Estimation of cost synergies from mergers without cost data: Application to U.S. radio *

Przemysław Jeziorski †

April 15, 2011

Abstract

This paper develops a new way to estimate fix cost synergies from mergers without using cost data. The estimator uses a structural model in which companies play a dynamic game with endogenous mergers and product repositioning decisions. This formulation has several benefits over the standard static merger analysis. In particular, it corrects for a sample selection of more profitable mergers and captures follow-up mergers as well as a post-merger product repositioning. The basic idea behind the estimator is treating mergers as endogenous. This allows for a comparison of mergers observed in the data with counterfactual ones based on simulated long-run gains for different levels of cost synergies. A bounds estimator is produced which, depending on variation in the data, provides partial or pointwise identification of the parameters. The framework is applied to estimate cost efficiencies after the de-regulation of U.S. radio in 1996. I find that average cost savings (across all owners) over 1996-2006 amount to \$0.5M (22%) per owned station. Total cost savings only from mergers after 1996 amount to about \$1B per year. This outweighs the loss of advertiser surplus caused by increased market power,which has been estimated to be \$223M by Jeziorski (2011).

*I would like to thank: Lanier Benkard and Peter Reiss, Joseph Harrington, Elena Krasnokutskaya, Ilya Segal, Alan Sorensen, Andrew Sweeting, Benjamin Van Roy and Ali Yurukoglu, as well as participants of numerous seminars.

†Department of Economics, Johns Hopkins University

1 Introduction

The extent to which a potential merger[1] generates cost efficiencies is often mentioned by managers as a major motivation to merge. Moreover, potential fixed cost savings generated by a merger are recognized by the Horizontal Merger Guidelines as a factor that can benefit both consumers and firms. Thus, for antitrust purposes one should evaluate cost savings in addition to measuring the decrease in competition. However, this approach is rarely used in practice, because in most cases reliable cost data is unavailable. This paper provides a solution to this problem, by proposing a method to estimate cost synergies without using data on cost. The method treats observed merger decisions as endogenous and compares them with counterfactual decisions that are based on simulated long-run gains from mergers for different levels of cost synergies. When the model predicts a merger but one is not seen in the data, I infer that the presumed cost synergies are too large. In the opposite case, when the model predicts no merger, but one is observed, I infer that presumed cost synergies are too small. The model produces a bound estimator which identifies the parameters partially or pointwise depending on variation in the data .

The proposed bound estimator of causes of ownership consolidation relies on long-run predictions of gains from merger. Estimating such gains requires a dynamic model in which both mergers and product characteristics are endogenous. However, most past empirical work analyzes mergers in a static framework and treats market structure as given. Papers by Nevo (2000), Pinkse and Slade (2004) and Ivaldi and Verboven (2005) exogenously impose changes in market structure on a static equilibrium model and calculate counterfactual changes in prices and welfare. These models are very useful in addressing the short run impacts of mergers but do not account for changes in market structure that might happen as a result of a merger. 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. Neither of these approaches allows for estimating the supply side determinants of mergers, such as cost synergies. Furthermore, the assumption that mergers are exogenous abstracts from selection bias that may result in overestimating the cost synergies (we might pick up other unobserved components correlated with the propensity to merge). In this paper, I propose a model (in the spirit of Gowrisankaran (1999)) in which merger and reposition-

[1] In this paper terms merger and acquisition would be used exchangeably to mean any change of ownership of a part or a whole company.

ing decisions are endogenous and happen sequentially. This formulation fixes the aforementioned issues by allowing for follow-up mergers and post-merger repositioning, as well as correcting for sample selection both of merger and repositioning decisions.

The model is subsequently applied 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 had the effect of eliminating 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 radio come from the cost side. Average cost savings (across all owners) from 1996 to 2006 amount to \$0.5m (22%) per owned station. Total cost side savings amount to \$1b per year, constituting about 5% of total industry revenue. Such cost synergies are higher than the anti-competitive effects of these mergers identified by Jeziorski (2011). The fact that consolidation leads to substantial cost side synergies leads me to conclude that the Telecom Act made radio more competitive against other media, such as TV or the Internet.

On the technical side, my model shares some similarities with Sweeting (2011), which uses a dynamic oligopoly model to investigate an impact of fees for music performance rights on equilibrium format switching in radio. This question is different from the one posed in this paper and so is the applied model of the industry. The biggest difference is that while Sweeting's model allows for endogenous product repositioning, my model additionally allows for endogenous mergers. Work by Stahl (2010) analyzes cross-market cost synergies from common ownership in the TV industry in the absence of market power incentives to merge. The main difference in methodology is that Stahl models merger decisions indirectly by allowing companies to choose characteristics of their portfolios (for example total population coverage or number of owned niches) that are treated as continuous variables. The advantage of a direct model of mergers used in this paper is that it acknowledges that acquisitions affect not only portfolios of buyers but also of sellers. Additionally it allows to account for endogenous cash transfers between a buyer and a seller. Another paper on a similar topic is O'Gorman and Smith (2008). They use a static oligopoly model to estimate the cost curve in radio and find that the fixed cost savings when owning two stations is bounded between between 20% and 50% of per-station costs (I estimate this number to be about 50%). The main advantage of my approach is that it allows firms to be forward looking when making decisions to merge as well as to endogenously reposition products in response to a merger. This paper also builds on the empirical literature on demand and cost curve estimation (started with Rosse (1970) and Rosse (1967)) by accounting explicitly for the demand side incentives to merge.

To my knowledge, Gowrisankaran (1999) is the only applied paper that uses a dynamic framework to directly endogenize mergers. His analysis argues that merger dynamics are very important for incentives to invest. The main drawback of his analysis is that it was never fit to real data. This was due in part to the complexity of his model and in part to the lack of a good dataset. To solve the complexity problem, I utilize the latest developments in the dynamic-games literature. These developments allow to estimate very complicated models without explicitly solving them (Bajari, Benkard, and Levin (2004)). In the application of my method I utilize a rich dataset on radio industry containing thousands of merger and repositioning events.

This paper is organized as follows. Section 2 contains a flexible, structural merger model that can applied to many industries. The estimation procedure is discussed in Section 3. Section 4 describes the application of the framework to analyze the merger wave in the U.S. radio industry. Section 5 concludes the paper.

2 Merger and repositioning framework

This section presents the dynamic oligopoly model of an industry with differentiated products in the spirit of Ericson and Pakes (1995). The industry is modeled as a dynamic game and the players are companies holding portfolios of different products (niches). The modeling effort emphasizes the actions of companies changing the portfolio of owned products, specifically repositioning and acquisitions. The model is general enough to encompass a number of different industries and types of competition, by allowing for a large range of different single-period profit functions and cost structures.

2.1 Industry basics

The industry is composed of M different markets that operate in discrete time over an infinite horizon. The payoff relevant market characteristics at time t for market m are fully characterized by a set of covariates $d^{mt} \in \mathcal{D}$ that include demand shifters. In each market m, there are up to K_m operating firms and up to J_m active products. Let $o_j \in K_m$ be the owner of the product j. I assume that each product $j \in J_m$ is characterized by a triple $s_j^t = (f_j^t, \xi_j^t, o_j^t)$. $f_j^t \in F$ is a discrete characteristic of a product (good examples are horizontal characteristics that appeal to different demographic groups like color, design or a type of broadcast in case of radio) and $\xi_j^t \in \Xi$ is a continuous characteristic of the product (good examples are vertical characteristics like quality). I deal with entry/exit by assuming that there exists an inactive state $\bar{f} \in F$ in which the niche brings no profit but also does not cost anything to operate. The state of the industry at the beginning of each period is therefore a pair $(s^t, d^t) \in \mathcal{S} \times \mathcal{D}$.

The following assumptions restrict the dynamics of ξ.

Assumption 2.1. ξ_j^t *evolves as an exogenous Markov process.*

One example can be an AR(1) process

$$\xi_j^t = \rho \xi_j^{t-1} + \zeta^t \tag{2.1}$$

where ζ^t is a mean zero IID random variable.

Market covariates are also assumed to be exogenous and Markov.

Assumption 2.2. d^t *evolves as an exogenous Markov process.*

These two assumptions are necessary to maintain Markovian structure of the game, so that Markov Perfect Equilibrium is a Nash Equilibrium. They can be partially relaxed to encompass n-th order Markov processes by including more history in the current state (s^t, d^t).

2.2 Static payoffs and costs

Conditional on the state of the industry (s^t, d^t) each firm k gets a one shot payoff $\bar{\pi}_k(s^t, d^t)$. This payoff is in most cases an outcome of a static pricing or quantity setting game. A functional forms of a payoff function can be fairly nonrestrictive and specific assumptions that ensure the existence of a equilibrium of a dynamic game are listed in Doraszelski and Satterthwaite (2010).

Each firm k, in order to maintain its product portfolio s_k^t, has to pay a per-period fixed cost $F(s_k^t)$. Estimating the properties of F, in particular potential cost efficiencies of owning multiple products, is a central question of this paper.

2.3 Players' actions

Firms can undertake two types of actions: product acquisitions and product 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. Each repositioning action from f_j^t to f_j^{t+1} incurs a cost $C(f_j^t, f_j^{t+1})$. Firms can potentially acquire any subset of competitors' products, as well as choose characteristics of owned and newly acquired products. Additionally, these decisions can be correlated. For example, acquiring product of type f might be negatively correlated with repositioning of owned product into type f. To accommodate all these important features, in the most general setting, acquisition and repositioning actions have to be modeled as complicated set valued correspondences. However, such formulation is analytically and numerically intractable even for small industries. To solve this I put more structure on a sequence of actions.

Assumption 2.3. *Product acquisition takes place before product repositioning. Players observe the state of the market after acquisitions, before deciding about product repositioning.*

The above assumption divides each period into an acquisition and a repositioning stage and is motivated by a premise that mergers are usually harder to execute than repositioning. Moreover, an alternative to this assumption which is a simultaneous acquisition and repositioning, causes technical issues and requires additional structure for estimation. For example, joint mergers and repositioning in combination with a sequential moves by players (introduced in a following section) would allow for a product to be repositioned twice during the same sample period: by an old and new owner. Since the actual repositioning and acquisition time stamps are rarely observed, it would create ambiguity about assignment of actions to players during the estimation stage. In contrast to that, Assumption 2.3 always assigns repositioning action to the new owner. In case the actual time stamps are observed this assumption is easy to relax.

2.3.1 Acquisition stage

In this stage players can decide to acquire any set of products owned by competitors. Following Gowrisankaran (1999) I assume that, within an acquisition stage, players move according to a sequence.

Assumption 2.4. *Owners move in a sequence specified by a permutation $\sigma(s^t, d^t)$ of the active owners' index $\{1, \ldots, K\}$. σ is allowed to depend on a state however its functional form is fixed over time and markets and is common knowledge.*

The sequence $\sigma(s^t, d^t)$ is allowed to be endogenous because it changes from period to period as a function of the industry state, which is controlled by the actions of the firms. A sequential formulation of moves solves the problem of multiple companies trying to acquire the same product in the same period. This formulation is necessary to reduce the number of potential equilibria [2], and make a simulation of the dynamics of the industry computationally feasible. Since, the sequence of moves is predetermined at the beginning of the period, it resolves all the potential merger conflicts. Also, it enables me to separately simulate a move of each player conditional on the players that already moved. It is a key to computational tractability of the model.

Before making its move each player k observes the current state of the market (s^t, d^t) and acquisition actions of all players that moved before it $A_{-k}(s^t, d^t) = \{a^t_{k'} : k' < \sigma_k(s^t, d^t)\}$. Each player knows its own position in the sequence as well as the position of all the other players (because it depends only on a starting state).

In addition to (s^t, d^t, A^t_{-k}) each player observes acquisition prices that are acceptable to the seller $P^t_k \subset \mathbb{R}^J_+$ and a stochastic one-time payoff shocks of integrating any competitor product into the portfolio $\phi^t_k \subset \mathbb{R}^J$. [3] The following assumption restricts the dynamics of payoff shocks.

[2] It guarantees the uniqueness of an equilibrium is a stage game. Although it greatly reduces the number of equilibria, it does not guarantee uniqueness in the dynamic game. Because of that, identification requires extra assumptions about equilibrium selection during the estimation stage.

[3] Information sets of players consists of pieces irrelevant to company payoffs - prices P^t_{kj} and integration payoff shocks ϕ^t_{kj} of products already owned by the firm. This formulation is chosen for notational simplicity and does not introduce any new equilibria. The reason is that these variables cannot serve as an additional correlation device because ϕ^t is private information (by Assumption 2.5) and prices are deterministic and can be computed from a state (by Assumption 2.6).

Assumption 2.5. *ϕ_k^t is IID across players and time and is private information of acquirer k. The seller has no private information.*

The independence across time is made mainly to simplify the exposition and lower the data requirements to estimate the model. It could potentially be relaxed by introducing latent state variables describing persistent unobserved heterogeneity. However, in most cases, given the complicated nature of merger decisions, the available data will not have enough variation to identify the extended version of the model.

The following assumption restricts the merger bargaining process in order to make the problem Markovian.

Assumption 2.6. *Acquisition prices P_{kj}^t are determined by a static bargaining process (resolved given the current state), and are only a function of a public information available during the transaction (s^t, d^t, A_{-k}^t).*

This is equivalent to assuming that ϕ_k^t cannot be credibly signaled during the merger negotiations. If both players believe that the market is in a Markov Perfect Equilibrium, a merger surplus frontier will only depend on the current state (s^t, d^t, A_{-k}^t), because it depends only on the buyers' and sellers' value functions. If the bargaining is an outcome of a static non-cooperative solution, for example, the Nash Bargaining Solution, the price acceptable to the seller will be indeed only a function of a current state (s^t, d^t, A_{-k}^t). This shows that the model is internally consistent and compatible with the MPE solution concept. [4]

Let $\mathcal{A} = \{0,1\}^J$ be a class of all possible action sets that can be observed by a player in an acquisition stage. In a similar way define $\Phi = \mathbb{R}^J$ to be a class of all possible shocks that a player can observe. The acquisition strategy[5] is a mapping from observables to actions

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

[4] The downside of this pricing formulation is that it assumes away selection of mergers on prices. Instead it allows for selection on privately observed integration payoff shocks ϕ. In theory one can introduce a second shock P_ϵ that reflects determinants of price unobserved to the econometrician. Identification of such shocks separately from ϕ would be possible if one observed prices of all potential mergers, not only the ones that are realized (or exogenous exclusion restrictions). This would also require explicit assumptions about the bargaining procedure that take a stand on how P_ϵ is generated, and how it feeds back into payoffs of a buyer and a seller. Given the presented framework such an extension is possible; however, it is beyond the scope of this paper.

[5] I restrict my attention to Markov strategies; however, it can be shown that a Markov equilibrium is a Nash

Note that if product j already belongs to a owner k, action $a_k(j)$ does not matter for payoffs.

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. In practice this means that in addition to a current state and shocks (s^t, d^t, ϕ_k^t), actions of player k can depend only a vector A_{-k}^t describing all acquisitions by players earlier in the sequence. Formally, for any (s,d) take two actions sets A and A' such that all observable past actions are the same $A_{-k}(s,d) = A'_{-k}(s,d)$. A feasible strategy is restricted to prescribe the same action for these sets i.e. $g_k^A(s,d,A,\phi) = g_k^A(s,d,A',\phi)$ for all ϕ. Additionally, the merger strategy of a player k cannot depend on private information of other players. Formally, if ϕ and ϕ' are such that $\phi_k = \phi'_k$, actions prescribed by a feasible strategies must be the same i.e. $g_k^A(s,d,A,\phi) = g_k^A(s,d,A,\phi')$ for all s,d,A.

When all players move make their move, the acquisition stage ends, and the game proceeds to the repositioning stage.

