

SUPPLEMENT TO “DYNAMIC PRODUCT POSITIONING IN
DIFFERENTIATED PRODUCT MARKETS: THE EFFECT OF FEES
FOR MUSICAL PERFORMANCE RIGHTS ON THE COMMERCIAL
RADIO INDUSTRY”

(*Econometrica*, Vol. 81, No. 5, September 2013, 1763–1803)

BY ANDREW SWEETING

APPENDIX A: STATE VARIABLES

TABLE A-I LISTS ALL OF THE STATE VARIABLES IN THE MODEL. As some station characteristics are assumed to be fixed and ownership is assumed not to change over time, a state is firm-specific, and the value of the state variables will depend on the characteristics of the stations owned by the firm (some of which are fixed), the characteristics of the firm’s competitors (the set of competitors is fixed), and market characteristics. The table notes (i) whether a variable is fixed over time, (ii) the information assumption made about the variable, and (iii) whether the value is observed or estimated by the researcher in the first stage.

APPENDIX B: SOLUTION METHOD FOR COUNTERFACTUALS

This appendix details the method for solving the model for the counterfactuals. As markets are independent in my model, I solve the model “market-by-market,” so I do not impose that the approximation to the value function is the same across markets that differ greatly in size, although the structural parameters only vary by market-size group. I also fix market demographics at their initial values.

Selection of States. It is necessary to solve for values and policies at a fixed subset of N states because some state variables (e.g., unobserved station quality) are continuous and the state space is large. For each market, I choose the states that are observed in the data, and then create 499 duplicates of the observed states, where the formats of the stations and their unobserved qualities are perturbed. In each duplication, the unobserved quality of each station is chosen as a uniform random draw on $[-2, 2]$, a range which comprises almost all of the values of ξ_{st} in the data. When there are no fees, the probability that a station’s format is the same as in the data is 0.3, the probability that the station is Dark is 0.05, and otherwise a new active format is drawn where the probability of each format is the same. With fees, it is likely that markets will evolve to situations with more non-music stations, and it is desirable to approximate the value function more accurately in these states. When the format of a station is to be changed, I therefore make the probability of choosing

TABLE A-I
STATE VARIABLES

	Evolution	Information?	Observed?
<i>Market Variables</i>			
Population in $d = 1, \dots, 18$ mutually exclusive groups	Ethnic group size evolves with growth rates	Public	Observed
Growth rates for black, Hispanic and white populations	AR(1), i.i.d. innovations	Public	Observed, AR(1) process estimated
Advertising prices per listener	Fixed	Public	Estimated
<i>Station Variables (for firm's own stations)</i>			
Format	Changes with choice	Public	Observed
Changed Format in Previous Period	Changes with choice	Public	Observed
Station Quality:			
observed characteristic component (e.g., based on signal coverage, station location)	Fixed	Public	Characteristics observed, quality coefficients estimated
band	Fixed	Public	Observed (interaction with format estimated)
ξ_{st}	AR(1), i.i.d. innovations	Public	Implied by demand estimates AR(1) process estimated
<i>Station Variables (for each station owned by competitors)</i>			
Owner	Fixed	Public	Observed
Format	Changes with choice	Public	Observed
Changed Format in Previous Period	Changes with choice	Public	Observed
Station Quality:			
observed characteristic component (e.g., based on signal coverage, station location)	Fixed	Public	Characteristics observed, quality coefficients estimated
band	Fixed	Public	Observed (interaction with format estimated)
ξ_{st}	AR(1), i.i.d. innovations	Public	Estimated
<i>Choice-Specific Payoff Shocks</i>			
ε for each choice for each firm	i.i.d. across firms, choice & time	Private	No, scale estimated

each non-music format twice as large as the probability of choosing each music format.

In the description of the solution procedure, a particular state j is denoted as $\mathcal{M}_{j,o,t}$, where o indicates the firm of interest in the state and t denotes the initial period.

Variables Used in Approximating the Value Function for the Counterfactual. I assume that firm value functions can be approximated by a linear parametric function of functions of the state variables. Given the size of the state space, it is not possible to include interactions between all of the state variables listed in Table A-I. Instead, I use a smaller number of functions, which I now detail. In the descriptions of the variables and the solution method, I will refer to three alternative measures of revenues:¹

- station current revenues (R_s),
- station “no ξ ” revenue: a station’s predicted revenues in a format given demographics, formats, and station characteristics when the time-varying ξ_{st} ’s of all stations are set equal to zero,
- “*revenue*”: a measure of the station’s average revenue potential excluding the ξ_{st} ’s formed by averaging the revenues that it would get across a large set of format configurations for all stations.

These last two measures can provide somewhat better measures of a station’s long-term revenue potential.

