999): "Free entry and social inefficiency in radio broadcasting," *The RAND Journal of Economics*, 30(3), 397–420.

BERRY, S. T., AND J. WALDFOGEL (2001): "Do Mergers Increase Product Variety? Evidence From Radio Broadcasting," *The Quarterly Journal of Economics*, 116(3), 1009–1025.

CRA INTERNATIONAL (2007): "Expost Merger Review: An Evaluation of Three Competition Bureau Merger Assessments," Discussion paper, CRA International.

DEPARTMENT OF JUSTICE (1997): "Horizontal Merger Guidelines," Discussion paper, Department of Justice.

DORASZELSKI, U., AND M. SATTERTHWAITE (2010): "Computable Markov-perfect industry dynamics," *The RAND Journal of Economics*, 41(2), 215–243.

DRUSHEL, B. E. (1998): "The Telecommunications Act of 1996 and Radio Market Structure," *Journal of Media Economics*, 11.

Dubé, J.-P., J. T. Fox, and C.-L. Su (2012): "Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation," *Econometrica*, 80(5), 2231–2267.

Ellickson, P., and A. Beresteanu (2005): "The Dynamics of Retail Oligopolies," 2005 Meeting Papers 829, Society for Economic Dynamics.

Ericson, R., and A. Pakes (1995): "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62(1), 53–82.

Federal Communications Commision (2008): "FCC's Review of the Broadcast Ownership Rules," `http://www.fcc.gov/cgb/consumerfacts/reviewrules.html (accessed May 7, 2008)`.

Gal-Or, E., and A. Dukes (2006): "On the Profitability of Media Mergers," *Journal of Business*, 79(2), 489–526.

Gandhi, A., L. Froeb, S. Tschantz, and G. J. Werden (2008): "Post-Merger Product Repositioning," *The Journal of Industrial Economics*, 56(1), 49–67.

Gowrisankaran, G. (1999): "A Dynamic Model of Endogenous Horizonal Mergers," *RAND Journal of Economics*, 30(1), 56–83.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–161.

Hotz, V. J., R. A. Miller, S. Sanders, and J. Smith (1994): "A Simulation Estimator for Dynamic Models of Discrete Choice," *The Review of Economic Studies*, 61(2), 265–289.

Ivaldi, M., and F. Verboven (2005): "Quantifying the effects from horizontal mergers in European competition policy," *International Journal of Industrial Organization*, 23(9-10), 669–691.

Jeziorski, P. (2013): "Effects of mergers in two-sided markets: Examination of the U.S. radio industry," Discussion paper, UC Berkeley.

Klein, J. I. (1997): "DOJ Analysis of Radio Mergers," *speech delivered in Washington DC*, http://www.usdoj.gov/atr/public/speeches/1055.pdf.

Krattenmaker, T. G. (1996): "The Telecommunications Act of 1996," *Connecticut Law Review*, 29, 123–174.

Leeper, S. E. (1999): "Game of Radiopoly: An Antitrust Perspective of Consolidation in the Radio Industry," *Fed. Comm. L.J.*, 52(473).

Menard, S. W. (2002): *Applied logistic regression analysis*. SAGE.

Nevo, A. (2000): "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," *RAND Journal of Economics*, 31(3), 395–421.

Newey, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62(6), 1349–82.

Nocke, V. (2008): "mergers, endogenous," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E. Blume. Palgrave Macmillan, Basingstoke.

O'Gorman, C., and H. Smith (2008): "Efficiency Gain from Ownership Deregulation: Estimates for the Radio Industry," CEPR Discussion Papers 6699, C.E.P.R. Discussion Papers.

Pakes, A., M. Ostrovsky, and S. Berry (2007): "Simple Estimators for the Parameters of Discrete Dynamic Games (With Entry/Exit Examples)," *The RAND Journal of Economics*, 38(2), pp. 373–399.

Park, M. (2011): "Understanding Merger Incentives and Outcomes in the US Mutual Fund Industry," .

Pesendorfer, M. (2005): "Mergers Under Entry," *RAND Journal of Economics*, 36(3), 661–679.

Pinkse, J., and M. E. Slade (2004): "Mergers, brand competition, and the price of a pint," *European Economic Review*, 48(3), 617 – 643.