2.3.2 Repositioning actions

In the repositioning stage I impose the same sequence of moves as in the acquisition stage (it is not necessary, however it simplifies the exposition). Before submitting a repositioning action b_k^t a player observes the current state of the market (s^t, d^t) and all acquisition actions and repositioning actions of players that moved before him $B_{-k}(s^t, d^t) = A^t \cup \{b_{k'}^t : k' < \sigma_k(s^t, d^t)\}$. In addition, players observe a set of additive shocks $\psi_k^t \subset \mathbb{R}^{JF}$ to per-period costs of maintaining any product with any characteristic.

Assumption 2.7. *ψ_k^t is IID across players and time, uncorrelated with ϕ_k^t and is private information of an owner k.*

Let $\mathcal{B} = \mathcal{A} \times J^F$ be a set of past actions that a player can observe before deciding about repositioning. Define a repositioning strategy

$$g_k^B : \mathcal{S} \times \mathcal{D} \times \mathcal{B} \times \Psi \to F^J$$

where $\Psi = \mathcal{R}^J$. The strategies do not need to explicitly depend on acquisition prices, because they are a sunk cost. Similarly, as in the definition of an acquisition strategy, a player faces the

equilibrium of a game in which we allow the actions to depend on all history of past actions and states. This critically depends on the Markovian structure of the game ensured by Assumptions 2.4 and 2.6.

informational constraint that his strategy needs to be measurable with respect to his information σ-field. For any two actions sets B and B' such that all observable past actions are the same $B_{-k}(s,d) = B'_{-k}(s,d)$, a feasible repositioning strategy is restricted to prescribe the same action i.e. $g_k^B(s,d,B,\psi) = g_k^B(s,d,B',\psi)$ for all ψ. Similarly, a player can condition only on his private information. Formally, if ψ and ψ' are such that $\psi_k = \psi'_k$, actions prescribed by a feasible strategies must be the same i.e. $g_k^B(s,d,B,\psi) = g_k^B(s,d,B,\psi')$ for all s,d,B.

2.3.3 Timing of the model

Timing of the action sequence is as follows:

1. The move sequence $\sigma(s^t,d^t)$ is determined. k is set to 1.
2. Firm σ_k starts its acquisition action. Shocks $\phi^t_{\sigma_k}$ are realized and actions $a^t_{\sigma_k}$ are revealed to everyone. Acquisition prices are transferred.
3. The process goes back to step 2 allowing $k+1$ firm in a sequence to move. If there are no more firms the game proceeds to the repositioning stage and k is reset to 1.
4. Firm σ_k starts its repositioning action. Shocks $\psi^t_{\sigma_k}$ are realized and actions $b^t_{\sigma_k}$ are revealed to everyone.
5. The process goes back to step 4 allowing $k+1$ firm in a sequence to move until there are no more firms.
6. New state (s^{t+1},d^{t+1}) is revealed and one shot payoffs $\pi(s^{t+1},d^{t+1})$ as well as fixed and repositioning costs are realized.
7. The game proceeds to the next period.

2.4 Equilibrium

Let $\mathbf{g} = (g_1^A,\ldots,g_K^A,g_1^B,\ldots,a_K^B)$ be a stationary Markov strategy profile. It can be shown that this profile and an initial condition (s^0,d^0) determine the unique, controlled Markov process over states (s^t,d^t), actions a^t, acquisition prices P^t, and payoff shocks (ψ^t,ϕ^t).

$$\mathcal{P}(\mathbf{g},s^0,d^0) \in \Delta(\mathcal{S}\times\mathcal{D}\times\mathcal{A}\times P\times\Psi\times\mathcal{B}\times\Phi\times\mathcal{T})$$

where $\mathcal{T}$ is a time horizon, and Δ is a set of probability measures. $\mathcal{P}$ is therefore a discrete time stochastic process on $\mathcal{S} \times \mathcal{D} \times \mathcal{A} \times P \times \Psi \times \mathcal{B} \times \Phi$. 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

$$\begin{aligned} \pi_k(s^t, s^{t+1}, d^{t+1}, P^t, \psi^t, \phi^t) = \bar{\pi}_k(s^{t+1}, d^{t+1}) - F(s_k^{t+1}) + \sum_{j: o_j^t \neq k, o_j^{t+1} = k} (\phi_{kj}^t - P_{kj}^t) + \\ + \sum_{j: o_j^t = k, o_j^{t+1} \neq k} P_{o_j^{t+1} j}^t + \sum_{j: o_j^{t+1} = k} \left[\psi_{kj f_j^{t+1}}^t - \mathbf{1}(f_j^{t+1} \neq f_j^t) C(f_j^t, f_j^{t+1}) \right] \end{aligned} \tag{2.2}$$

where a third term is composed of outcoming cash flows resulting from acquisitions, fourth term of incoming cash flows from selling products, and the last term of cash flows from repositioning.

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

$$V_k(s, d | \mathbf{g}_k, \mathbf{g}_{-k}) = E_{\mathcal{P}(\mathbf{g}, s, d)} \sum_{t=0}^{\infty} \beta^t \pi_k(s^t, s^{t+1}, d^{t+1}, P^t, \psi^t, \phi^t) \tag{2.3}$$

It is assumed that the markets are in a Markov Perfect Equilibrium, i.e., firms choose strategy profile $\mathbf{g}^*$, such that for all k

$$V_k(s, d | \mathbf{g}_k^*, \mathbf{g}_{-k}^*) \geq V_k(s, d | \mathbf{g}_k, \mathbf{g}_{-k}^*) \quad \forall \mathbf{g}_k. \tag{2.4}$$

For simplicity, I restrict my attention to symmetric equilibria. The next section describes the estimation procedure.

3 Estimation

Consider parameterizations of a fixed cost $F(s_k | \theta^{FIX}, \theta^{SYN})$ and repositioning costs $C(f, f' | \theta^{REPCOST})$. To facilitate identification I make a series of simplifying assumptions about the structure of a fixed cost function. The estimation procedure does not crucially depend of these assumptions, hence they could be easily modified to fit a particular application, as long as one could argue identification.

Let n_k be a number of products owned by an owner k, that is prescribed by s_k.

Assumption 3.1. *Let C be a cost function such that for all s:*

(a) *Operating only one product in an inactive state does not cost anything (equivalent of exit). Formally, for s_k such that $n_k = 1$ and $f_j = \bar{f}$ if j is owned by k,*

$$F(s_k|\theta^{FIX}, \theta^{SYN}) = 0, \quad \forall \theta^{FIX}, \theta^{SYN} \tag{3.1}$$

(b) *Cost of operating one niche does depend on θ^{SYN}, i.e. for any s_k such that $n_k = 1$*

$$F(s_k|\theta^{FIX}, \theta^{SYN}) = F(s_k|\theta^{FIX}, \bar{\theta}^{SYN}), \quad \forall \theta^{FIX}, \theta^{SYN}, \bar{\theta}^{SYN} \tag{3.2}$$

(c) *Cost is increasing in θ^{FIX}, i.e.*

$$\bar{\theta}^{FIX} > \theta^{FIX} \Rightarrow F(s_k|\bar{\theta}^{FIX}, \theta^{SYN}) > F(s_k|\theta^{FIX}, \theta^{SYN}) \tag{3.3}$$

and is sufficiently unbounded, i.e. there exists θ^{FIX} such that $F(s_k|\theta^{FIX}, \theta^{SYN}) > \pi(s_k, d)$ for all s_k, d.

(d) *Additional cost of extra product beyond 1 is deceasing in θ^{SYN}. Formally, if $\bar{\theta}^{SYN} > \theta^{SYN}$, $n_k > 1$ and $n'_k = n_k + 1$*

$$F(s'_k|\theta^{FIX}, \bar{\theta}^{SYN}) - F(s_k, \theta^{FIX}, \bar{\theta}^{SYN}) < F(s'_k|\theta^{FIX}, \theta^{SYN}) - F(s_k|\theta^{FIX}, \theta^{SYN}) \tag{3.4}$$

Parameter θ^{FIX} determines the level of fixed cost, and θ^{SYN} determines the level cost synergies. This section outlines a procedure to obtain consistent estimators $(\hat{\theta}^{FIX}, \hat{\theta}^{SYN})$ and $\hat{\theta}^{REPCOST}$. The procedure has two stages. In the first stage, I estimate the equilibrium strategies, dynamics of exogenous state variables and acquisition pricing equation. In the second stage, I estimate cost parameters by imposing the dynamic game equilibrium moment inequalities in (2.4) using a simulated value function.

3.1 Basic setup

Data is given by the set $X = \{x^{tm} : 1 \leq m \leq M, 1 \leq t \leq T\}$. Each point in the data 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 [6] in market m and time t. The data does not have to contain any direct information on the cost. This is convenient since in many cases the data on cost suffers from accounting issues or direct cost estimates are unavailable.

To facilitate the inference process I make an assumption that the data is generated by a single MPE strategy profile $\mathbf{g}^*$. Moreover, the dataset needs to contain a reasonable amount of within market acquisitions and repositioning in order to provide enough information about equilibrium merger and repositioning strategies. Since for most industries such data sets are unavailable, it is possible to pool similar industries to construct one dataset. However, this requires a stronger assumption that equilibrium behavior is the same across the pooled industries.

In order to simplify the exposition all state variables are assumed to be observed. In many cases one can infer the unobserved state variable from a static estimation of a one-shot profit function $\bar{\pi}$. One example of such a case is the Berry, Levinsohn, and Pakes (1995) estimator, which uses variation of static market shares that cannot be explained by observables to identify unobserved product quality. Moreover, there are numerous ways to proceed in case one cannot directly infer all the latent state variables. For example, one could extend the procedure from this paper with an algorithm proposed by Arcidiacono and Miller (2010).

3.2 First step

The first step of the procedure is constructing three auxiliary data sets using a sequential structure of the acquisition and repositioning process. For each t, the predefined sequence of player moves $\sigma(s_t, d^t)$ specifies a mapping

$$(s^t, s^{t+1}) \overset{\sigma}{\mapsto} (s^t, a^t_{\sigma_1}, \ldots, a^t_{\sigma_k}, \ldots, a^t_{\sigma_K}, b^t_{\sigma_1}, \ldots, b^t_{\sigma_k}, \ldots, b^t_{\sigma_K})$$

This information can be used to construct a data set of acquisitions prices and its determinants

$$X^P = \{(s^{tm}, d^{tm}, P^t_k, A^t_{-k}) : k \in K^m, 1 \le m \le M, 1 \le t \le T\}$$

The data set X^P is used to estimate an equilibrium price $P_j(s^{tm}, d^{tm}, A^t_{-k}|\theta^P)$ of station j. Because of Assumption 2.6, equilibrium pricing does not depend on the components unobserved by the

[6] The transaction prices are helpful but not necessary to identify the cost parameters. Estimation without prices requires more assumptions about the merger bargaining process, as well as a large amount of additional computing power. The extra steps needed to proceed without the prices are mentioned in Appendix A.

econometrician; therefore it can be inferred from observed prices and variation in (s^t, d^t, A^t_{-k}). Next, I construct a data set of acquisition actions

$$X^A = \{(s^{tm}, d^{tm}, a^t_k, A^t_{-k}) : k \in K_m, 1 \leq m \leq M, 1 \leq t \leq T\}$$

and repositioning actions

$$X^B = \{(s^{tm}, d^{tm}, b^t_k, B^t_{-k}, A^{tm}) : k \in K_m, 1 \leq m \leq M, 1 \leq t \leq T\}$$

Equilibrium strategies are not estimated directly. Instead I estimate conditional choice probabilities (CPP) for mergers

$$\text{Prob}^A(a_k | s^{tm}, d^{tm}, A^t_{-k}) \in \Delta(\{0,1\}^J)$$

and repositioning

$$\text{Prob}^B(b_k | s^{tm}, d^{tm}, A^{tm}, B^t_{-k}) \in \Delta(J^F)$$

conditional on the information observed by a player. With the perfect data these CCPs can be identified in a fully non-parametric way. In a finite sample, an econometrician is likely to face two major problems: large dimensionality of information sets and action spaces $\{0,1\}^J$, F^J.

In the equilibrium, CCPs depend on the distributions of unobservables ψ and ϕ and differences between conditional value functions for available actions. Therefore, a dependence of the CCP on a large dimensional state space is described by an index function (which is a value function, because the decisions depend on the state only through the value function), whose exact form is unknown. To deal with this, one could construct a semi-parametric[7] estimator by making distributional assumptions about ψ and ϕ and allowing for a non-parametric single index. In practice this means using flexible parametric functions[8] $\widehat{\text{Prob}}^A(a_k | s^t, A^t_{-k}, d^t, \theta^{AQ})$ and $\widehat{\text{Prob}}^B(b_k | s^t, B^t_{-k}, A^t, d^t, \theta^{REP})$. The asymptotics of such estimators (as the size of a dataset and a dimensionality of pseudo-parameter vector goes to infinity) is well behaved by Newey (1994) and was studied in a context of dynamic games by Bajari, Chernozhukov, Hong, and Nekipelov (2009).

[7]The original BBL estimator does not permit continuous state space and non-parametric first stage. Even though ξ is modeled as continuous in a theory part, in practice I discretize it. It is due to limitations in how computers store real numbers.

[8]An alternative method would be the semi-parametric approach using quasi-likelihood proposed by Severini and Staniswalis (1994).

Large dimensionality of an action space, in most cases, cannot be handled without additional assumptions. In a true equilibrium conditional acquisition and repositioning actions are correlated. For example, a company's acquisition action of a product with a characteristic f might be negatively correlated with acquisition of another product with characteristic f. For a large number of products, the full covariance matrix is infeasible to estimate and the researcher will be forced to make compromises. In the remainder of this section, I propose alternative sets of assumptions that make this problem tractable while keeping the model internally consistent. The proposed solutions are only one option to follow and in practice can be modified according to the demands of the particular application.

The following three examples show how to handle correlations between actions of the same type for different types of industries.

Example 3.1 (Independent actions)**.** *Before making decisions whether to acquire a product j, a company observes only ϕ_{kj}. When making repositioning decisions for product j, a company observes only vector ψ_{kj}. In this case where the acquisition decisions are uncorrelated conditional on the state space, that allows for direct application of discrete choice estimators.*

The example applies to industries with large, decentralized companies that make many small and frequent acquisitions and repositioning decisions. The next example is intended for large and infrequent actions.

Example 3.2 (One action per period)**.** *If the data suggests that the acquisitions and repositioning are fairly rare, one could allow only one acquisition and repositioning per player each period. This reduces the decision space for each type of action into one dimension.*
In particular, if $\phi^t_{jk} = \epsilon^k_{jk} - \epsilon^k_{0k}$, where epsilons are independent extreme value random variables, acquisitions can be consistently estimated using a multinomial logit with a non-parametric index. If $\psi^t_{jkf} = \bar{\epsilon}^t_{jk} + \bar{\epsilon}^t_{jkf}$, where $\bar{\epsilon}^t_{jk}$ is Gumbel and $\bar{\epsilon}^t_{jkf}$ is extreme value distributed, repositioning actions can be estimated as a nested logit (with a non-parametric index), in which a nest is a product chosen to be repositioned.

This method could be used if the acquisitions and repositioning are fairly rare. In such cases it might be the only way to proceed, since limited data will not provide evidence for more complicated correlation patterns.

The next example combines the large and small action approach and is intended for the industries that are in the middle. It makes more assumptions about the structure of the acquisition and repositioning decision making within the firm.