The approximating function for a state $\mathcal{M}_{j,o,t}$ includes the following variables² (note that demographics and market size are not included because they are held fixed over time in the counterfactual):

- current firm no ξ revenues,
- the following variables on their own and interacted with the number of stations owned by firm o :
 - sum of *revenues* for firm o ’s stations
 - the sum of the $\exp(\xi_{st})$ measures for firm o ’s stations
 - the sum of the $\exp(X_{st}\gamma^S)$ measures for firm o ’s stations (excluding the AM * format component)
 - the sum of the AM dummies for firm o ’s stations and the sum of those dummies for stations that o has in non-music formats
 - the sum of the AM dummies interacted with $\exp(X_{st}\gamma^S)$ (excluding the AM * format component) for firm o ’s stations
 - the sum of the AM dummies interacted with $\exp(\xi_{st})$ for firm o ’s stations

¹When performing counterfactuals with fees (calculated as a proportion of revenues), revenues are calculated net of fees.

²These variables were chosen based primarily on a set of Monte Carlo experiments, using a simplified model, described in the Appendices of [Sweeting \(2011\)](#). These experiments revealed that solutions often become less accurate, due to overfitting, when too many variables were included in the approximation, which can also affect numerical stability of the iterative procedure. As the number of states used in the approximation increases, more variables can be included.

- the sum of the $\exp(X_{st}\gamma^S)$ (excluding the AM * format component) interacted with $\exp(\xi_{st})$ for firm o 's stations
 - the sum of $\overline{revenue}$ interacted with $\exp(\xi_{st})$ for firm o 's stations
 - for the largest competitor faced by the firm (call it firm x):³
 - the number of stations owned by x
 - the sum of the $\overline{revenue}$ measures for firm x 's stations
 - the sum of the $\exp(\xi_{st})$ measures for firm x 's stations
 - the sum of interactions between the $\overline{revenue}$ measures and $\exp(\xi_{st})$ for firm x 's stations
 - for the second largest competitor faced by the firm (call it firm y):
 - the number of stations owned by firm y
 - the sum of the $\overline{revenue}$ measures for firm y 's stations
 - the sum of the $\exp(\xi_{st})$ measures for firm y 's stations
 - the sum of interactions between the $\overline{revenue}$ measures and $\exp(\xi_{st})$ for firm y 's stations
 - for each active format (i.e., the coefficients can vary freely across formats):
 - number of rival stations in the format
 - number of rival stations in the format owned by firms that own more than one station
 - sum of rival stations' $\exp(\xi_{st})$'s
 - sum of rival stations' $\exp(X_{st}\gamma^S)$'s (excluding the AM * format component)
 - sum of the $\overline{revenue}$ measures for rival stations
 - sum of the $\overline{revenue}$ measures for rival stations that are in the AM band
 - based on a calculation of the gains in no ξ revenues that firm o could make by moving one of its stations, holding the formats of other stations fixed, a dummy for whether o could make any gains, and the average size of these revenue-gaining opportunities (i.e., the \$ gains are added together and divided by the number of gains).

In the following description of the solution procedure, $\Phi_{j,k}(\mathcal{M}_{j,o,t})$ is the value of the k th approximating variable in state j and Φ is the matrix where these variables are stacked for the N states.

Solution Procedure. An iterative procedure is used to find the coefficients of the parametric function that is used to approximate the value function. I now detail each of the steps in a particular iteration i . In state $\mathcal{M}_{j,o,t}$, $P_o^i(a|\mathcal{M}_{j,o,t})$ is the iteration i guess of the probability that firm o chooses action a , $P_o^i(\mathcal{M}_{j,o,t})$ is the collection of these probabilities, and $P_{-o}^i(\mathcal{M}_{j,o,t})$ is the set of choice probabilities of o 's competitors in the state.

³The size of competitors is determined by the number of stations owned and, where this is equal, the sum of the $\overline{revenue}$ measures for the different firms.

Step 1. For each of the N states, $\tilde{\pi}(P_o^i(\mathcal{M}_{j,o,t}))$ is calculated as

$$(A-1) \quad \tilde{\pi}(P_o^i(\mathcal{M}_{j,o,t})) = \sum_{s \in S^o} R_s(\mathcal{M}_{j,o,t} | \gamma) \\ + \sum_{a \in A_o(\mathcal{M}_{j,o,t})} P_o^i(a | \mathcal{M}_{j,o,t}) (\beta C_o(a) \theta^C \\ - W_o(a) \theta^W + \theta^\varepsilon (\varkappa - \log(P_o^i(a | \mathcal{M}_{j,o,t}))))),$$

where the θ 's are the structural parameters appropriate for the market, \varkappa is Euler's constant, $C_o(a, \mathcal{M}_{j,o,t})$ is the number of stations that the firm will have operating in the same format as one of its other stations in the next period if it chooses action a , and $A_o(\mathcal{M}_{j,o,t})$ is o 's choice set. β is the discount factor. $\tilde{\pi}(P_o^i(\mathcal{M}_{j,o,t}))$ is a function only of o 's choice probabilities. $\tilde{\pi}(P^i)$ is the vector that stacks these values for the N states.