Rodrigues, V. (2001): "Endogenous mergers and market structure," *International Journal of Industrial Organization*, 19(8), 1245–1261.

Rosse, J. N. (1967): "Daily Newspapers, Monopolistic Competition, and Economies of Scale," *The American Economic Review*, 57(2), 522–533.

——— (1970): "Estimating Cost Function Parameters without Using Cost Data: Illustrated Methodology," *Econometrica*, 38(2), 256–75.

Rubinstein, A. (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica*, 50(1), pp. 97–109.

Ryan, S. P. (2012): "The costs of environmental regulation in a concentrated industry," *Econometrica*, 80(3), 1019–1061.

Ryan, S. P., and C. Tucker (2011): "Heterogeneity and the dynamics of technology adoption," Discussion paper, National Bureau of Economic Research.

Rysman, M. (2004): "Competition between Networks: A Study of the Market for Yellow Pages," *The Review of Economic Studies*, 71(2), 483–512.

Salant, S. W., S. Switzer, and R. J. Reynolds (1983): "Losses from Horizontal Merger: The Effects of an Exogenous Change in Industry Structure on Cournot-Nash Equilibrium," *The Quarterly Journal of Economics*, 98(2), 185–199.

Severini, T. A., and J. G. Staniswalis (1994): "Quasi-likelihood Estimation in Semiparametric Models," *Journal of the American Statistical Association*, 89(426), pp. 501–511.

Stahl, J. (2010): "A dynamic analysis of consolidation in the broadcast television industry," Working paper, Federal Reserve Board.

Sweeting, A. (2009): "The strategic timing of radio commercials: An empirical analysis using multiple equilibria," *RAND Journal of Economics*, 42(2).

——— (2010): "The Effects of Horizontal Mergers on Product Positioning: Evidence from the Music Radio Industry," *RAND Journal of Economics*, 40(4).

——— (2011): "Dynamic Product Positioning in Differentiated Product Industries: The Effect of Fees for Musical Performance Rights on the Commercial Radio Industry," Discussion paper, Duke University.

Williamson, O. (1968): "Economies as an antitrust defense: The welfare tradeoffs," *The American Economic Review*, 58(1), 18–36.

Yi, S.-S. (1997): "Stable Coalition Structures with Externalities,," *Games and Economic Behavior*, 20(2), 201 – 237.

A Radio acquisition and format-switching algorithms

This section contains detailed flows of the algorithms used to simulate the value function from section 5.

Algorithm 1: Merger algorithm

Let $\omega_1^r = s^r$;
foreach *firm* k *in a sequence* $I(s^r)$ **do**
    Let J_{-k} be a set of stations not owned by k sorted by ξ_j^r;
    foreach *station* j *in* J_{-k} **do**
        Set purchase price $P_{jk}^r = \bar{P}^m$;
        Compute acquisition probability $\widehat{\text{Prob}}^M(\omega_k^r, d^t)$;
        Draw a random number u from $U[0,1]$;
        if $u \leq \widehat{\text{Prob}}^M$ **then**
            Increase $A_{\text{old owner}}^r$ by $\beta^{r-t} P_{jk}^r$;
            Decrease A_k^r by $\beta^{r-t} P_{jk}^r$;
            Update ω_k^r for acquisition;
            Increase B_k^r by $\beta^{r-t} E[\phi|\text{acquisition}]$;
        end
    end
    Let $\omega_{k+1}^r = \omega_k^r$;
end

Algorithm 2: Format-switching algorithm

Let $\tilde{\omega}_1^r = \omega_{K+1}^r$;
foreach *firm* k *in a sequence* $I(s^r)$ **do**
    Let J_k be a set of stations owned by k sorted by ξ_j^r;
    foreach *station* j *in* J_k **do**
        Compute repositioning probabilities $\widehat{\text{Prob}}_k^R(\tilde{\omega}_k^r, d^r)$;
        Simulate the future characteristic f_j^{r+1};
        Increase C_k^r by $\beta^{r-t} E[\psi|f_j^r]$;
        if *the* f_j *changed* **then**
            Update $\tilde{\omega}_k^r$;
            Remember the repositioning for a computation of D_k^r ;
        end
    end
    Let $\tilde{\omega}_{k+1}^{tm} = \tilde{\omega}_k^{tm}$;
end