Example 3.3 (Sequential actions)**.** *Suppose that the acquisition decisions are made in a sequence, i.e., after observing ψ_{kj} for a particular product, the firm decides about its acquisition without knowing shocks for other products higher in a sequence. However, it conditions on the shocks for products lower in a sequence as well previous acquisition decisions in the same period. It allows for acquisition decisions of multiple products by the same firm in the same time period to be correlated. The same applies to repositioning and shocks ϕ_{kj}. In this case, the estimation is similar to the one in Example 3.2; however, the non-parametric index depends in addition on all previously made decisions. Under this assumption $\vec{\epsilon}^{t}_{jk}$ drops out because of a sequential structure and repositioning can be consistently estimated as a multinomial logit instead of a nested logit.*

An example of practical implementation of Example 3.3 is contained in Section 4. In the next subsection I describe a second stage of the cost function estimator that uses the estimators of acquisition pricing and equilibrium strategies. I also assume that one can consistently estimate the exogenous Markov processes describing the dynamics of ξ and d^t.

3.3 Minimum distance estimator

For the second stage the parameters of the fixed cost $(\theta^{FIX}, \theta^{SYN})$ and repositioning cost $\theta^{REPCOST}$ are estimated using a minimum distance estimator. The estimator is constructed using the MPE inequalities (2.4). The remainder of this section describes how I obtain estimates of the value functions in those inequalities.

The value function V_k (defined in equation (2.3)) 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

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

is the expected stream of revenues,

$$B_k^t = E\sum_{r=t}^{\infty}\beta^{r-t}\sum_{j:o_j^r\neq k, o_j^{r+1}=k}\phi_{kj}^r$$

is the expected stream of acquisition payoff/cost shocks,

$$C_k^t = E\sum_{r=t}^{\infty}\beta^{r-t}\sum_{j:o_j^{r+1}=k}\psi_{kjf_j^{r+1}}^t$$

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

$$D_k^t = E\sum_{r=t}^{\infty}\beta^{r-t}\left[F(s_k^r|\theta^{FIX},\theta^{SYN}) + \sum_{j:o_j^{r+1}=k}\mathbf{1}(f_j^{r+1}\neq f_j^r)C(f_j^r, f_j^{r+1}|\theta^{REPCOST})\right]$$

is the expected stream of fixed costs and repositioning costs. The extra parameters θ^ϕ and θ^ψ are needed because the first stage estimation requires normalization of the variances of ϕ and ψ.

Accounting for B_k^t in the simulation of profits from a merger takes care of selection on unobservables. 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, it would be the case that $E[\phi_{jk}^{tm}|a_{jk}^{tm} = 1] > 0$. [9]. Assuming that $E[\phi_{jk}^{tm}|a_{jk}^{tm} = 1] = E[\phi_{jk}^{tm}] = 0$ would cause underestimation of profits from mergers and could result in overestimation of fixed cost synergies. The same point can be made about the selection on unobservables when repositioning products and inclusion of C_k^t.

Note that only the last part of D_k^t depends on the parameters of interest $(\theta^{FIX}, \theta^{SYN})$ and $\theta^{REPCOST}$ and the value function is linear 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. [10] This saves a large amount of processing power and makes the estimator feasible using today's computers.

[9] If one was to adopt assumptions from Examples 3.1, 3.2 or 3.3, and adopt 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.

[10] In most cases A_k^t is the hardest to compute because computing $\bar{\pi}$ may involve solving a one-shot Nash equilibrium price or a quantity setting game.

Let V^n be an equilibrium value function, where n indexes players, states, markets and time periods. Consider C types[11] of suboptimal strategies $\mathbf{g}_c^n$. For each $c \in C$ compute a suboptimal value function $\tilde{V}_c^n(\mathbf{g}_c^n, \mathbf{g}^{-n})$, where $\mathbf{g}^{-n}$ is a equilibrium strategy for the competitors of a player prescribed by an index n.

Consider a random sample of size N of indexes (one could sample states, players, time periods, or markets). Following the equation (2.4) I define a minimum distance estimator

$$(\hat{\theta}^{FIX}, \hat{\theta}^{SYN}, \hat{\theta}^{REPCOST}, \theta^{\phi}, \theta^{\psi}) \in \operatorname{argmin} \frac{1}{N \times C} \sum_{n,c} W_c \left(\max\{\tilde{V}_c^n - V^n, 0\} \right)^2 \tag{3.5}$$

where W_c is a set of positive weights. According to the results in Bajari, Benkard, and Levin (2004) this estimator is consistent and asymptotically normal. This finishes the description of the estimator.

3.4 Identification

The econometrician needs to identify three main parameters: repositioning cost $\theta^{REPCOST}$, fixed cost level θ^{FIX} and fixed cost synergies θ^{SYN}. The key identifying equation is the equilibrium restriction (2.4) applied to different off-equilibrium counterfactual strategies. I discuss identification of θ^{FIX} and θ^{SYN} assuming that $\theta^{REPCOST}$ is identified.

It is useful to think of the identification of cost parameters by using a simple static model of mergers, because the intuition carries over to a full dynamic model. Suppose that the game has one period that consists of two-stages: merger stage and profit stage. In a merger stage every company can acquire one additional niche for price P.

Denote the starting state as s^0. Choose any company k. Let $\bar{s}^1$ and s^1 be states in the profit stage conditional respectively on the acquisition action of some active product j by a company k

[11] An introduction of multiple types of sampled off-equilibrium policies is made only to stress the argument validating identification of the model (different types identify lower and upper bound on cost synergies). The original Bajari, Benkard, and Levin (2004) setup has only one type being sampled, however it does not specify the sampling type. It makes my estimator a special case of their framework and enables me to invoke their limiting results. For the same reason, my second stage estimator is consistent even if c is smaller than the number of parameters (for the BBL case it would be $c = 1$). The intuition is contained in the proofs in the BBL paper, and amounts to thinking of using $N \times C$ randomly sampled moments.

and no acquisition action by company k. To simplify the notation denote the additional revenue from acquisition net of acquisition price as

$$\Delta\pi(s^0, d, P) = E\left[\pi(\bar{s}_k^1, \bar{s}_{-k}^1, d) - P - \pi(s_k^1, s_{-k}^1, d)|s^0\right]$$

and additional fixed cost after acquisition as

$$\Delta F(s^0|\theta^{FIX}, \theta^{SYN}) = E\left[F(\bar{s}_k^1|\theta^{FIX}, \theta^{SYN}) - F(s_k^1|\theta^{FIX}, \theta^{SYN})|s^0\right]$$

The expectations are taken with respect to acquisition actions of the opponents. I assume that the equilibrium strategies of the opponents are known or can be consistently estimated from the data. In the remainder of this section I fix these strategies. I also assume that acquisition prices P are observed. If it is not the case, P can be inferred from a profit function of a seller and a buyer. This would however require additional structure on a bargaining procedure, and might make the identification argument more complicated.

For a given (s^0, d, P) denote by $M(s^0, d, P) \in \{0, 1\}$ an observed decision of company k to acquire any active product j. Consider two types of suboptimal strategies for k: always merge and never merge. These strategies are indeed suboptimal if one observes variation in M and the market is in Nash Equilibrium. First, I discuss the identification of θ^{FIX} parameters using these suboptimal strategies. Take s^0 such that the company k owns only products in an inactive state, i.e. $f_j^0 = \bar{f}$ for all $j \in s_k^0$. Suppose there is enough variation[12] in d or P to generate observations $M(s^0, d, P) = 0$ if $(d, P) \in Z_1$ and $M(s^0, d, P) = 1$ if $(d, P) \in Z_2$. Note that in case $(d, P) \in Z_2$ it must be that $\Delta\pi(s^0, d, P) > 0$ if the model is correct.

In case of no merger $(d, P) \in Z_1$ the strategy to always merge would be suboptimal, because one does not observe a merger in the data. Nash Equilibrium requires

$$\Delta F(s^0|\theta^{FIX}, \theta^{SYN}) - \Delta\pi(s^0, d, P) \geq 0 \tag{3.6}$$

It means that the additional fixed cost generated by the merger must be greater than the additional profits. By assumption 3.1 (a) and (b) G_M does not depend on θ^{SYN} because the firm k initially owns only products in an inactive state. By assumption 3.1 (c) ΔF is increasing in θ^{FIX}. Therefore, there exists $\underline{\theta^{FIX}}$ such that condition (3.6) is satisfied for all observations in Z_1 and all $\theta^{FIX} > \underline{\theta^{FIX}}$.

[12]In case this condition is not satisfied the identified set will consist of just lower or upper bound.

Similarly for $(d, P) \in Z_2$ "never merge" strategy is suboptimal and Nash Equilibrium requires

$$\Delta\pi(s^0, d, P) - \Delta F(s^0|\theta^{FIX}, \theta^{SYN}) \geq 0 \tag{3.7}$$

It means that the additional profit generated by the merger is greater than the additional fixed costs. Using a similar reasoning based on assumption 3.1 one can obtain an upper bound on $\overline{\theta^{FIX}}$ that ensures condition (3.7). If the model is not misspecified $\underline{\theta^{FIX}} \leq \overline{\theta^{FIX}}$ and the true cost parameter θ^{FIX} must lie within the set $[\underline{\theta^{FIX}}, \overline{\theta^{FIX}}]$.

Given the identification of θ^{FIX}, we proceed to identify θ^{SYN} by rationalizing acquisitions resulting in owning multiple products. Take s^0 such that the company k owns at least one product in an active state, and any value $\theta^{FIX} \in [\underline{\theta^{FIX}}, \overline{\theta^{FIX}}]$. Suppose that there is enough variation in the (d, P) to generate non-merge events for $(d, P) \in Z_3$ and merge events for $(d, P) \in Z_4$.

For $(d, P) \in Z_3$, Nash equilibrium requires condition (3.6) to be satisfied. Given that we fixed θ^{FIX} we can use assumption 3.1 to produce an upper bound $\overline{\theta^{SYN}}$ because ΔC is decreasing in θ^{SYN}. Similarly one can produce a lower bound $\underline{\theta^{SYN}}$ for $(d, P) \in Z_4$ using the equation (3.7). As a result we have a collection of sets $[\underline{\theta^{SYN}}, \overline{\theta^{SYN}}]$ for each $\theta^{FIX} \in [\underline{\theta^{FIX}}, \overline{\theta^{FIX}}]$. The estimator of θ^{SYN} is defined as an union of that collection. It must be non-empty if the model is not misspecified and its tightness depends on the variation in the data [13]. Note that for θ^{FIX} that is not a true value, the set $[\underline{\theta^{SYN}}, \overline{\theta^{SYN}}]$ might indeed be empty. Such cases provide additional information that leads to exclusion of θ^{FIX} from an identified set $[\underline{\theta^{FIX}}, \overline{\theta^{FIX}}]$. It suggests that joint estimation of both cost parameters can lead to sharper identification and efficiency gains.

It is important to note that a full dynamic model has multiple stages and companies can acquire more than one niche at once. The theoretical complication is that $\Delta F(\cdot)$ is replaced by the expected stream of additional fixed cost, and $\Delta\pi(\cdot)$ by the expected stream of additional revenue. Theses streams incorporate both future actions of a player as well as complicated responses of competitors. While the intuition from a simple example carries over, one needs to apply a structural model to a concrete data set to verify if the comparative statics results facilitating the identification still hold. Moreover, it is likely that such computations will require strong parametric assumptions.

[13] In case of a richer model that contains additional stochastic unobservables it is possible to obtain a point identification. It would be the case if there is no θ that would satisfy all the inequalities with probability one. To obtain an estimate of θ one could minimize a penalty function of a form 3.5, which would produce a valid estimator of θ. Econometric details of such approach are contained in Bajari, Benkard, and Levin (2004).

An example of such analysis for the post 1996 merger wave in radio is contained in subsection 4.5.

The intuition behind the identification of the repositioning cost parameter is similar to cost parameters. The estimation procedure obtains bounds on repositioning cost by inducing more and less than optimal repositioning, and comparing it with equilibrium outcomes observed in the data. Similar identification argument can be made for an estimator of repositioning cost obtained by Sweeting (2011)[14] and the details are not discussed in this paper.

4 Application

In this section, I describe how to use the above framework to estimate merger synergies from ownership consolidation in the U.S. radio industry. In the next subsection I give a brief review of the industry. The second subsection presents the tailored version of the estimation algorithm. The last subsection presents and discusses the results.

4.1 Industry and data description

Radio is an important medium in the U.S., reaching about 94% of Americans aged 12 and older. Moreover, the average consumer listens to about 20 hours of radio per week and between 6am and 6pm more people use radio than TV or print media.[15] There are about 13,000 commercial radio stations that broadcast in about 350 strictly defined local markets nationwide. I use an Arbitron definition of local markets that is widely adopted as an industry standard.

Before 1996, this industry had ownership limitations both nationally and locally, preventing

[14] It is important to note that Sweeting (2011) uses a Maximum Likelihood Estimator with a nested Parametric Policy Iteration (PPI). Using such estimator in this paper could improve efficiency, however it is computationally infeasible. The main reason is that combined repositioning and merger action space is large, so best responses would be too costly to compute. Moreover, in case of my model, PPI step requires many more evaluations of static payoffs than value function simulation (payoffs are used in a Bellman equation step as well as to construct a problem specific projection basis). Therefore, I would have to solve a computationally intense static pricing game too many times. One solution would be to use an approximation of a static game instead of a full Nash equilibrium. However, in practice it is hard to come up with a tractable reduced form approximation that would provide good predictions of an impact of mergers on market power in a differentiated product industry.

[15] Source: A.Richter (2006)

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 triggered an unprecedented merger and product repositioning wave that completely reshaped the industry. Figure 1 contains the average percentage of stations that switched owners and that switched 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 in 2006.

The impact of this consolidation on consumer surplus has been studied before using a static demand and supply approach. For example, Jeziorski (2011), finds that consolidation of ownership in this industry was harmful to advertisers, causing \$223 million loss in advertiser surplus, but beneficial to listeners, raising the listener welfare by 0.2%.

In order to analyze the supply side effects of this consolidation, I compiled a dataset[16] on stations in the 88 markets. The data contains ownership for each station o_j and station format f_j. I use the estimates of station quality ξ_j that are obtained using the procedure described in the on-line appendix.

For the purpose of this paper, a merger activity observed in a BIA dataset is interpreted in a particular way. Mergers involving many markets are treated 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 other company (for example in case a full merger violates the ownership cap). Divestitures of stations from the portfolio that were forced by the regulator are treated as regular sales. When estimating merger policy function I do not include entry of new owners. It is equivalent to assuming that, that players' beliefs when merging are consistent with no entry of new owners[17]. Entry of

[16]Data is constructed using the software provided by *BIA Financial Network Inc.* and Media Market Guides by SQAD

[17]Under the assumption that there are no cross-market cost synergies, entry of new owners through acquisition of an existing owner, is equivalent relabeling the owner. Since I concentrate on symmetric equilibria such relabeling does not affect my results.

radio stations is handled through repositioning from an inactive state.

Because many stations were purchased in cash, I also observe an acquisition price for about about 40% of deals. Part of the remaining 60% of the deals either spanned across multiple stations and the individual prices were not specified, or a payment was not made in cash, but in other instruments (station swaps, other equity or debt transfers). The dataset contains some information about these transactions; nevertheless it is very hard to extract from it exact station prices. As a result I use only a subset of acquisition deals to estimate a pricing equation. However, during the estimation I argue that the selection on unobservables is not an issue using Heckman 2-step procedure.

4.1.1 Evidence of cost synergies

Preliminary evidence of cost synergies can be obtained by looking at the acquisition prices and comparing them the stream of profits of individual stations. First, I compute a very rough estimate, market by market, of expected average stream of station revenues. To do that, I take an average (1996-2006) yearly station revenue in a given market and multiply it by 0.95. Next, I obtain an average industry profit margin over years 1995-2003 from FCC Radio Report 2003. The report gives two different measures: median net profit margin equal to 20%, and median EBIT profit margin equal to 12%. I multiply the computed revenue stream by both margins to obtain an estimate of an average discounted profits of stations in each market. I divide that profit by an average acquisition price in each market and take an average of that ratio across markets.