For each of the N states, the choice probabilities of all firms are used to calculate $E_{pi} \Phi^i$, a vector that contains the expected value of each of the approximating variables for the following period given strategies. For a particular k ,

$$(A-2) \quad E_{pi} \phi_{j,k}^i = \int \phi_{h,k}(\mathcal{M}_{h,o,t+1}) \\ \times g(\mathcal{M}_{h,o,t+1} | P_o^i(\mathcal{M}_{j,o,t}), P_{-o}^i(\mathcal{M}_{j,o,t}), \mathcal{M}_{j,o,t}) d\mathcal{M}_{h,o,t+1},$$

where g is the transition density. This integral is approximated by reweighting variables for a prespecified sample of $t + 1$ states, as calculating $\phi_{h,k}(\mathcal{M}_{h,o,t+1})$ requires solving a random coefficients demand model. Specifically, for a given state $\mathcal{M}_{j,o,t}$, I consider a set of H states $\mathcal{M}_{h,o,t+1}$ which is equal to the set of states that can be reached by any move by o , a set of S^ξ draws for innovations in ξ , and $S^{-o,m}$ moves by other firms in the same local market. During the solution procedure, the integral is approximated by

$$(A-3) \quad E_{pi} \phi_{j,k}^i \approx \sum_{h=1}^H \phi_{h,k}(\mathcal{M}_{h,o,t+1}) \\ \times \frac{g(\mathcal{M}_{h,o,t+1} | P_o^i(\mathcal{M}_{j,o,t}), P_{-o}^i(\mathcal{M}_{j,o,t}), \mathcal{M}_{j,o,t})}{\sum_{h'=1}^H g(\mathcal{M}_{h',o,t+1} | P_o^i(\mathcal{M}_{j,o,t}), P_{-o}^i(\mathcal{M}_{j,o,t}), \mathcal{M}_{j,o,t})}.$$

$S^\xi = 10$ and $S^{-o,m} = 500$. To be accurate, the $S^{-o,m}$ moves should include those that are most likely to be made. With no fees, I choose the ones that are most likely to be made based on the first-stage estimates of the conditional choice probabilities. With fees, non-music formats will be more likely to be chosen,

all else equal, and this needs to be accounted for. The first step is to estimate a simpler multinomial logit model using the observed data, where the covariates are the elements of $W_o(a_o)$ and $C_o(a_o)$ and the revenues that the firm would earn in the next period given the particular choice if no other firms made a format switch and unobserved station qualities and demographics remained unchanged. The estimated coefficients in this model are consistent with stations moving toward increased revenues. I then recompute the next period revenue variable taking into account the effects of fees, giving a new set of choice probabilities. I then use these choice probabilities to choose the $S^{-o,m}$ moves that are most likely to be made.⁴

Step 2. Create matrices $(\Phi - \beta E_{P^i} \Phi)$ ($\beta = 0.95$). As the parameters λ are overidentified ($N > K$), use an OLS regression to calculate the coefficients $\widehat{\lambda}^{P^i}$,

$$(A-4) \quad \widehat{\lambda}^{P^i} = ((\Phi - \beta E_{P^i} \Phi)'(\Phi - \beta E_{P^i} \Phi))^{-1} (\Phi - \beta E_{P^i} \Phi)' \widetilde{\pi}(P^i).$$

Step 3. New choice probabilities for each state are calculated using the fixed parameters θ and the multinomial logit choice formula

$$(A-5) \quad P'_o(a|\mathcal{M}_{j,o,t}) \\ = \exp\left(\frac{\text{FV}(a, \mathcal{M}_{j,o,t}, P^i_{-o}(\mathcal{M}_{j,o,t})) - W_o(a)\theta^W + \beta C_o(a)\theta^C}{\theta^\varepsilon}\right) \\ / \sum_{a' \in A_o(\mathcal{M}_{j,o,t})} \exp((\text{FV}(a', \mathcal{M}_{j,o,t}, P^i_{-o}(\mathcal{M}_{j,o,t})) \\ - W_o(a')\theta^W + \beta C_o(a')\theta^C)/\theta^\varepsilon),$$

where

$$(A-6) \quad \text{FV}(a, \mathcal{M}_{j,o,t}, P^i_{-o}(\mathcal{M}_{j,o,t})) \\ = \sum_{h=1}^H \sum_{k=1}^K \phi_{h,k}(\mathcal{M}_{h,o,t+1}) \\ \times \left\{ \frac{g(\mathcal{M}_{h,o,t+1}|a, P^i_{-o}(\mathcal{M}_{j,o,t}), \mathcal{M}_{j,o,t})}{\sum_{h'=1}^H g(\mathcal{M}_{h',o,t+1}|a, P^i_{-o}(\mathcal{M}_{j,o,t}), \mathcal{M}_{j,o,t})} \right\} \widehat{\lambda}_k^{P^i},$$

that is, it reflects the states that can be reached given that action a is chosen. The same formulas are used to calculate updated choice probabilities of

⁴It is also possible to solve the model once using a prespecified set of moves, and then use the implied set of choice probabilities to create a new set of draws. Some experimentation indicated that this approach gives similar results.

competitors, although, in subsequent iterations, the set of competitor moves $S^{-o,m}$ used to calculate expected payoffs is kept the same (i.e., they are simply reweighted).