I find that, using a net profit margin measure, the acquisition prices are on average 43% higher than expected profits. The number grows to 130% if I use an EBIT profit margin. These numbers are hard to rationalize by market power alone since during the consolidation period the standard deviation of radio industry net profit margins amounted to merely 7%. One of the possible explanations could be the anticipated cost synergies. This simple calculation does not take into account many issues like potential selection problems (mergers occur between companies with higher profit margin), or the fact that the stations' revenues are not constant over time. However, it does provide a motivation for further and more detailed investigation.

Cost synergies from consolidation has been acknowledged by 1998 Occupational Outlook Handbook released by U.S. Department of Labor. For example according to that publication: "Employment of announcers is expected to decline slightly through 2008 due to the lack of growth of new radio and television stations... Changes in station ownership, format, and ratings frequently cause periods of unemployment for many announcers. Increasing consolidation of radio and television stations, new technology, and the growth of alternative media sources will contribute to the expected decline in employment of announcers. Consolidation in broadcasting may lead to increased use of syndicated programming and programs originating outside a station's viewing or listening area.". Similarly 2004 edition of Occupational Outlook Handbook mentions common ownership within local markets as a major factor reducing employment: "Employment in broadcasting is expected to increase 11 percent over the 2004-14 period, more slowly than the 14 percent projected for all industries combined.... Consolidation of individual broadcast stations into large networks, especially in radio, has increased as the result of relaxed ownership regulations. This trend will continue to limit employment growth as networks use workers more efficiently. For example, 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 advertising sales workers are pooled to work for several stations simultaneously. ". Motivated by these studies I proceed to quantify cost synergies within local markets using a structural model.

4.2 Static profits

The static profit function is described by a simplified version of the model used by Jeziorski (2011). Radio stations provide free programming for listeners and get revenue from selling advertising. However, there is a direct relationship between listenership market share of the station and its revenue, because the advertising slots are priced on a per-listener basis. The pricing is done according to market shares of each station reported by a rating company Arbitron.

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 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_{1i}^L is a set of format fixed effects, θ_{2i} is a disutility of advertising, and θ_3^L is an AM/FM fixed effect. I assume that the random coefficients can be decomposed as

$$\theta_{1i}^L = \theta_1^L + \Pi D_i + \nu_{1i}, \quad D_i \sim F_m(D_i|d), \quad \nu_{1i} \sim N(0, \Sigma_1)$$

and

$$\theta_{2i}^L = \theta_2^L + \nu_{2i}, \quad \nu_{2i} \sim N(0, \Sigma_2)$$

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

$$r_j(q|s,d) = \int\int \frac{\exp(u_{ij})}{\sum_{j' \in s} \exp(u_{ij'})} dF(\nu_i) dF_m(D_i|d) \tag{4.1}$$

On the advertiser side, the pricing is done on a per-listener basis. The price for a 60sec slot of advertising is the product of cost-per-point (CPP) and station rating (market share in percentage). To capture potential market power of radio stations over advertisers CPP is allowed to be a decreasing function of the ad quantities offered by a station and its competitors. The slope of this function is to be estimated. The simplest model that captures these features and is a good approximation of the industry is a linear inverse demand for advertising, such as

$$p_j = \theta_1^A r_j \left(1 - \theta_2^A \sum_{f' \in \mathbb{F}} \omega_{ff'}^m q_{f'}\right) \tag{4.2}$$

where f is the format of station j, θ_1^A is a scaling factor for the value of advertising, θ_2^A is a market power indicator and $\omega_{ff'} \in \Omega$ are weights indicating competition closeness between formats f and f'. Such formulation, due to data limitations, abstracts from the strategic choice of advertising time. For a detailed study of this topic the reader is referred to Sweeting (2010).

Given the advertising quantity choices of competing owners q_{-k}, each radio station owner k chooses q_k to maximize total profit before subtracting the fixed cost

$$\bar{\pi}_k(q_k|q_{-k}, s, d) = \max_{\{q_j; j \in s_k\}} \sum_{j \in s_k} r_j(q|\xi, \theta^L) p_j q_j - \mathrm{MC}_j q_j \tag{4.3}$$

I assume a constant marginal cost and allow for a firm level of unobserved additive marginal cost heterogeneity. The market is assumed to be in a Nash Equilibrium.[18] The augmented profit function is given by $\bar{\pi}_k(s,d) = \bar{\pi}_k(q_k^*|q_{-k}^*, s, d)$

4.3 Estimation details

The first piece of the model that needs to be specified is the function $\sigma(s^t, d^t)$, that prescribes the sequence of moves in the merger and repositioning process. I assume that the biggest firms move first and I use market share in the previous period as a measure of size. This is motivated by the fact that the bigger players in the market might a have first-mover advantage over smaller players.

To estimate the acquisition probabilities I use the method outlined in Example 3.3. Each owner considers stations to acquire, one at a time, starting from the one with the highest quality measure ξ_j, and moving down according to ξ_j. A detailed description of the merger process is presented in Appendix B. Following Example 3.3, assume that shocks to payoffs of acquiring station j have a form $\phi_{jk}^t = \epsilon_{jk1}^t - \epsilon_{jk2}^t$, where epsilons are IID Type 1 extreme value random variables. Because the shocks are independent between stations, the conditional acquisition decisions will be correlated only with past acquisition decisions. Therefore, I can write the probability of acquiring station j as

$$\widehat{\text{Prob}}_j^A(\cdot|\tilde{a}_{-j}^t, s^t, A_{-k}^t, d^t|\theta^{AQ})$$

where $\tilde{a}_{-j}^t$ are acquisition decisions already made in this period by firm k. CCPs can now be estimated using a logit model with a flexible index function. Because of the large size of the state space I use a linear link function of several statistics about the state space computed from the data (a similar approach can be found in Ryan (2005), Ellickson and Beresteanu (2005), and Ryan and Tucker (2006)).

A similar strategy can be employed to estimate conditional format switching probabilities. The flow chart describing a format switching process is contained in Appendix B. Under the sequentiality assumption CCPs can be expressed as a one-dimensional measure

$$\widehat{\text{Prob}}_j^B(\cdot|\tilde{b}_{-j}^t, s^t, B_{-k}^t, A^t, d^t|\theta^{REP})$$

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

where $\tilde{b}_{-j}^t$ are past repositioning actions of owner k. I assume that format maintenance cost shocks ϕ_{jk}^t are IID Type 1 extreme value random variables. In this case, $\hat{\theta}^{REP}$ can be consistently estimated as a multinomial logit.

In the second stage of the estimation, I parametrize the fixed cost function

$$F_m(n|\theta^{FIX}, \theta^{SYN}) = \theta_m^F \sum_{q=1}^{n} MFC(q) \tag{4.4}$$

where n is the number of stations owned and $MFC(q)$ is the marginal fixed cost of operation of q-th station. I allow for four different values of cost level θ^{FIX} depending on a market population size, i.e. I allow for 4 different values of θ^{FIX}: for markets with greater than 2.5M, between 1M and 2.5M, between 0.5M and 1M and less than 0.5M population. It is motivated by the fact that Occupational Outlook Handbook quotes much larger salaries in broadcasting industry in larger markets. $MFC(1)$ is normalized to 1 so that θ^{FIX} is the cost of operating one station. I set $MFC(2) = \theta_1^{SYN}$, $MFC(8) = \theta_2^{SYN}$ and compute MFCs for intermediate values by linear interpolation. This formulation allows for increasing ($\theta_1^{SYN} > \theta_2^{SYN}$) and decreasing ($\theta_1^{SYN} < \theta_2^{SYN}$) amount of cost synergies, as the company owns more stations.

Current formulation of cost synergies does not account for the fact that there might be additional synergies between stations of the same format. If it is the case, it might affect my estimates of the switching cost. It can also lead me to overestimating cost synergies between stations of different formats. Acknowledging this fact, I decided to compromise on identifying cost synergies within format to get reasonable confidence bound on total cost synergies. Such trade-offs are forced by the fact that the accuracy of a BBL estimator is largely affected by the number of parameters in the second stage. It is caused by the fact that the estimation errors from the first stage add up with the errors in the second stage. Moreover, identification of within format cost synergies separately from overall cost synergies and format switching cost is likely to heavily rely on functional form assumptions.

The paper abstracts from potential cross-market cost synergies. It can be an issue for large owners like Clear Channel, however introducing such synergies would greatly increase the number of parameters to be estimated. Entry by acquisition would not be equivalent to relabeling the existing owner, since a new owner might own an extensive network of stations outside of the market. To account for that one would need to endogenize entry decisions of large owners. While, it is theoretically possible, in practice it would be infeasible (given a current dataset) to additionally estimate a flexible entry strategy in the first stage, while keeping the precision of cost estimates at the reasonable level.

To allow for heterogeneity of repositioning cost across markets and to keep the number of estimated parameters small I set $\theta_m^{REPCOST} = \theta^S \theta_m^{FIX}$. This assumption means that all heterogeneity across markets is captured in the heterogeneity in operations costs. Richer heterogeneity is not identified given the available data and produces large confidence bounds.

The standard deviation of unobserved profit from mergers θ_m^{ϕ} and switching θ_m^{ψ} is assumed to be proportional to the average observable per-period market revenue of the owner, i.e $\theta_m^{\phi} = \theta^{\phi}(1-\beta)A_m$. 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.

In the second stage, I simulate the value function only for the owner with the biggest market share at each data point (s^{tm}, d^{tm}). These simulations are done according to Algorithms 2 and 3 from the Appendix B. The suboptimal value function $\tilde{V}_k$ is computed under 4 suboptimal strategies:

1. "More mergers": Increase the probability of merger by 50% (not by 50 percentage points), until first suboptimal merger happens
2. "Less mergers": Prevent a first merger
3. "More format switching": Switch the first station into the random format
4. "Less format switching": Prevent any switches in the first year

These four strategies produce 4 counterfactual value functions $\tilde{V}_k$. To increase efficiency the parameters are estimated jointly using equation (3.5). The condition that value function V cannot be negative is included and a fifth set of restrictions. Because the nominal deviations for larger markets are higher I weight each deviation from equilibrium by a level of counterfactual revenues $(\tilde{A}_k^{tm})$. These weights do not depend on parameters and in practice prevent the outliers from dominating the results. Moreover, to make sure that neither of the restrictions is dominating others in an estimation process I normalize the penalty from each restriction at a starting point (all parameters are zero) to 1.

4.4 Results

This subsection describes the results of the estimation. The exposition is divided into two parts. Since the main goal of the application is to estimate the extent of fixed cost synergies in the radio market, I start with presenting results of a second stage: fix cost and repositioning cost parameters followed by counterfactuals. I perform a full correction of standard errors using a parametric bootstrap from a joint asymptotic distribution of profit function, station quality ξ, quality auto-correlation parameter ρ, and first-stage estimates. Standard errors were also corrected for a second stage simulations by using independent draws for each bootstrap iteration[19].

Second, I present the first stage estimates: acquisition pricing, acquisition strategy and format switching strategy. Even thought the exact values of these estimates are not of a main concern, it is important that they are economically plausible, since they determine the second stage results. The standard errors were corrected for the fact that unobserved station quality (which enters as a covariate) is estimated beforehand, jointly with one-shot profit function parameters.[20] This was done by expressing a profit function estimation and a first-stage estimation as a joint GMM objective function. Even though the first-stage estimation is semi-parametric this correction is valid by the results of Ai and Chen (2007) and Ackerberg, Chen, and Hahn (2009).

The transition of ξ prescribed by the equation (2.1), as well as a distribution of ζ (non-parametric) is estimated jointly with a profit function. I find that $\hat{\rho} = 0.56$ with 0.09 standard deviation. During the simulation I draw from an empirical distribution of ζ controlling for different variance in each market.

4.4.1 Second stage: Fixed and switching cost

The estimates of fixed cost parameters can be found in Table 2. In the estimation algorithm θ_2^{SYN} was restricted to be less than or equal to 1 [21]. The cost of operating one station θ^{FIX} is

[19]Standard errors are based on 70 parametric boostrap draws. Reestimation of the second stage takes about 8 hours using a 48 CPU cluster running a highly optimized and parallelized C code.

[20]Details of one-shot profit and unobserved station quality estimation can be found in an on-line appendix.

[21]This restriction means that I prohibit dis-economies of scale when owning large number of stations. The available data variation does not allow to test against dis-economies of scale on the margin when owning large stations. The procedure with unrestricted parameter value produces a large value for a θ_2^{SYN} with a large standard error. In effect, I can only test for economies of scale against constant returns to scale on margin when owning

decreasing with the size of the market. Its level is slightly bigger than the revenue of medium and big stations in those markets, which suggests that owners need to achieve an efficient scale to earn profits. Interpretation of parameters θ^{SYN} is presented in Table 3. I find that there are extensive synergies of operating multiple stations early on; however, as the portfolio grows savings are quickly vanishing. This is consistent with the results of O'Gorman and Smith (2008) which find a similar relationship for the radio industry using a static model. I find that average cost savings (across all owners) between 1996 and 2006 amount to \$0.51m (with a \$0.195m standard error) or 22% (with a standard error of 6 percentage points) per owned station. The economic significance of these parameters is presented in Table 5. I compare average, post 1996 Telecom Act, per-year cost savings for the whole country with changes in consumer and producer surplus computed by Jeziorski (2011). Since 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. Mergers that occurred after 1996 provided an additional \$988m (with a standard error of \$382m) of fix cost savings (about 5% of total industry revenue). This outweighs the changes in producer surplus by about \$750m per year (the difference is significant with a 5% test size). Therefore, I conclude that despite the drop in advertiser surplus caused by smaller competition, the post 1996 merger wave raised total surplus.

The last column of Table 2 is an estimate of the standard deviation of an unobserved one-time merger revenue/cost distribution. It is estimated to be about 130% of an average per-period revenue of a radio owner. It has to be noted that those revenues/costs represent an aggregate value on 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%. This measures the extent of selection on unobservables during the merger process.

Table 4 presents format switching costs. These numbers are fairly large, which is consistent 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, the optimal thing to do would be to purchase stations close to your portfolio to get rid of competition and

a large number of stations. Note, because I do not need restrictions θ_1^{SYN}, I can still test for economies of scale against dis-economies of scale (including constant returns) on the margin for small amount of stations owned.

reposition them to avoid cannibalization. However, if the switching costs are high, it might be optimal to avoid paying them and purchase a station further away. The previous subsection and Sweeting (2009) present the evidence of the latter type of behavior, reinforcing the finding of high switching cost estimates.

The second row of Tables 2 and 4 present the results of a robustness check to changing the station acquisition sequence for each owner (ordering by ξ in Example 3.3). Since I do not observe that sequence in the data, I cannot directly justify ordering by quality to be the correct sequence to use. Instead I re-estimate the second stage of the model under the alternative fully random sequence (each period a new random sequence is drawn). I find that the results are largely robust to these perturbations. This can be explained by the fact that even though the perturbed model predicts acquisition of different stations, the prices paid are also adjusted and account for the difference in net acquisition profit.

4.4.2 First stage: Demographic dynamics

Predicting the demographic trends is a very complicated and challenging task. For the purposes of this paper, I am interested in capturing only the first order mid and long run trends that might effect format switching. When simulating the value function, each period I record the share of different demographic groups in all the markets (groups can be found in Table 12). For periods before 2009 I compute these shares using CPS. For periods 2009 and after I use national census projections of growth rates to compute new shares (for education and income groups I compute the mean 1996-2006 shares). These shares are used when computing the integral (4.1) and enter as a series of independent binomial random variables.

4.4.3 First stage: Acquisition pricing

The results of OLS regression of acquisition prices on a chosen statistic from the information set can be found in Table 7. The top part of the table contains market level covariates. Listeners population is a big driver of an acquisition price, since per-listener ad prices are largely dependent on the size of the market. Dummies as well as the coefficient on the population size are positive and highly significant. Percentage of stations in the format of 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 suggests high switching cost. I find limited evidence that demographics affects acquisition price. For example, interactions between percentage of Hispanic population of price of Hispanic stations is positive and significant.