Step 4. If the maximum absolute difference between P' and P^i is less than $1e-5$, the procedure stops, and the values of $\widehat{\lambda}^{P^i}$ are saved as λ^* . Otherwise, P^i (i.e., both P_o and P_{-o}) is updated as a weighted combination of P^i and P' ,

$$(A-7) \quad P^{i+1} = \psi P' + (1 - \psi)P^i,$$

where $\psi = 0.1$, and Step 1 is repeated for iteration $i + 1$.

To start the procedure, it is necessary to have an initial set of guesses P^1 . I use the first-stage multinomial logit approximation to the CCPs.

Forward Simulation. The solution procedure gives conditional choice probabilities for each firm in the initial (observed) state of the market (period t). These choice probabilities and the AR(1) process for ξ are used to simulate the model forward one period to $t + 1$. In this new configuration, it is necessary to solve for a new set of choice probabilities. This involves a further iterative procedure. Before this procedure begins, a set of H states for $t + 2$ is drawn, and the approximating variables are calculated for these states. The H states are chosen as described in Step 1 above. A set of initial choice probabilities to start the iterative procedure is also required, and, once again, I use the choice probabilities implied by the first-stage multinomial logit approximation to the CCPs. At iteration i , with choice probabilities P^i , the following scheme is followed:

Step 1. $P_{-o}^i(\mathcal{M}_{j,o,t+1})$ is used to calculate the FV value for each possible action by each firm,

$$(A-8) \quad \begin{aligned} & \text{FV}(a, \mathcal{M}_{j,o,t+1}, P_{-o}^i(\mathcal{M}_{j,o,t+1})) \\ &= \sum_{h=1}^H \sum_{k=1}^K \phi_{h,k}(\mathcal{M}_{h,o,t+2}) \\ & \quad \times \left\{ \frac{g(\mathcal{M}_{h,o,t+2}|a, P_{-o}^i(\mathcal{M}_{j,o,t+1}), \mathcal{M}_{j,o,t+1})}{\sum_{h'=1}^H g(\mathcal{M}_{h',o,t+2}|a, P_{-o}^i(\mathcal{M}_{j,o,t+1}), \mathcal{M}_{j,o,t+1})} \right\} \lambda^*, \end{aligned}$$

and the multinomial logit formula is used to calculate new choice probabilities,

$$(A-9) \quad \begin{aligned} & P'_o(a|\mathcal{M}_{j,o,t+1}) \\ &= \exp\left(\frac{\text{FV}(a, \mathcal{M}_{j,o,t+1}, P_{-o}^i(\mathcal{M}_{j,o,t+1})) - W_o(a)\theta^W + \beta C_o(a)\theta^C}{\theta^\varepsilon}\right) \end{aligned}$$

$$\begin{aligned} & / \sum_{a' \in A_o(\mathcal{M}_{j,o,t+1})} \exp((FV(a', \mathcal{M}_{j,o,t+1}, P_{-o}^i(\mathcal{M}_{j,o,t+1}))) \\ & - W_o(a')\theta^W + \beta C_o(a')\theta^C) / \theta^e). \end{aligned}$$

Step 2. If the maximum absolute difference between P' and P^i is less than $1e-6$, the procedure stops. Otherwise, P^i (i.e., both P_o and P_{-o}) is updated as a weighted combination of P^i and P' ,

$$(A-10) \quad P^{i+1} = \psi P' + (1 - \psi) P^i$$

where $\psi = 0.1$, and Step 1 is repeated for iteration $i + 1$.

The converged choice probabilities are used to simulate the model forward to the next period and the procedure is repeated, until the market is advanced for the desired number of periods (40 in the paper).

APPENDIX C: ESTIMATION

As explained in the text, estimation is separated into two stages. The parameters of the listener demand model, the revenue function, and the process governing demographic growth rates are estimated in the first stage, along with a set of initial estimates of firms' conditional choice probabilities (CCPs) that are based on a multinomial logit choice model. In the second stage, these estimates are used to estimate the remaining parameters (repositioning costs, economies of scope, and the scale of the payoff shocks associated with each format choice) using the dynamic model. This appendix provides full details of these procedures.

C.1. *First Stage: Estimation of the Listener Demand Model and the Evolution of Unobserved Station Quality (ξ)*

The listener demand model is a random coefficients demand model. There is no price variable, but there is a potential endogeneity problem, as unobserved station quality may affect firms' format choices. I avoid this problem by forming quasi-differenced moments based on innovations in station quality (v_{st}^ξ), that are assumed to be unknown to firms when period t format choices are made in period $t - 1$. The model has 37 nonlinear parameters (ρ^ξ , γ^σ , and 35 demographic taste parameters), collectively labeled γ^{NL} , and a set of linear parameters (γ^{L}) that capture format tastes, time effects, and observable differences in station quality. Estimation involves minimizing a GMM objective function based on three sets of moments.