The second part of the table consists of station level covariates. I find that station quality positively affects price; however, the effect is diminishing. It is consistent with the fact that in the assumed profit function, station quality has a decreasing effect on revenues. I find that Dark (in a inactive state $\bar{f}$) stations are on average cheaper than their active counterparts and FM stations are on average 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 the buyer and seller characteristics. I find that the price is positively affected by number of stations already owned. This is preliminary evidence for increasing market power or cost synergies. The coefficient on the dummy controlling for the bargaining power of the seller (proxied by the fact that it is in top3 in a move sequence) is positive. This suggests that, controlling for station covariates, higher ranked sellers obtain higher prices. This might be explained by a greater amount of business stealing if buying from a bigger competitor or better negotiation power of bigger sellers. At the same time I find no effect of the ranking of the buyer on price (coefficient not included in the final regression).

4.4.4 First stage: Acquisition strategy

The results on determinants of acquisition strategy are grouped in Tables 8-12. Table 8 contains values of intercepts on acquisition strategy for different seller and buyer ranking. I find that higher ranked buyers are more likely to acquire new stations. This can be caused by either an increasing amount of market power or cost synergies. These two stories can be disentangled by the structural estimation in the second step. Additionally, I find that companies are less likely to purchase stations from higher ranked sellers. This can be connected to the fact that higher ranked sellers quote on average higher prices for similar stations.

The impact of market structure on acquisition decision can be investigated using Table 10.

It presents coefficients on the percentage of the total number of stations that are owned by the acquiree and top competitors. The percentage numbers were disaggregated by format. I find that own effects are much stronger than competitive effects. Most of the coefficients in the higher part of the table are highly significant. The diagonals of both parts of the table indicate impact of own format on purchase decision. Coefficients for own effect are negative and highly significant, which suggests cannibalization and high switching cost. The sum of all own portfolio coefficients column by column is the effect of a total percentage of owned stations on acquisition decision. These numbers are negative and highly significant, that suggests that the closer the owner is to the ownership cap (roughly maximum percentage of stations that it can own) the lower incentives it has to acquire an extra station. It is consistent with an intuition that owners do not want the caps to be binding in order to have an option value of acquiring stations in response to changes in demographics or quality. The effects of the portfolio owned by competitors are presented in the lower part of the table. Half of the coefficients are positive and three of them are significant, which suggests that these effects are weaker but still present.

Table 9 presents the impact of chosen station characteristics and past actions on the propensity to acquire. I find that 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 very limited and the price of such small stations is much lower, this can be seen as a way to enter or introduce new stations.

In the second part of the Table 9 I present the coefficient on dummies indicating the number of past acquisitions. As explained in Section 3, this formulation allows controlling for correlation between acquisition decisions of multiple stations while maintaining computational feasibility of the estimator. The correlation structure is summarized by allowing conditional probabilities of acquiring additional station to depend on number of previous acquisitions in the same period. In case of acquiring multiple stations in the same local market in one period, the numbers in Table 10 should be interpreted in conjunction with those with Table 9. The logit index for probability of acquiring a first station is affected by a relevant number in Table 9 that is decreased by 5.32 ("None" column in Table 10). Logit index probability of acquiring second station conditional on one acquisition is decreased by only 2.05, which picks up the fact that acquisitions are highly correlated in the data. For third station the index is decreased by 1.02, for fourth station by 0.67

and is not decreased in case of fifth or more stations.

Table 11 presents the impact of station quality on acquisition decision. Controlling for the position of the station in queue (Table 9), the station quality has no impact on acquisition decision. This can be explained by the fact, described in the previous section, that higher quality stations cost more to acquire. On the other hand, I find that the average quality of the owned stations in the format of a potential acquiree increase the propensity to merge. This can be evidence of quality spillovers; however, those are beyond the scope of this paper.

Table 12 constrains interactions between a format of a potential acquiree and the percentage number of different demographic groups. I find that demographics is not a big driver of acquisitions. Most of the numbers are small and not significant. A notable exception is the percentage of low income people which is correlated with acquisition in almost any format.

To check for goodness of fit I compute the average probability of making merger/no merger decision for each iteration of the merger algorithm. Results are reported in Table 6. MLE that is used to estimate the model maximizes the total average probability. It is reported in 'All' column and amounts to 0.98. Subsequently, I check the robustness of this probability to slicing the data into subsamples. I compute the average probability of a correct decision only for stations of a particular format. The numbers are stable and consistently high. Even though it is a within sample robustness analysis, it suggests that there is a good fit across many heterogeneous subsamples.

4.4.5 First stage: Format switching

Table 13 presents a format switching matrix. Each column of the table represents the target format (relative to DARK). The first row of the table contains an AM dummy. The results suggest that AM stations are most likely to switch to News/Talk format and least likely to switch to Rock and CHR format (AM stations would more likely switch to DARK than these formats). This is connected with the fact that AM stations tend be less commercial and are used to serve niche markets. The diagonal of the table is positive and highly significant, capturing the fact that stations are more likely to stay in their current format than switch. Many off-diagonal coefficients are also positive and significant, which suggests certain general trends in format switching. For example, Rock stations are likely switch to CHR or Urban and Country stations are likely switch to AC. Coefficients in the DARK row are negative and significant which means that DARK stations are more likely to turn-on than to stay DARK.

Table 14 presents the impact of station characteristics on the probability to stay in the current format. Each row represents a current format and each column represents a station characteristic. I find that AM stations are more likely to stay in their formats, with an exception of News/Talk and Other. However, because these numbers have to be viewed in conjunction with the previous table, they do not mean that an AM News/Talk station is more likely to be of another format than News/Talk. The second row of the table presents the impact of the acquisition on staying in the current format. The highly significant negative numbers capture the fact that in the data the conditional probability of format switching on acquisition is much higher than the unconditional one. 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, while the fact that competitors own them increases the probability of switching out.

Table 15 presents portfolio effects in format switching. The diagonal of the upper part of the table contains own format effects. For CHR format I find a strong negative effect which can be an evidence of cannibalization of stations of this format. On the other hand Spanish format has a strong positive own effect that is capturing specialization of some owners in this type of format that is present in the data. Many off-diagonal elements are also significant. For example, the News/Talk row contains mostly positive and significant coefficients. This suggests that each owner tries to have a News/Talk station and it does not depend on a portfolio. The lower part of the table presents competitive effects. Since most of the coefficients are insignificant, I conclude that these effects are not a significant driver of format switching (as opposed to mergers).

Table 16 presents the relationship between the current demographic composition of the market format switching decisions. One can observe many patterns that suggest firms respond to the current state of population demographics according to demographic tastes for formats. For example, a larger current population of Hispanics 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 patters largely reflect correlations between tastes for formats and demographics described in Jeziorski (2011).

4.5 Identification

This section contains an analysis of the ideas presented in section 3.4. Table 17 presents the objective function values disaggregated by the impact of each counterfactual policy. The implemented counterfactual policies are equivalents of simple policies discussed in section 3.4. The policy of "more mergers" is a version of "always merge", and the policy of "less mergers" is a version of "never merge". I find that all parameters are identified pointwise, because of large variation in revenue shifters and functional form assumptions. The value of parameters is chosen to minimize the penalty function and reach a right balance between violating different sub-optimality conditions (equivalents of G functions from section 3.4). The remainder of this section discusses the comparative statics induced by selection, fixed cost and switching cost that enable identification.

The first row in Table 17 contains values of penalty functions accounting only for fixed cost. Turning off selection and synergies results in increasing the penalty for "more mergers" and decreasing the penalty for "less mergers". In the former policy the company owns more stations on average than observed in the data and in the latter on average less. This is equivalent to the impact of the fixed cost level θ^{FIX} on equations (3.6) and (3.7). In effect the "less merger" policy has too big of a penalty. Additionally the last term penalizing a negative value function is binding.

To counterbalance that effect, I introduce selection and fixed cost synergies. Selection increases the profits of "more mergers" and decreases profit of "less mergers" because it takes into account unobserved profitability of mergers. Cost synergies have the same effect and equivalent to the impact of θ^{SYN} on equations (3.6) and (3.7). The parameters are obtained by jointly minimizing the countering effects of cost level and synergies/selection on total penalty function. The only caveat left is to distinguish between selection and cost synergies. This distinction is made by the structure of the model and the fact that unobserved profit is assumed to be IID and uncorrelated with the number of stations already owned, while synergies are persistent and are a function of an owned portfolio. Therefore in practice changing those parameters will have a different impact on the penalty function for different starting portfolios.

The last two rows present identification of format switching cost. This is very similar reasoning involving "less switching" and "more switching" policies. Because all parameters are estimated jointly, format switching cost affects penalty of suboptimal merger policies, since more mergers induce more format switching. This gives additional estimation power that increases the efficiency of all estimates.

5 Conclusions

This paper proposed a new estimator of a production cost curve that enables the identification of cost synergies 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. Since reliable cost data is very hard to obtain, the cost side analysis of mergers was very hard to perform. This method is able to solve this problem, and it provides a powerful tool for policy makers to improve their merger assessments. The estimates can be used for retrospective merger analysis, as well as to compute cost savings from future mergers.

Since the proposed method is based on a fully dynamic framework, it additionally solves many of the problems of static merger analysis. First of all, endogenizing the merger decision allows for sample selection on unobservables in the estimation and correcting for the fact that only the most profitable mergers are carried out. Moreover, I allow for follow-up mergers and merger waves. Additionally, endogenizing product characteristics enables correction for post-merger product repositioning.

The estimator belongs to a class of indirect estimators proposed by Hotz, Miller, Sanders, and Smith (1994) and Bajari, Benkard, and Levin (2004). 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.

The estimator was applied to analyze the cost side benefits of a de-regulation of the U.S. radio industry. It turns out that the consolidation wave in that industry between 1996 and 2006 provided substantial cost synergies. I find that average cost savings (across all owners) between 1996-2006 amount to \$0.5m (22%) per owned station. Total cost savings only from mergers after 1996 amounted to about \$1b of cost savings, which outweighs the loss of advertiser surplus caused by increased market power (this loss was estimated to be \$300m). This provides a significant argument for the supporters of a de-regulation 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 (2009): "A Practical Asymptotic Variance Estimator for Two-Step Semiparametric Estimators," Discussion paper.

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.

Arcidiacono, P., and R. Miller (2010): "CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," Discussion paper, Duke University.

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 (2004): "Estimating Dynamic Models of Imperfect Competition," NBER Working Papers 10450, National Bureau of Economic Research, Inc.

Bajari, P., V. Chernozhukov, H. Hong, and D. Nekipelov (2009): "Semiparametric Estimation of Dynamic Games of Incomplete Information," Discussion paper, Working Paper.

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 (2006): "A Theory of Strategic Mergers," Discussion paper, EFA Zurich Meetings Paper.

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): "Mergers, Station Entry, and Programming Variety in Radio Broadcasting," Working Paper 7080, National Bureau of Economic Research.

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.

Dube, J.-P. H., J. T. Fox, and C.-L. Su (2009): "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation," NBER Working Papers 14991, National Bureau of Economic Research, Inc.

Ellickson, P., and A. Beresteanu (2005): "The Dynamics of Retail Oligopolies," Discussion paper.

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. K., L. M. Froeb, S. T. Tschantz, and G. J. Werden (2005): "Post-Merger Product Repositioning," *SSRN eLibrary*.

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. (2011): "Merger enforcement in two-sided markets," Discussion paper, Johns Hopkins University.

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

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

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.

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.

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.

Ryan, S. (2005): "The Costs of Environmental Regulation in a Concentrated Industry," Working Papers 0510, Massachusetts Institute of Technology, Center for Energy and Environmental Policy Research.

Ryan, S., and C. Tucker (2006): "Heterogeneity and the Dynamics of Technology Adoption," Working Papers 06-26, NET Institute.

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

——— (2010): "The strategic timing of radio commercials: An empirical analysis using multiple equilibria," *RAND Journal of Economics*, 42(2).

——— (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.

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

Appendices

A Estimation without acquisition prices

In case the pricing function $\hat{P}^r_{jk}$ cannot be estimated in the first state because of data constraint, one could employ a bargaining model for infer it. Suppose one employs a parametrization $\hat{P}(\omega|\theta_P)$. For an initial value of parameters θ^0_P one could compute a surplus from acquisition of the product j by an owner k using simulated $\hat{V}^t_k$ and $\hat{V}^t_{k'}$ where k' is the current owner of product j. Then using a bargaining model one could infer prices and fit a new parametrization θ^1_P. If repeating this procedure leads to convergence, then obtain a parametrization $\hat{\theta}_P$ and value functions $\hat{V}^t_k$ that are consistent with each other. The detailed description of this procedure is given in Algorithm 1. The big dowside of this approch is that one needs resolve this procedure for any set of cost parameters and cannot take advantage of the linear nature of the value function. It makes the procedure infeasible to use for large datasets because of the computational burden. However, given the rapid development of hardware it is reasonable to think it it would be feasible in the near future.

Algorithm 1: Estimator without price data

Take any θ^0_P;

Let $r = 0$;

repeat

- Simulate the value functions $\hat{V}^r$ using pricing process $\hat{P}(\omega|\theta^r_P)$;
- Compute surplus from any acquisition using the simulated value functions;
- Compute acquisition prices $\hat{P}_{jm}$ by applying any bargaining game;
- Fit new parameters θ^{r+1}_P using $\hat{P}_{jm}$;

until *convergence of* θ^r_P;

B Radio acquisition and format switching algorithms

This section of the appendix contains a detailed flows of the algorithms used to simulate the value function from section 4.