C.1.1. *Quasi-Differenced Moments*

The quasi-differenced moments are formed from the mean utility equations for listener mean utilities and the AR(1) process that determines the evolution

of the component of station quality that is not associated with observed station characteristics, ξ . For stations that do not change format,

$$(A-11) \quad v_{st}^{\xi} = \delta_{st}(q, \gamma^{\text{NL}}) - \rho^{\xi} \delta_{st-1}(q, \gamma^{\text{NL}}) - (1 - \rho^{\xi}) X_{st} \gamma^{\text{L}} - (1 - \rho^{\xi}) F_{st} \overline{\gamma^{\text{F}}},$$

where the mean utility δ_{st} is uniquely defined by observed market shares q and the nonlinear taste parameters (Berry (1994), Berry, Levinsohn, and Pakes (1995)). The X variables include station characteristics and format * AM interactions and, per the discussion in Section 3, time dummies. The assumption that the quality innovations v_{st}^{ξ} are unknown when format choices are made implies that v_{st}^{ξ} will be uncorrelated with X . The moments are formed as

$$(A-12) \quad E(Z' v^{\xi}(\rho^{\xi}, \gamma^{\text{NL}}, \gamma^{\text{L}})) = 0,$$

where the instruments Z include X_{st} , F_{st} , and the log of the station's market share in the initial period of the data, which should be correlated with δ_{st-1} . As in Nevo (2000, 2001), given estimates of the nonlinear parameters, the linear parameters γ^{L} can be estimated by linear regression where the dependent variable is $\delta_{st}(q, \gamma^{\text{NL}}) - \rho^{\xi} \delta_{st-1}(q, \gamma^{\text{NL}})$. $\sigma_{v^{\xi}}^2$ is estimated using the residuals from this regression.

I assume that the AR(1) process that governs the evolution of quality is the same for stations that change format, apart from a fixed quality change γ^{ξ} . This is potentially controversial, so I choose to estimate the model using only stations that stay in the same format, and then estimate γ^{ξ} using the residuals implied by the estimated coefficients and the mean utilities of stations that do switch formats. I then examine how well the model does at matching the distribution of share changes for switching stations in the data. As shown in the text, the model does very well in this dimension, providing support for my assumption.

C.1.2. Demographic Moments

Petrin (2002) illustrated how the accurate estimation of coefficients for demographic tastes using aggregate market share data can be aided by using demographic-specific moments. I form this type of moment based on the average demographic composition of the audience of different formats reported in Arbitron's annual *Radio Today* reports. Specifically these reports list the average proportion of a format's listeners who are in particular age (12–24, 25–49, 50 plus), gender, and ethnic/racial (white, black, or Hispanic) categories based on a particular set of markets. I specify 35 moments (which match the 35 demographic taste parameters) based on the difference between these reported averages and the averages predicted by my model for the quarters used by Ar-

bitron and the set of markets that are common to my sample and Arbitron's calculations:⁵

$$(A-13) \quad E(\text{prop}_{ftd}^{\text{ARB}} - \text{prop}_{ftd}(\delta(q, \gamma^{\text{NL}}), \gamma^{\text{NL}})) = 0,$$

where prop is the proportion of a format's listeners who are in a particular demographic group.

C.1.3. *One Additional Moment*

The quasi-differenced moment with the lagged share instrument and the demographic moments provide only 36 moments for identifying 37 nonlinear parameters. Intuitively, the parameter that lacks an obvious identifying moment is γ^σ , which determines how much substitution takes place between radio listening and the outside good when the number or quality of stations changes. For example, a high value of γ^σ implies that, all else equal, listening will increase slowly as the number of stations increases. To provide an additional moment, I assume that the expected value of ξ_{st} , which could also affect how audiences increase with the number of stations in a market, is independent of market size, measured by log population.⁶ This is similar in spirit to Berry and Waldfogel's (1999) use of population as an instrument to identify the nesting parameter in a nested logit model of station listenership.⁷

C.1.4. *Estimation Algorithm*

Berry, Levinsohn, and Pakes (1995) and Nevo (2000) outlined a simulation-based estimator for random coefficients demand models. I follow their algorithm, adding the additional moments outlined above. The algorithm involves solving for values of δ for each guess of the nonlinear parameters, using analytic formulas for the gradients of the objective function. The tolerance on this contraction mapping is set equal to $1e-12$. Predicted shares for given nonlinear parameters are calculated using 25 Halton draws of v^R for each of the 18

⁵Arbitron uses different markets for its age/gender and ethnic calculations. There are some markets included in Arbitron's calculations which are not in my sample. The demographic taste coefficients remain similar if I include all of the markets used by Arbitron in the demand estimation. Creating the moments requires aggregating some of the formats used in Arbitron's reports, which is done by weighting these formats by their average listenership.

⁶Specifically, I assume that the vector of ξ_{st} for stations that are based inside the market should be independent of market size. The assumption would likely not hold for stations located outside of the market, as their signals are likely to cover less of the market in larger markets. The signal coverage of these stations is not observed in the data, so this difference would not be controlled for by the included X_{st} variables.

⁷One might object to this moment on the basis that, in larger markets, where fixed costs can be spread across more listeners, investment in quality is likely to be larger. However, if this objection was correct, audiences would increase with market size (correlated with the number of stations), and I would likely underestimate γ^σ . In practice, the estimated value of γ^σ is very high, implying that there is little substitution with the outside good.

demographic groups. The shares for each of the 450 simulated individuals are then weighted using the frequency of each demographic group in the population in the market of interest to calculate the predicted market share (results using more draws are almost identical).