Algorithm 2: Merger algorithm

```
Let ω_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 = P̄^m;
        Compute acquisition probability Prob^M(ω_k^r, d^t);
        Draw a random number u from U[0,1];
        if u ≤ Prob^M then
            Increase A_{old owner}^r by β^{r-t} P_{jk}^r;
            Decrease A_k^r by β^{r-t} P_{jk}^r;
            Update ω_k^r for acqusition;
            Increase B_k^r by β^{r-t} E[φ|acquisition];
        end
    end
    Let ω_{k+1}^r = ω_k^r;
end
```

Algorithm 3: Format switching algorithm

```
Let ω̃_1^r = ω_{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 Prob_k^R(ω̃_k^r, d^r);
        Simulate the future characteristic f_j^{r+1};
        Increase C_k^r by β^{r-t} E[ψ|f_j^r];
        if the f_j changed then
            Update ω̃_k^r;
            Remember the repositioning for a computation of D_k^r ;
        end
    end
    Let ω̃_{k+1}^{tm} = ω̃_k^{tm};
end
```

C Tables and graphs

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

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

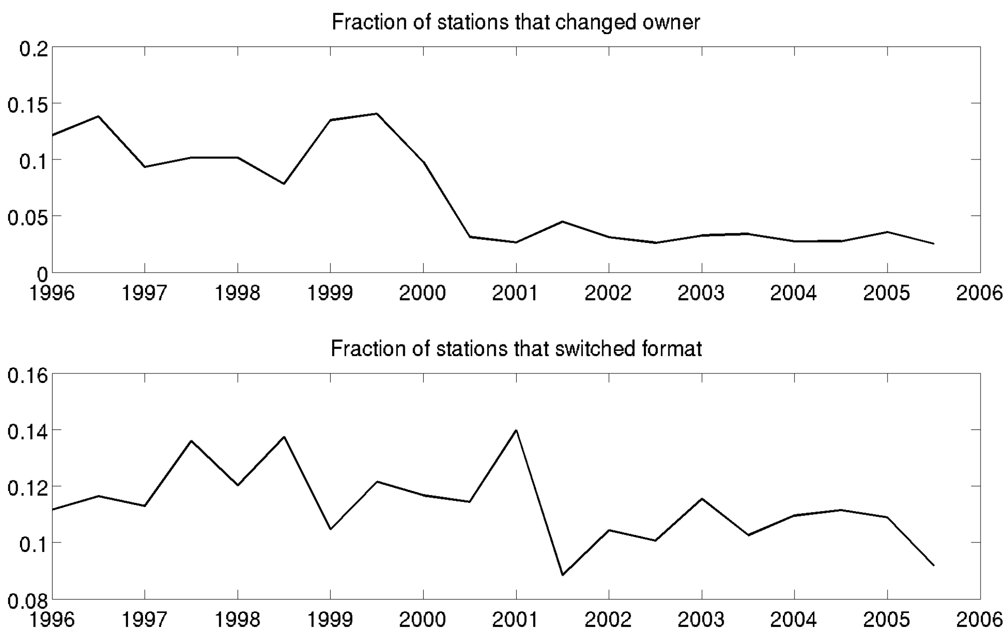

Figure 1: Dynamics of station acquisition and format switching

	Fixed cost θ_m^{FIX}				Synergies θ_1^{SYN}	Synergies θ_2^{SYN}	θ^{ϕ}
	>2.5M	1M-2.5M	0.5M-1M	<0.5M			
Main estimates	12.64*** (1.91)	2.40** (0.98)	1.47*** (0.34)	0.00 (0.12)	0.04 (0.28)	1.00*** (0.00)	1.37*** (0.45)
Random sequence	10.67	1.92	1.37	0.00	0.00	1.00	1.38

Standard errors (full parametric bootstrap) in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2: Second stage: Fixed cost parameters

	Number of stations owned							
	1	2	3	4	5	6	7	8
MFC	100.00%	4.00%	19.98%	35.96%	51.93%	67.91%	83.89%	99.87%
	(0.00)	(27.80)	(23.16)	(18.52)	(13.88)	(9.24)	(4.61)	(0.41)
Cumulative cost	100.00%	104.00%	123.98%	159.93%	211.87%	279.78%	363.66%	463.53%
	(0.00)	(27.80)	(50.96)	(69.48)	(83.36)	(92.60)	(97.21)	(97.17)

Standard errors (full parametric bootstrap) in parentheses

Table 3: 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

	Switching cost $\theta_m^{REPCOST}$				θ^{ψ}
	>2.5M	1M-2.5M	0.5M-1M	<0.5M	
Main estimates	55.63**	10.56	6.45	0.00	0.00
	(25.42)	(13.32)	(6.17)	(5.00)	(0.01)
Random sequence	56.56	10.20	7.26	0.00	0.00

Standard errors (full parametric bootstrap) in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Second stage: Switching cost parameters

	Consumer Surplus	Advertiser Surplus	Fixed Cost
Impact of Telecom Act	+0.2%	-$223m	-$987.66 (382.18)

Table 5: Total cost savings created (bootstrap standard errors in parenthesis) by mergers after 1996, compared to demand effects from Jeziorski (2011)

Station format	All	AC	Rock	CHR	Urban Alt.	News Talk	Country	Spanish
Average probability of a correct merger decision	0.978	0.976	0.977	0.979	0.980	0.980	0.982	0.979

Table 6: Goodness of fit of the merger strategy measured by average probability of a correct merger decision. The data was sliced by format check for robustness.

	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)
	Popilation 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 7: Determinants of acquisition price conditional on a merger

Buyer ranking			Seller ranking		
1st	2nd	3rd	1st	2nd	3rd
1.26^{***}	0.82^{***}	0.44^{***}	-0.34^{***}	-0.07	-0.08
(0.12)	(0.11)	(0.10)	(0.06)	(0.06)	(0.06)

Standard errors (corrected for sequential estimation) in parentheses

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

Table 8: Acquision: Seller and buyer rankings

AM	No report	Station ranking	Past acq.			
			None	One	Two	Three
-0.18^{***}	0.44^{***}	-0.04^{***}	-5.32^{***}	-2.05^{***}	-1.02^{***}	-0.67^{***}
(0.05)	(0.14)	(0.00)	(0.11)	(0.11)	(0.10)	(0.11)

Standard errors (corrected for sequential estimation) in parentheses

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

Table 9: Acquision: Station characteristics and past actions

	Portfolio	AC	Rock	CHR	Urban Alt.	News Talk	Country	Spanish	Other
Owner	AC	$\mathbf{-12.27}^{***}$	-5.02^{**}	-10.21^{**}	-7.35^{**}	-10.32^{***}	-8.25^{***}	-68.00^{***}	-22.65^{***}
		(2.56)	(2.42)	(4.05)	(3.77)	(2.31)	(2.38)	(11.81)	(2.50)
	Rock	-6.55^{***}	$\mathbf{-2.06}$	-14.09^{***}	-8.01^{**}	-6.22^{***}	-8.01^{***}	-44.14^{***}	-10.60^{***}
		(2.41)	(2.71)	(4.73)	(3.94)	(2.24)	(2.71)	(9.80)	(2.29)
	CHR	-6.99^{*}	-3.68	$\mathbf{-17.26}^{**}$	-5.87	-3.33	-10.99^{**}	-37.20^{***}	-19.95^{***}
		(3.69)	(4.13)	(7.69)	(6.37)	(3.50)	(4.44)	(10.73)	(3.92)
	Urban Alt.	-10.74^{***}	-10.66^{***}	-6.82	$\mathbf{-1.90}$	-18.33^{***}	-16.01^{***}	-76.93^{***}	-18.90^{***}
		(3.13)	(3.51)	(5.11)	(3.54)	(3.39)	(3.57)	(15.18)	(2.71)
	News Talk	-6.01^{***}	-6.23^{**}	-1.11	-18.35^{***}	$\mathbf{-9.08}^{***}$	-7.19^{***}	-61.68^{***}	-19.62^{***}
		(2.20)	(2.54)	(3.62)	(4.00)	(2.47)	(2.38)	(8.30)	(2.38)
	Country	-3.80^{*}	-5.82^{**}	-3.98	-19.13^{***}	-13.45^{***}	$\mathbf{-10.75}^{***}$	-48.98^{***}	-17.46^{***}
		(2.27)	(2.53)	(3.80)	(4.49)	(2.59)	(2.74)	(9.09)	(2.42)
	Spanish	-17.65^{***}	-24.90^{***}	-20.49^{**}	-29.21^{***}	-38.96^{***}	-33.17^{***}	$\mathbf{-13.93}^{***}$	-43.45^{***}
		(4.44)	(6.64)	(8.03)	(7.63)	(6.20)	(6.42)	(3.38)	(5.47)
	Other	-19.62^{***}	-16.19^{***}	-15.03^{***}	-27.03^{***}	-19.23^{***}	-18.30^{***}	-57.41^{***}	$\mathbf{-23.55}^{***}$
		(2.45)	(2.71)	(3.97)	(3.83)	(2.37)	(2.56)	(6.50)	(2.25)
Top2	AC	$\mathbf{3.37}^{***}$	-4.22^{**}	-2.35	-0.97	-0.12	2.40^{*}	7.03^{***}	0.04
		(1.26)	(1.69)	(2.40)	(2.22)	(1.38)	(1.44)	(2.39)	(1.18)
	Rock	0.19	$\mathbf{-1.25}$	-1.54	-1.19	-2.49	1.01	-1.91	1.20
		(1.54)	(1.72)	(2.59)	(2.56)	(1.62)	(1.65)	(2.59)	(1.38)
	CHR	0.62	3.15	$\mathbf{-1.01}$	2.33	-2.52	4.25^{**}	1.25	0.52
		(1.85)	(2.29)	(2.81)	(3.02)	(1.94)	(2.14)	(3.40)	(1.58)
	Urban Alt.	-1.17	-2.94	-8.21^{**}	$\mathbf{-3.94}^{*}$	-3.00	-1.60	1.87	-2.67^{*}
		(1.93)	(2.29)	(3.78)	(2.33)	(1.92)	(2.07)	(3.11)	(1.53)
	News Talk	-2.35^{*}	2.53^{*}	0.15	0.94	$\mathbf{3.05}^{**}$	0.17	-5.60^{**}	2.05^{*}
		(1.33)	(1.54)	(2.34)	(1.95)	(1.31)	(1.41)	(2.35)	(1.17)
	Country	2.41^{*}	-0.78	1.11	-0.69	4.89^{***}	$\mathbf{1.54}$	11.84^{***}	0.88
		(1.36)	(1.58)	(2.21)	(2.31)	(1.30)	(1.37)	(2.49)	(1.18)
	Spanish	3.09	2.07	-11.44^{*}	0.45	4.55^{**}	-4.25	$\mathbf{2.53}$	0.65
		(2.68)	(3.52)	(6.33)	(4.04)	(2.30)	(3.58)	(2.28)	(2.31)
	Other	-1.30	2.64^{*}	2.36	2.17	-1.35	-0.58	5.23^{**}	$\mathbf{5.15}^{***}$
		(1.34)	(1.53)	(2.24)	(1.97)	(1.35)	(1.45)	(2.47)	(1.03)

Standard errors (corrected for sequential estimation) in parentheses

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

Table 10: Acquisions: portfolio interactions

Station quality	Average quality in the format		
	Owner	Top2	Others
0.03	0.19***	−0.00	0.06***
(0.04)	(0.02)	(0.02)	(0.02)

Standard errors (corrected for sequential estimation) in parentheses

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

Table 11: Acquision: Quality

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

Standard errors (corrected for sequential estimation) in parentheses

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

Table 12: Acquisitions: demographics interactions

	AC	Rock	CHR	Urban Alt.	News Talk	Country	Spanish	Other
AM	-1.54^{***} (0.31)	-3.11^{***} (0.43)	-3.04^{***} (0.47)	-1.84^{***} (0.35)	1.60^{***} (0.31)	-1.05^{***} (0.34)	0.03 (0.32)	-0.07 (0.30)
AC	**6.71***** (0.51)	1.58^{***} (0.54)	1.78^{***} (0.55)	1.69^{***} (0.55)	0.13 (0.55)	1.19^{**} (0.55)	1.18^{**} (0.57)	2.45^{***} (0.51)
Rock	2.78^{***} (0.69)	**7.12***** (0.69)	0.96 (0.75)	2.81^{***} (0.71)	0.55 (0.72)	1.32^{*} (0.73)	1.30^{*} (0.77)	2.81^{***} (0.69)
CHR	2.15^{***} (0.66)	1.22^{*} (0.70)	**6.67***** (0.67)	2.36^{***} (0.68)	-0.93 (0.77)	0.07 (0.77)	1.42^{*} (0.73)	1.98^{***} (0.67)
Urban Alt.	1.59^{***} (0.57)	1.90^{***} (0.58)	0.86 (0.62)	**6.34***** (0.56)	-0.29 (0.60)	0.22 (0.64)	0.86 (0.64)	1.99^{***} (0.55)
News Talk	1.27^{**} (0.54)	0.59 (0.65)	0.41 (0.68)	0.35 (0.66)	**5.68***** (0.54)	0.81 (0.57)	1.00^{*} (0.57)	1.36^{***} (0.52)
Country	1.60^{***} (0.50)	1.09^{**} (0.52)	0.70 (0.55)	0.98^{*} (0.56)	0.11 (0.51)	**6.49***** (0.49)	1.12^{**} (0.55)	1.59^{***} (0.49)
Spanish	0.84 (0.60)	-1.56 (1.13)	0.31 (0.66)	0.61 (0.63)	-0.34 (0.53)	0.37 (0.63)	**5.10***** (0.53)	1.24^{**} (0.51)
Other	2.82^{***} (0.45)	1.76^{***} (0.48)	0.99^{*} (0.52)	2.07^{***} (0.48)	0.49 (0.47)	1.38^{***} (0.48)	1.37^{***} (0.50)	**6.69***** (0.44)
Dark	-1.81^{***} (0.51)	-2.38^{***} (0.59)	-3.55^{***} (0.61)	-3.34^{***} (0.55)	-3.49^{***} (0.57)	-1.99^{***} (0.59)	-2.74^{***} (0.47)	-1.62^{***} (0.46)

Standard errors (corrected for sequential estimation) in parentheses

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

Table 13: Format switching: switching patterns

	Stay in the current format				
	If AM	If acquired	Avg. quality in format		
			Owner	Top2	Others
AC	0.50^{**} (0.21)	-0.76^{***} (0.16)	0.05^{*} (0.03)	-0.01 (0.04)	-0.04 (0.03)
Rock	0.99^{**} (0.43)	-0.70^{***} (0.21)	0.16^{***} (0.05)	-0.12^{*} (0.07)	-0.13^{***} (0.05)
CHR	0.82 (0.55)	-0.85^{***} (0.26)	0.16^{***} (0.05)	-0.02 (0.10)	-0.06 (0.06)
Urban Alt.	1.17^{***} (0.30)	-0.63^{***} (0.23)	0.20^{***} (0.05)	-0.03 (0.08)	-0.01 (0.04)
News Talk	-1.53^{***} (0.25)	-1.22^{***} (0.18)	0.17^{***} (0.03)	-0.10^{**} (0.04)	-0.02 (0.03)
Country	0.07 (0.25)	-1.07^{***} (0.18)	0.07^{**} (0.03)	-0.09 (0.06)	-0.04 (0.03)
Spanish	-0.23 (0.29)	-1.74^{***} (0.23)	0.08^{**} (0.04)	-0.06 (0.10)	-0.05 (0.04)
Other	-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 14: Format switching: stay in format

	Portfolio	AC	Rock	CHR	Urban Alt.	News Talk	Country	Spanish	Other
Owner	AC	**−0.10**	8.28*	2.17	4.49	5.16	6.53	−0.72	−0.29
		(4.72)	(4.88)	(5.09)	(5.03)	(4.75)	(4.87)	(5.45)	(4.67)
	Rock	1.70	**−1.87**	8.58*	0.61	3.61	1.19	−8.06	−2.66
		(4.80)	(4.93)	(5.16)	(5.16)	(4.84)	(4.98)	(5.81)	(4.74)
	CHR	−2.59	−3.20	**−15.52****	−1.78	4.08	−1.79	−12.38*	−11.40*
		(6.20)	(6.52)	(6.74)	(6.79)	(6.23)	(6.42)	(7.50)	(6.13)
	Urban Alt.	1.15	−0.59	3.01	**2.95**	1.16	0.97	−7.58	−0.76
		(5.06)	(5.36)	(5.66)	(5.24)	(5.15)	(5.41)	(6.71)	(4.88)
	News Talk	11.79**	10.45**	13.20**	5.02	**8.38***	11.06**	4.33	8.17
		(5.03)	(5.17)	(5.27)	(5.28)	(5.08)	(5.18)	(5.79)	(5.00)
	Country	−0.10	4.49	4.05	−7.81	0.13	**−2.21**	−4.70	0.77
		(4.48)	(4.65)	(4.83)	(5.02)	(4.52)	(4.55)	(5.48)	(4.41)
	Spanish	−11.55**	0.32	4.58	−5.63	−5.84	−10.12*	**15.08*****	−4.11
		(5.43)	(5.91)	(5.63)	(5.75)	(5.25)	(5.97)	(4.40)	(4.93)
	Other	−4.03	−4.27	−3.84	−2.83	−1.47	−3.31	−9.04**	**−2.56**
		(3.45)	(3.62)	(3.84)	(3.80)	(3.53)	(3.60)	(4.15)	(3.34)
	Dark	−0.53	−0.55	0.16	0.27	−0.46	−0.60	−0.02	−0.53
		(0.37)	(0.43)	(0.36)	(0.30)	(0.42)	(0.45)	(0.24)	(0.33)
Top2	AC	**−0.46**	1.69	1.22	1.66	2.63	2.43	3.03	1.98
		(2.30)	(2.46)	(2.62)	(2.57)	(2.31)	(2.40)	(2.55)	(2.20)
	Rock	3.42	**−0.20**	4.83	2.93	5.11*	4.46	3.07	3.45
		(2.80)	(3.00)	(3.16)	(3.05)	(2.82)	(2.92)	(3.04)	(2.71)
	CHR	3.07	4.53	**−1.04**	4.99	4.64	3.93	5.11	3.99
		(3.34)	(3.59)	(3.90)	(3.67)	(3.38)	(3.52)	(3.70)	(3.22)
	Urban Alt.	0.20	0.94	−4.48	**−0.58**	0.42	0.46	3.66	1.52
		(3.10)	(3.30)	(3.59)	(3.29)	(3.11)	(3.23)	(3.39)	(2.97)
	News Talk	−0.23	0.35	−3.91	−2.64	**−2.02**	−1.13	−1.65	0.28
		(2.28)	(2.46)	(2.63)	(2.52)	(2.30)	(2.39)	(2.52)	(2.19)
	Country	−1.72	0.10	0.79	−1.57	−0.55	**−5.42****	−1.51	−1.06
		(2.45)	(2.63)	(2.74)	(2.72)	(2.49)	(2.57)	(2.74)	(2.35)
	Spanish	4.03	1.70	3.47	−0.24	0.98	4.07	**−1.77**	3.67
		(3.95)	(4.31)	(4.48)	(4.29)	(3.88)	(4.14)	(3.85)	(3.72)
	Other	5.00**	3.64	5.99**	4.38*	4.85**	4.93**	1.79	**3.98***
		(2.32)	(2.47)	(2.62)	(2.53)	(2.33)	(2.43)	(2.62)	(2.24)
	Dark	−0.15	0.23	0.18	0.19	−0.08	−0.21	−0.24	−0.11
		(0.43)	(0.48)	(0.48)	(0.48)	(0.43)	(0.43)	(0.44)	(0.41)