The Berry/Nevo algorithm has been criticized based on examples where it fails to find the minimum of the objective function (Dube, Fox, and Su (2012)). However, a feature of my model is that it is exactly identified, so I know that the minimized value of the objective function should be equal to zero (up to numerical tolerance). At the parameter estimates, the value of the objective function is $2.90e-12$, and the algorithm converges to the same estimates from a number of different starting values.

C.2. First Stage: Estimation of the Revenue Model

The revenue model assumes that station s 's revenues for a listener with demographics D_d are

$$(A-14) \quad r_{st}(Y_s, D_d, \gamma) = \gamma_{my}(1 + Y_{st}\gamma^Y)(1 + D_d\gamma^D),$$

where γ_{my} are market-year effects. However, only annual station revenues are reported in the data, so, for estimation, I assume that the mean annual revenues per listener, derived from BIAfn's revenue and share estimates, are

$$(A-15) \quad \overline{r_{sy}^{BIA}} = \frac{\sum_{t \in y} \sum_{\forall d} r_{st}(Y_s, D_d, \gamma) \widehat{l}_{sdt}(\delta, \widehat{\gamma}^{NL}, \widehat{\gamma}^L)}{\sum_{t \in y} \sum_{\forall d} \widehat{l}_{sdt}(\delta, \widehat{\gamma}^{NL}, \widehat{\gamma}^L)} + \varepsilon_{sy}^R,$$

where $\widehat{l}_{sdt}(\delta, \widehat{\gamma}^{NL}, \widehat{\gamma}^L)$ is the estimated listener demand model's prediction of s 's audience in demographic group d in period t . The residual ε_{sy}^R is assumed to be uncorrelated with station characteristics, local tastes, or format choices, as if, for example, it is random measurement error in BIAfn's revenue formula. The model is estimated using nonlinear least squares, and the standard errors are corrected, by expressing the first-order conditions as moments, for uncertainty in the estimated demand parameters.

C.3. First Stage: Estimation of Demographic Transition Process

The population of ethnic group e in market m is assumed to evolve according to the process

$$(A-16) \quad \begin{aligned} & \log(\text{pop}_{met}) - \log(\text{pop}_{met-1}) \\ &= \tau_0 + \tau_1(\log(\text{pop}_{met-1}) - \log(\text{pop}_{met-2})) + u_{met}, \end{aligned}$$

where pop is the level of population. This model cannot be estimated directly, as the *County Population Estimates* are annual (July each year), so they are only observed every other period. However, adding the equations for t and $t - 1$ and substituting for $(\log(\text{pop}_{\text{met}-1}) - \log(\text{pop}_{\text{met}-3}))$ in the resulting equation yields

$$(A-17) \quad \log(\text{pop}_{\text{met}}) - \log(\text{pop}_{\text{met}-2}) \\ = 2\tau_0(1 + \tau_1) + \tau_1^2(\log(\text{pop}_{\text{met}-2}) - \log(\text{pop}_{\text{met}-4})) + \tilde{u}_{\text{met}},$$

where $\tilde{u}_{\text{met}} = u_{\text{met}} + (1 + \tau_1)u_{\text{met}-1} + \tau_1 u_{\text{met}-2}$. The population numbers in this equation are observed, but $(\log(\text{pop}_{\text{met}-2}) - \log(\text{pop}_{\text{met}-4}))$ will be correlated with \tilde{u}_{met} . I estimate (A-17) by 2SLS using $(\log(\text{pop}_{\text{met}-4}) - \log(\text{pop}_{\text{met}-6}))$ as an instrument for $(\log(\text{pop}_{\text{met}-2}) - \log(\text{pop}_{\text{met}-4}))$.⁸ I estimate this equation using data on the black, white, and Hispanic populations in all radio markets (not just the 102 markets in the sample) from 1996 to 2006, where the particular ethnic/racial group makes up at least 10% of the market population.⁹ The estimates are $\hat{\tau}_0 = 0.00014$ (0.000039) and $\hat{\tau}_1 = 0.96968$ (0.00335). The standard deviation of the innovations u_{met} is 0.0027.

C.4. First Stage: Estimation of Firm CCPs

I calculate initial estimates of firms' choice probabilities by estimating a multinomial logit model, where, as in the true model, the choices for each firm are to keep its stations in the same format or to move one of them to a new format. In an ideal world, these CCPs would be estimated nonparametrically, but this is not possible given the size of the state space, the large number of choices that each firm has, and the size of the observed sample. The small number of observations where a firm moves more than one station are not included when calculating the likelihood. The following explanatory variables are included in the logit model for each choice:

- a dummy for whether the choice involves a station moving to an active format, and interactions with a measure of market revenues per share point (to capture market size effects), and for station being moved: the current period revenue, $\exp(\xi_{st})$, the exponent of the fixed quality component (e.g., based on signal coverage), the interaction of these two exponentiated qualities, and an interaction of $\exp(\xi_{st})$ with a dummy for whether the station is an AM station
- a dummy for whether the choice involves a station moving from an active format to Dark, and interactions with the prior revenue of the station being moved and the measure of market revenues per share point

⁸The instrument will be correlated with the endogenous variable if $\tau_1 \neq 0$ (serial correlation in population growth rates) and it should be uncorrelated with \tilde{u}_{met} if the innovations in growth rates are independent.

⁹Including observations on smaller population groups leads to more volatile growth rates, which can create some implausible population changes when applied to larger populations.

- a dummy for whether the moving station has made a format switch in the previous period and an interaction with market revenues per share point
- a dummy for whether the moving station is coming from the inactive dark format, and an interaction with market revenues per share point
- a count of how many of the owner's stations will be located in a format with another station with the same owner and the interaction of this variable with market revenues per point;
 - interactions of these five sets of variables with the total number of stations that the firm owns and the current period revenues of the firm
- a dummy for the format the station might be moved from, an interaction with a dummy for whether the station is an AM station, and interactions with several demographic measures (proportions black, Hispanic, 12–24 50+, and the growth rates of the black and Hispanic populations relative to the white population)
- a dummy for the format the station might be moved to, an interaction with a dummy for whether the station is an AM station, and interactions with several demographic measures (proportions black, Hispanic, 12–24 50+, and the growth rates of the black and Hispanic populations relative to the white population)
- measures of the intensity of competition that the firm faces both in the format that it is moving the station from and in the format which it might move the station to, specifically: the number of competitor stations, the number owned by multi-station firms, and the sum of their $\exp(\xi)$, fixed quality, and $\overline{revenue}$ (see Appendix B) measures.

In practice, estimation with the observed data requires dropping a small number of format dummies to avoid almost perfect multicollinearity. With 160 variables, the model is flexible but still quite coarse given the richness of the state space. Table A-II gives the coefficient estimates from this model. Relative to a baseline model where the only dummies are for a switch to active format, a switch to Dark, and a switch from Dark, the pseudo- R^2 of the estimated model is 0.142.

C.5. Second Stage: Estimation of the Dynamic Model Using Value Function Approximation

As explained in the text, I consider a number of different estimators of the dynamic model. In this appendix, I detail the estimators that use parametric approximations to the value function. Appendix D details the estimators that use forward simulation to approximate the value function. I begin the discussion by specifying features that are common to all of the parametric approximation estimators.

Selection of States. While I can only estimate the model using firms' observed choices, I am not limited to using only observed states when I approximate the value function. I therefore use the 6,061 observed states from the 612 observed market-quarters where I observe firms' choices for the next period, and

TABLE A-II
 FIRST STAGE MULTINOMIAL LOGIT MODEL OF FORMAT CHOICE
 (STANDARD ERRORS IN PARENTHESES)

(a) Coefficients on Switch Characteristics and Interactions

| | | * Firm's
Revenue | * Number of
Stations Owned By Firm |
|--|--------------------|---------------------|---------------------------------------|
| Switch to an Active Format | -2.000
(4.788) | 0.021
(0.034) | 0.132
(0.043) |
| * Revenue of Moving Station | 0.134
(0.157) | 0.012
(0.009) | -0.105
(0.043) |
| * Market Revenue per Point | -0.643
(2.409) | -0.170
(0.143) | 1.162
(0.802) |
| Moving Station Also Moved Last Period | -0.134
(0.708) | 0.384
(0.216) | -0.414
(0.274) |
| * Market Revenue per Point | 9.061
(5.479) | -0.879
(0.812) | -3.446
(3.959) |
| Switch From Dark Format | -4.628
(8.436) | -1.518
(1.077) | 1.634
(0.730) |
| * Market Revenue per Point | 6.256
(9.821) | 8.740
(6.073) | -8.090
(5.173) |
| Switch to Dark Format | -0.051
(8.366) | -0.633
(1.359) | -0.240
(0.707) |
| * Revenue of Moving Station | -4.921
(4.468) | -0.244
(3.054) | 0.771
(2.168) |
| * Market Revenue per Point | -9.608
(10.131) | -2.421
(10.427) | 8.865
(6.901) |
| Count of Stations in the Same Format
After Switch | 0.471
(0.366) | -0.062
(0.037) | -0.117
(0.073) |
| * Market Revenue per Point | 0.083
(2.938) | -0.044
(0.111) | 1.353
(0.883) |

(Continues)

then create 9 duplicates of each of these market-quarters ($N = 60,610$ states in total) where station formats, unobserved qualities, and market demographics can take on different values from those that are observed. Ownership and observed station characteristics are held fixed, as this is assumed in the model. In each duplication, the unobserved quality of each station is chosen as a uniform random draw on $[-2, 2]$, a range which comprises almost all of the values implied by the estimated demand model. The probability that a station's format is the same as in the data is 0.3, and with probability 0.7 it receives a new format draw. If this happens, the probability of Dark is set equal to 0.05 and the probabilities of each of the other formats are set to be equal to each other. Demographics are altered by varying the size of the white, Hispanic, and black populations by i.i.d. draws uniformly drawn from $[-20\%, 20\%]$ (the size of each age-gender group within the ethnic/racial group changes by the same