Standard errors (corrected for sequential estimation) in parentheses

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

Table 15: Format switching: porfolios

	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 16: Format switching: demographics

	Total	More mergers	Less mergers	More switching	Less switching	Nonnegative profits
Starting point	4.00	1.00	1.00	1.00	1.00	0.00
Fixed cost	7.91	0.68 ↓	1.78 ↑	0.96	0.91	1.43 ↑
Selection + Fixed cost	4.26	0.92 ↑	0.90 ↓	0.98	0.94	0.20 ↓
Selection + Fixed cost + Cost synergies	3.83	1.15 ↔	0.68 ↔	1.01	0.99	0.00 ↔
Selection + Fixed cost + Cost synergies + Switching cost	3.67	0.99 ↔	0.79 ↔	1.11 ↔	0.78 ↔	0.00 ↔

Table 17: Identification of model parameters; Objective function for different counterfactual strategies.

D On-line appendix

D.1 Static model

D.1.1 Listeners

I assume that each listener chooses only one radio station to listen to at a particular moment. Suppose that s is a set of active stations in the current market at a particular time. For any radio station $j \in s$, I define a vector $\iota_j = (0, \ldots, 1, \ldots, 0)$ where 1 is placed in a position that indicates the format of station j.

The utility of listener i listening to station $j \in s$ is given by

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

where θ^L_{2i} is the individual listener's demand sensitivity to adverting, q_j the amount of advertising, ξ_j the unobserved station quality, ϵ_{ji} an unobserved preference shock (distributed type-1 extreme value), and finally θ^L_{1i} is a vector of the individual listener's random effects representing preferences for formats.

I assume that 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 unobserved taste shock, and Π is the matrix representing the correlation between demographic characteristics and format preferences. I assume that draws for ν_i are uncorrelated across time and markets.

The random effects model allows for fairly flexible substitution patterns. For example, if a particular rock station increases its level of advertising, the model allows for consumers to switch proportionally to other rock stations depending on demographics.

Following Berry, Levinsohn, and Pakes (1995), I can decompose the utility into a part that does not vary with consumer characteristics

$$\delta_j = \delta(q_j|\iota_j, \xi_j, \theta^L) = \theta^L_1 \iota_i - \theta^L_2 q_j + \theta^L_3 \mathrm{FM}_j + \xi_j$$

an interaction part

$$\mu_{ji} = \mu(\iota_j, q_j, \Pi D_i, \nu_i) = (\Pi D_i + \nu_{1i})\iota_j + \nu_{2i} q_j$$

and error term ϵ_{ji}.

Given this specification, and the fact that ϵ_{ji} is distributed as an extreme value, one can derive the expected station rating conditional on a vector of advertising levels q, market structure s, a vector of unobserved station characteristics ξ, and market demographic characteristics d,

$$r_j(q|s, \xi, d, \theta^L) = \int\int \frac{\exp[\delta_j + \mu_{ji}]}{\sum_{j' \in s} \exp[\delta_{j'} + \mu_{j'i}]} dF(\nu_i) dF_m(D_i|d)$$

D.1.2 Advertisers

In this subsection I present the details of the demand for advertising. The model captures several important features specific to the radio industry. In particular, the pricing is done on a per-listener basis, so that the price for a 60sec slot of advertising is a product of cost-per-point (CPP) and station rating (market share in percents). Moreover, since radio stations have direct market power over advertisers CPP is a decreasing function of the ad quantities offered by a station and its competitors. The simplest model that captures these features and is a good approximation of the industry is a linear inverse demand for advertising, such as

$$p_j = \theta_1^A r_j \left(1 - \theta_2^A \sum_{f' \in \mathbb{F}} \omega_{ff'}^m q_{f'}\right) \tag{D.2}$$

where f is a format of station j, θ_1^A is a scaling factor for the value of advertising, θ_2^A is a market power indicator and $\omega_{ff'} \in \Omega$ are weights indicating competition closeness between formats f and f'.

The weights ω are a key factor determining competition between formats and thus market power. They reflect the fact that some formats are further and others are closer substitutes for advertisers because of differences in the demographic composition of their listeners. In principle, one could proceed by estimating these weights from the data. However, here it is not feasible to do that because the available data do not contain radio station level advertising prices. Instead, I make additional assumptions that will enable me to compute the weights using publicly available data. The remainder of this subsection discusses the formula for the weights and provides an example supporting this intuition.

Let there be $\mathcal{A}$ types of advertisers. Each type $a \in \mathcal{A}$ targets a certain demographic group(s) a. I.e. an advertiser of type a gets positive utility only if a listener of type a hears an ad. Denote $r_{f|a}$ to be the probability that a listener of type a chooses format f and $r_{a|f}$ to be the probability that a random listener of format f is of type a. Advertisers take these numbers, along with station ratings r_j, as given and decide on which station to advertise. This assumption is motivated by the fact that about 75% is purchased by small local firms. Such firms' advertising decisions are unlikely to influence prices and station ratings in the short run.

This decision problem results in an inverse demand for advertising with weights $\omega_{jj'}$, that are given by

$$\omega_{ff'} = \frac{1}{\sum_{a \in \mathcal{A}} r_{a|f}^2} \sum_{a \in \mathcal{A}} r_{a|f} \left(r_{a|f} r_{f'|a} \right) \tag{D.3}$$

The intuition behind this is that the total impact on the per-listener price of an ad in format f is a weighted average of impacts on the per-listener value of an ad for different types of advertisers. The weighting uses the conditional probabilities of advertisers' arrival, which are equal to the conditional probability of listeners' arrival $r_{a|f}$. For each advertiser of type a the change of value of an ad in format f, in response to a change of total quantity supplied in format f', is affected by two things: it is proportional to the probability of correct targeting in format f, given by $r_{a|f}$, because advertisers are expected utility maximizers; and it is proportional to the share of advertising purchased by this advertiser in format f', given by $r_{f'|a}$. Assembling these pieces together and normalizing the weights to sum to 1 gives equation (D.3). Extensive discussions with examples of the weights as well micro foundations for the model can be found in Jeziorski (2011).

In the next section I will combine demand for programming and advertising to compose the profits of the radio station owners.

D.1.3 Radio station owners

In this subsection I will describe a profit maximizing problem for the radio station owners. Given the advertising quantity choices of competing owners q_{-k}, the profit of radio station owner k is given by

$$\begin{aligned}\bar{\pi}_k(q_k|q_{-k},\xi,\theta) &= \max_{\{q_j;j\in s_k\}}\sum_{j\in s_k} r_j(q|\xi,\theta^L)p_jq_j - \mathrm{MC}_j(q_j) = \\ &= \theta_1^A \max_{\{q_j;j\in s_k\}}\sum_{j\in s_k} q_jr_j(q|\xi,\theta^L)\left(1-\theta_2^A\sum_{f'\in\mathbb{F}}\omega_{ff'}^m q_{f'}\right) + \mathrm{C}_j(q_j|\theta^A,\theta^C)\end{aligned} \tag{D.4}$$

where $C_j(q_j)$ is the total cost of selling advertising. I assume constant marginal cost and allow for a firm level of unobserved cost heterogeneity η_j, i.e. $\mathrm{C}_j(q_j|\theta^A,\theta^C) = \theta_1^A[\theta^C+\eta_j]q_j$.

I assume that the markets are in a Cournot Nash Equilibrium. The first order conditions for profit optimization become

$$r_jp_j + \sum_{j'\in s_k} q_{j'}\left[\frac{\partial r_{j'}}{\partial q_j}p_{j'} - r_{j'}\theta_2^A\omega_{jj'}^m\right] - \theta^C - \eta_j = 0 \quad \forall k \text{ and } j\in s_k \tag{D.5}$$

Additionally, I assume that station unobserved quality is exogenous but serially correlated. It evolves according to an AR(1) process such that

$$\xi_j^t = \rho\xi_j^{t-1} + \zeta_j^t \tag{D.6}$$

where ζ_j^t is an exogenous innovation to station quality.

D.2 Estimation

The estimation of the model is done in two steps. In the first step, I estimate the demand model that includes parameters of the consumer utility θ^L (see equation (D.1)) and the unobserved station quality lag parameter ρ (see equation (D.6)). In the second step, we recover parameters of the inverse demand for advertising θ^A, $w_{jj'}$ (see equation (D.2)) and cost parameters θ^C (see equation (D.4)).

D.2.1 First stage

First stage provides the estimates of demand for radio programming θ^L. Estimation is done using the generalized method of simulated moments. I use two sets of moment conditions. The first set is based on the fact that innovation to station unobserved quality ξ_j has a mean of zero conditional on the instruments:

$$E[\xi_{jt} - \rho\xi_{jt-1}|Z_1,\theta^L] = 0 \tag{D.7}$$

This moment condition follows Berry, Levinsohn, and Pakes (1995) and extends it by explicitly introducing auto-correlation of ξ. I use instruments for advertising quantity since it is likely to be correlated with unobserved station quality. My instruments include: lagged mean and second central moment of competitors' advertising quantity, lagged market HHIs and lagged number and cumulative market share of other stations in the same format. These are valid instruments under the assumption that ξ_t follows an AR(1) process and the fact that decisions about portfolio selection are made before decisions about advertising.

A second set of moment conditions is based on demographic listenership data. Let R_{fc} be the national market share of format f among listeners possessing certain demographic characteristics c. The population moment conditions are

$$\int_t \int_{(D_{ic}^t, m)} \int_{\nu_i} \frac{\exp[\delta_j^{mt} + \mu_{ji}^{mt}]}{\sum_{j' \in s^{mt}} \exp[\delta_{j'}^{mt} + \mu_{ij'}^{mt}]} dF(\nu_i) dF_c^t(D_{ic}^t, m) dt = R_{fc} \tag{D.8}$$

where $F_c^t(D_i, m)$ is a national distribution of people who possess characteristic c at time t. Each person is characterized by the demographic characteristics D_i and the market m they belong to.

For each time t and demographic characteristic c, I draw $\mathcal{I}$ observation pairs (D_{ic}^t, m) from the nationally aggregated CPS. Let $g = (g_1, g_2)$ represent the empirical moments and W be a weighting matrix. I estimate the model by using the constrained optimization procedure:

$$\min_{\theta^L, \xi, g} g' W g$$

Subject to:

$$\begin{aligned} &\hat{r}_{jmt}(q_{mt} | s_{mt}, \xi_{mt}, d_{mt}, \theta^L) = r_{jmt} \quad \forall t, m \\ &\frac{1}{\mathcal{T}\mathcal{I}} \sum_t \sum_{(D_{ic}^t, m)} \int_{\nu_i} \frac{\exp[\delta_j^{mt} + \mu_{ji}^{mt}]}{\sum_{j' \in s^{mt}} \exp[\delta_{j'}^{mt} + \mu_{ij'}^{mt}]} dF(\nu_i) - R_{fc} = g_1 \quad \forall c \\ &\frac{1}{\text{size of } \xi} Z_1(\xi - \rho L \xi) = g_2 \end{aligned} \tag{D.9}$$

where L is a lag operator that converts the vector ξ into one-period lagged values. If the radio station did not exist in the previous period, the lag operator has a value of zero. Integration with respect to demographics when calculating the first constraint is obtained by drawing from the CPS in the particular market and period. This way of integrating allows us to maintain proper correlations between possessed demographic characteristics. The same is true when obtaining the data set D_{ict}. When computing the interaction terms μ in the second constraint, I draw one vector ν_i from the normal distribution for each D_{ict}.

D.2.2 Second stage

The second stage of the estimation obtains the competition matrix Ω and the parameters of demand for advertising θ^A. The estimation is done separately for every market, thereby allowing for different Ω and θ^A.

To compute the matrices Ω^m for each market I use the specification layed out in section D.1.2. The elements of the matrix Ω are specified as

$$\omega_{ff'} = \frac{1}{\sum_{a\in\mathcal{A}} r_{a|f}^2} \sum_{a\in\mathcal{A}} r_{a|f}\left(r_{a|f} r_{f'|a}\right)$$

following equation (D.3). The $r_{f|a}$ are advertisers' beliefs about listeners' preferences for formats. These are constant across markets. To recognize that advertisers know the demographic composition of each market I allow for market specific conditional probabilities of listeners' arrival for each format $r^m_{f|a}$. However, I assume that the advertisers compute those values by using Radio Today reports and the Current Population Survey. After computing weights, I treat Ω^m as exogenous and fixed in all of the following steps.[22]

After computing matrices Ω, I estimate θ^A. Using estimates of demand for radio programming θ^L from the first stage, I compute ratings for each station conditioned on the counterfactual advertising quantities. I use the set of $3M$ moment conditions

$$E_m[\eta^m | Z_2, \theta^A, \theta^C] = 0 \quad \forall m \in \mathbf{M} \tag{D.10}$$

where the integral is taken with respect to time and stations in each market. η_j^{tm} is an unobserved shock to marginal cost defined in equation (D.2). The Z_2 are three instruments: a column of ones, the AM/FM dummy and the number of competitors in the same format. They are uncorrelated with η^m under the IID assumption, but are correlated with the current choice of quantity because they describe the market structure.

We back out η_j^{tm} using FOCs for owner's profit maximization (see equation (D.4))

$$\eta_j^t = r_j^t p_j^t + \sum_{j'\in s_k^{tm}} q_{j'}^t \left[\frac{\partial r_{j'}^t}{\partial q_j^t} p_{j'}^t - \theta_{2m}^A r_{j'}^t \omega_{ff'}^m\right] - \theta_m^C \quad \forall t \in \mathbf{T}, k \in \mathbf{K}^{tm}, j \in s_k^{tm} \tag{D.11}$$

[22] Such an approach potentially ignores possible variance of the Ω^m estimator. The source of this variance might come from the finiteness of the CPS dataset and the distribution of Arbitron estimates.

Since the equation does not depend on θ^A_{1m}, I can use it to estimate θ^A_{2m} and θ^C_m. During the estimation, I allow for a different value of marginal cost for each market. I allow for 3 different values for the slope of inverse demand depending on the population of the market (up to 500 people, between 500 and 1500, and 1500 or more). Ratings and derivatives of ratings in the equation (D.11) are calculated using the estimates of θ^L and ξ from the first stage. Demographic draws are taken from the CPS and are independent of those used in the first stage. Given the estimates of θ^A_{2m} and θ^C, I can back out θ^A_{1m} by equating the observed average revenue in each market with its predicted counterpart.

Next I discuss a variation in the data that identifies parameters θ^A and θ^C. The intuition for such identification is that estimating Equation D.11 can be regarded as a linear regression in which θ^C_m is an intercept and θ^A_2 is a coefficient of a variable that is a function of supplied quantity. In this case, the mean deviation of FOCs from zero in each market identifies the intercept θ^C_m. The slope parameter θ^A_2 is identified by the size of the response of the firm to changes in quantity supplied by its competitors due to change in the market structure or demographics. Such a response, is composed of listeners' demand feedback and the direct effect of quantity on CPP. Elasticity of listeners' demand, which determines the strength of the feedback, is consistently estimated in the first step. Therefore, one can subtract the difference out of the feedback effect from the total response observed in the data. This allows us to obtain the strength of the direct effect that directly identifies the slope of the CPP, θ^A_2. For example, if we look at the response of ad quantity reacting to the merger, the slope of listeners' demand alone predicts large increases in ad quantity. However, in the data we observe smaller increases or even decreases in the quantity supplied, depending on the market. Those differences are rationalized by a negative value of CPP slope, θ^A_2.

D.3 Results

This section presents estimates of the structural parameters. The next subsection discusses listeners' demand parameters. This is followed by results concerning advertisers' demand and marker power. The last subsection contains estimates of marginal cost and profit margin (before subtracting fixed cost).

D.3.1 Listeners' demand

Table 18 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 is consistent with the belief that radio listeners have a disutility for advertising. When it comes to the mean effects of programming formats, Contemporary Hit Radio format gives the most utility, while the News/Talk format gives the least.

The second column of Table 18 contains variances of random effects for station formats. The higher a format's variance, the more persistent are the tastes of listeners for that format. For example, in response to an increased amount of advertising, if the variance of the random effect for that format is high, listeners tend to switch to a station of the same format. The estimates also suggest that tastes for the Alternative/Urban format are the most persistent.

Table 19 contains estimates of interactions between listener characteristics and format dummies. The majority of the parameters are consistent with intuition. For example, younger people are more willing to choose a CHR format while older people go for News/Talk. The negative coefficients on the interaction of Hispanic format with education and income suggests that less educated Hispanic people with lower income 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 is consistent with the the fact that Urban radio stations play mostly rap, hip-hop and soul music performed by black artists.

D.3.2 Advertisers' demand

Tables 22 presents the weights for selected markets representing large, medium and small listener populations. They were computed using the 1999 edition of Radio Today publication and Common Population Survey aggregated from 1996 to 2006. I also compute a total impact coefficient that is the sum of all the columns of the table for each format. Not surprisingly, general interest formats like AC and News/Talk have the biggest impact on the price of advertising, while Spanish format has the smallest. The values on the diagonals of the matrices represent the formats' own effect of the quantity of advertising supplied on per-listener price. They are usually bigger than the off-diagonal values, that suggests that it is mostly the ad quantity in the same format that influences a per-listener price. In accord with an intuition, the formats with the most demographically

	Mean Effects (θ_1^L)	Random Effects (Σ_1)
Advertising	-1.106 (0.002)	0.030 (0.009)
AM/FM	0.861 (0.000)	-
AC, SmoothJazz, and New AC	-2.431 (0.008)	0.043 (0.004)
Rock	-1.559 (0.140)	0.004 (0.020)
CHR	-0.179 (0.025)	0.009 (0.006)
Alternative Urban	-2.339 (0.026)	0.348 (0.008)
News/Talk	-4.678 (0.010)	0.024 (0.002)
Country	-2.301 (0.006)	0.011 (0.003)
Spanish	-1.619 (0.004)	0.011 (0.001)
Other	-4.657 (0.004)	0.005 (0.002)
ρ	0.568 (0.091)	-

Table 18: Estimates of mean and random effects of demand for radio programming.

	Demographics characteristics (Π)					
	Age	Sex	Education	Income	Black	Spanish
AC, SmoothJazz, and New AC	−0.171 (0.001)	−0.341 (0.064)	0.602 (0.013)	−0.024 (0.003)	0.121 (0.012)	−1.014 (0.008)
Rock	−0.645 (0.072)	0.399 (0.031)	0.861 (0.006)	−0.147 (0.045)	−1.359 (0.007)	−1.643 (0.003)
CHR	−2.541 (0.015)	0.477 (0.080)	1.772 (0.006)	−0.291 (0.005)	1.946 (0.015)	0.463 (0.001)
Alternative Urban	−0.817 (0.008)	1.350 (0.018)	0.583 (0.025)	−0.141 (0.002)	3.152 (0.005)	0.267 (0.027)
News/Talk	0.329 (0.002)	1.228 (0.012)	0.237 (0.009)	0.093 (0.005)	−0.321 (0.001)	−1.649 (0.005)
Country	0.062 (0.004)	−0.149 (0.022)	0.133 (0.004)	−0.125 (0.003)	−1.548 (0.009)	−1.717 (0.002)
Spanish	−0.024 (0.013)	−0.908 (0.012)	−0.328 (0.018)	−1.140 (0.002)	−2.560 (0.004)	0.797 (0.003)
Other	0.263 (0.373)	0.624 (0.003)	0.338 (0.006)	−0.031 (0.063)	0.498 (0.001)	0.238 (0.002)

Table 19: Interaction terms between listeners' demographics and taste for radio programming.

homogenous listener pools, Urban/Alternative and Spanish, have the highest values of the own effects. On the other hand, general interest formats like CHR and Rock are charaterized by the smallest values of the own effect, measuring the fact that their target population of listeners is more dispersed across other formats. For cross effects, one notices that News/Talk is close to AC and Urban is close to CHR. This can be explained by, for example, the age of the listeners. In the former case the formats appeal to an older population while in the latter case to a younger one.

Estimates of the slope of inverse demand are presented in Table 20. In markets with less than 0.5m people radio stations have considerable control over the per-listener price. However, such control significantly drops in markets from 0.5m to 2m people, and it disappears completely in markets with more than 2m people, making radio stations essentially price takers.

D.3.3 Supply

The marginal costs of selling advertising minutes are presented in Table 21. The values of this cost range from $356 per minute of advertising sold in Los Angeles, CA to $11 in Ft. Myers, FL. 66% of the variation in marginal cost can be explained by variation in market population. A population increase of one thousand translates to about a 2 cent increase in marginal cost (with t-stat equal to 12). The high correlation between population and marginal costs can be explained by the fact that revenues per-minute of advertising are an increasing function of total market population. Suppose this surplus is split between radio station owners and advertisers' sales people according to the Nash Bargaining solution. In this case, the high correlation of revenue with population will translate into a high correlation of marginal cost with population.

From the revenues and marginal cost estimates, I can calculate variable profit margins. These are presented in the last last column of Table 21. The range is from 92% in Shreveport, LA to 15% in Honolulu, HI and Reno, NV. 38% of the profit margin variation can be explained by the variance in total ad quantity supplied and markets with high profit margins firms supply more advertising. The marginal effect of extra minute per day of broadcasted advertising translates into 0.6% of extra profit margin.

Market population	less than .5m	between .5m and 1.5m	more than 1.5m
	1.34 (0.046)	0.35 (0.026)	0.00 (0.008)

Table 20: Slope of the inverse demand for ads θ_2^A, by market size

Market	Population (mil)	Marginal cost ($ per-miute)	Profit margin	Market	Population	Marginal cost	Profit margin
Los Angeles, CA	13,155	356.4 (5.15)	30%	Tulsa, OK	856	72.8 (2.13)	21%
Chicago, IL	9,341	180.0 (2.70)	34%	Knoxville, TN	785	54.3 (1.99)	27%
Dallas-Ft. Worth, TX	5,847	198.6 (5.60)	28%	Albuquerque, NM	740	27.4 (1.04)	36%
Houston-Galveston, TX	5,279	199.7 (4.20)	28%	Ft. Myers-Naples-Marco Island, FL	737	11.3 (0.94)	57%
Atlanta, GA	4,710	95.4 (3.37)	43%	Omaha-Council Bluffs, NE-IA	728	48.0 (0.91)	28%
Boston, MA	4,532	172.2 (3.68)	33%	Harrisburg-Lebanon-Carlisle, PA	649	29.7 (1.44)	42%
Miami-Ft, FL	4,174	134.3 (3.70)	28%	El Paso, TX	619	41.8 (4.12)	20%
Seattle-Tacoma, WA	3,776	128.7 (2.21)	29%	Quad Cities, IA-IL	618	51.3 (1.30)	23%
Phoenix, AZ	3,638	63.7 (1.84)	39%	Wichita, KS	598	38.9 (0.85)	25%
Minneapolis-St. Paul, MN	3,155	160.8 (4.66)	26%	Little Rock, AR	577	45.2 (1.64)	26%
St. Louis, MO	2,689	190.6 (5.38)	18%	Columbia, SC	577	60.0 (2.10)	23%
Tampa-St, FL	2,649	102.7 (2.09)	26%	Charleston, SC	569	59.6 (1.74)	19%
Denver-Boulder, CO	2,604	99.9 (1.40)	32%	Des Moines, IA	564	21.3 (0.92)	40%
Portland, OR	2,352	48.6 (1.35)	41%	Spokane, WA	540	24.5 (0.63)	28%
Cleveland, OH	2,134	170.6 (3.34)	24%	Madison, WI	520	93.6 (3.02)	22%
Charlotte, NC-SC	2,127	67.1 (1.96)	38%	Augusta, GA	510	30.9 (0.60)	24%
Sacramento, CA	2,100	47.6 (1.30)	42%	Ft. Wayne, IN	509	37.8 (1.35)	27%
Salt Lake City, UT	1,924	58.1 (1.19)	26%	Lexington-Fayette, KY	495	36.8 (1.59)	35%
San Antonio, TX	1,900	75.0 (2.27)	24%	Chattanooga, TN	471	41.5 (2.53)	29%
Kansas City, MO-KS	1,871	152.5 (2.87)	19%	Boise, ID	469	46.2 (3.73)	30%
Las Vegas, NV	1,752	47.7 (1.49)	32%	Jackson, MS	453	18.6 (2.03)	59%
Milwaukee-Racine, WI	1,713	74.6 (1.27)	25%	Eugene-Springfield, OR	439	27.4 (1.29)	31%
Orlando, FL	1,686	42.4 (1.77)	41%	Reno, NV	400	99.7 (1.64)	15%
Columbus, OH	1,685	70.2 (1.53)	30%	Shreveport, LA	359	19.8 (4.25)	92%
Indianapolis, IN	1,602	86.8 (2.32)	26%	Fayetteville, NC	337	38.1 (2.48)	46%
Norfolk, VA	1,583	196.8 (4.64)	17%	Springfield, MA	336	20.8 (0.87)	55%
Nashville, TN	1,342	40.5 (1.84)	38%	Macon, GA	276	34.4 (2.29)	26%
Greensboro-Winston, NC	1,329	53.5 (2.34)	32%	Binghamton, NY	255	37.5 (1.51)	27%
New Orleans, LA	1,294	91.2 (2.44)	24%	Lubbock, TX	248	57.7 (1.98)	18%
Memphis, TN	1,278	53.2 (1.82)	30%	Odessa-Midland, TX	231	21.4 (0.99)	27%
Jacksonville, FL	1,271	66.1 (1.64)	29%	Fargo-Moorhead, ND-MN	200	48.6 (2.42)	25%
Oklahoma City, OK	1,268	75.6 (1.35)	25%	Medford-Ashland, OR	184	27.7 (0.90)	28%
Buffalo-Niagara Falls, NY	1,150	141.5 (3.63)	19%	Duluth-Superior, MN-WI	159	43.3 (0.79)	20%
Louisville, KY	1,100	92.9 (2.36)	21%	Parkersburg-Marietta, WV-OH	157	31.7 (1.41)	21%
Richmond, VA	1,066	55.3 (1.47)	28%	Abilene, TX	149	23.0 (1.14)	26%
Birmingham, AL	1,030	85.8 (2.50)	24%	Eau Claire, WI	149	31.6 (2.77)	28%
Honolulu, HI	938	78.2 (2.39)	15%	Williamsport, PA	130	31.0 (1.13)	23%
Albany, NY	909	113.9 (3.18)	16%	Monroe, LA	124	14.2 (1.49)	64%
Grand Junction, CO	902	24.5 (0.67)	24%	Sioux City, IA	118	26.1 (0.96)	24%
Tucson, AZ	870	41.1 (0.93)	27%	San Angelo, TX	104	26.4 (1.36)	16%
Grand Rapids, MI	864	37.9 (0.79)	38%	Bismarck, ND	99	32.8 (1.65)	22%

Table 21: Estimated marginal cost (in dollars per minute of broadcasted advertising) and profit margins (before subtracting the fixed cost) for a chosen set of markets

Los Angeles, CA

	AC SmoothJazz New AC	Rock	CHR	Alternative Urban	News/Talk	Country	Spanish	Other
AC SmoothJazz New AC	**0.22**	0.10	0.11	0.09	0.17	0.14	0.00	0.17
Rock	0.15	**0.21**	0.12	0.09	0.16	0.13	0.01	0.12
CHR	0.18	0.12	**0.16**	0.16	0.10	0.13	0.03	0.13
Alternative Urban	0.11	0.05	0.17	**0.44**	0.06	0.05	0.00	0.12
News/Talk	0.17	0.10	0.05	0.05	**0.30**	0.13	0.00	0.21
Country	0.16	0.10	0.09	0.07	0.15	**0.22**	0.01	0.21
Spanish	0.03	0.04	0.11	0.02	0.01	0.03	**0.72**	0.04
Other	0.18	0.07	0.06	0.08	0.20	0.17	0.00	**0.23**
Total impact	1.20	0.79	0.87	0.99	1.15	1.00	0.77	1.23

Atlanta, GA

	AC SmoothJazz New AC	Rock	CHR	Alternative Urban	News/Talk	Country	Spanish	Other
AC SmoothJazz New AC	**0.20**	0.10	0.12	0.09	0.14	0.18	0.00	0.18
Rock	0.14	**0.21**	0.13	0.10	0.12	0.17	0.01	0.13
CHR	0.17	0.13	**0.17**	0.14	0.09	0.17	0.01	0.13
Alternative Urban	0.11	0.06	0.16	**0.40**	0.06	0.08	0.00	0.13
News/Talk	0.16	0.10	0.05	0.05	**0.25**	0.17	0.00	0.22
Country	0.15	0.09	0.08	0.06	0.13	**0.26**	0.01	0.22
Spanish	0.04	0.04	0.12	0.02	0.01	0.03	**0.71**	0.03
Other	0.16	0.07	0.06	0.07	0.16	0.23	0.01	**0.25**
Total impact	1.11	0.78	0.88	0.94	0.95	1.31	0.75	1.29

Knoxville, TN

	AC SmoothJazz New AC	Rock	CHR	Alternative Urban	News/Talk	Country	Spanish	Other
AC SmoothJazz New AC	**0.20**	0.11	0.16	0.11	0.10	0.16	0.01	0.16
Rock	0.13	**0.21**	0.14	0.11	0.10	0.18	0.01	0.12
CHR	0.16	0.12	**0.18**	0.14	0.08	0.17	0.02	0.13
Alternative Urban	0.12	0.06	0.16	**0.38**	0.06	0.08	0.00	0.13
News/Talk	0.16	0.13	0.10	0.09	**0.17**	0.16	0.01	0.18
Country	0.15	0.13	0.14	0.10	0.09	**0.22**	0.01	0.16
Spanish	0.05	0.05	0.11	0.02	0.02	0.04	**0.66**	0.05
Other	0.17	0.09	0.11	0.12	0.12	0.18	0.01	**0.21**
Total impact	1.12	0.90	1.11	1.05	0.74	1.21	0.72	1.14

Table 22: Product closeness matrices for chosen markets