TABLE A-II—Continued

(b) Format Coefficients and Demographic Interactions: For Format Station Would Move From

| | Dummy | * AM
Station | * Proportion
Black | * Proportion
Hispanic | * Proportion
12–24 | * Proportion
50+ | * Black Growth/
White Growth | * Hispanic Growth/
White Growth |
|-------------------|-------------------|-------------------|-----------------------|--------------------------|-----------------------|---------------------|---------------------------------|------------------------------------|
| AC/CHR | – | –0.340
(1.160) | 3.812
(3.019) | 2.665
(2.788) | –8.782
(18.396) | –10.899
(12.949) | –0.031
(0.014) | 0.010
(0.005) |
| Country | –4.798
(5.014) | 1.744
(0.647) | 5.926
(3.253) | 2.779
(2.794) | 9.073
(18.456) | –10.711
(13.584) | –0.027
(0.015) | 0.010
(0.005) |
| Rock | –1.200
(4.929) | 4.294
(1.106) | 5.809
(3.168) | 3.252
(2.900) | –2.766
(19.412) | –13.019
(13.440) | –0.023
(0.014) | 0.008
(0.005) |
| Urban | –7.568
(5.841) | 0.361
(0.632) | –3.877
(3.761) | –0.316
(3.382) | 11.583
(19.817) | 1.666
(14.028) | 0.038
(0.038) | –0.013
(0.013) |
| News | –5.397
(5.367) | –0.412
(0.692) | 0.972
(3.728) | 3.136
(3.060) | 6.151
(19.239) | –3.611
(13.718) | –0.029
(0.015) | 0.010
(0.005) |
| Other Programming | –1.527
(4.086) | 1.021
(0.582) | 2.798
(2.921) | 0.673
(2.640) | 5.376
(17.344) | –14.540
(12.615) | –0.027
(0.014) | 0.009
(0.005) |
| Spanish | –0.621
(7.748) | 1.489
(0.831) | –1.431
(5.477) | –6.034
(3.775) | –4.229
(24.437) | –9.759
(15.085) | –0.052
(0.021) | 0.000
(0.009) |

(Continues)

TABLE A-II—Continued

(c) Format Coefficients and Demographic Interactions: For Format Station Would Move to

| | Dummy | * AM
Station | * Proportion
Black | * Proportion
Hispanic | * Proportion
12–24 | * Proportion
50+ | * Black Growth/
White Growth | * Hispanic Growth/
White Growth |
|-------------------|-------------------|-------------------|-----------------------|--------------------------|-----------------------|---------------------|---------------------------------|------------------------------------|
| AC/CHR | – | –4.359
(1.135) | –5.344
(3.166) | –0.477
(2.916) | 0.661
(18.023) | 10.540
(13.108) | 0.022
(0.016) | –0.007
(0.006) |
| Country | –3.543
(5.432) | –1.192
(0.655) | –0.826
(3.294) | 0.089
(3.049) | 0.982
(19.666) | 15.452
(13.584) | 0.028
(0.015) | –0.009
(0.005) |
| Rock | 5.746
(5.433) | –1.357
(0.666) | 9.824
(3.457) | 2.659
(3.169) | –27.486
(20.575) | 2.117
(13.499) | 0.026
(0.019) | –0.008
(0.007) |
| Urban | 1.019
(4.466) | 1.696
(0.574) | –2.092
(3.031) | –1.223
(2.705) | –4.628
(18.023) | 5.141
(12.714) | 0.011
(0.015) | –0.001
(0.005) |
| News | –2.131
(4.750) | –0.363
(0.603) | –3.853
(3.167) | –4.094
(3.142) | 6.715
(18.262) | 10.327
(13.201) | 0.030
(0.014) | –0.009
(0.005) |
| Other Programming | 0.519
(5.073) | –1.600
(0.668) | –4.248
(3.751) | 8.348
(2.935) | –3.824
(19.043) | 4.639
(13.558) | 0.030
(0.014) | –0.009
(0.005) |
| Spanish | 1.886
(4.382) | –3.487
(0.886) | –8.583
(3.192) | –4.041
(3.072) | 0.369
(17.905) | 6.097
(13.038) | 0.024
(0.015) | –0.008
(0.005) |

(Continues)

TABLE A-II—*Continued*

(d) Characteristics of Moving Station and Measures of Competition

| Characteristics of Station Being Moved | | Measures of Competition | Format Station | |
|---|-------------------|--|-------------------|-------------------|
| | | | Moving From | Moving to |
| exp(ξ) | −0.187
(0.100) | Number of Other Stations in Format | −0.098
(0.069) | 0.028
(0.066) |
| exp(fixed quality component) | −0.090
(0.035) | Number of Other Stations in Format That
Have the Same Owner | 0.097
(0.054) | −0.064
(0.048) |
| exp(ξ) * exp(fixed quality component) | −0.051
(0.034) | Sum of exp(ξ) of Other Stations | 0.048
(0.032) | −0.006
(0.031) |
| exp(ξ) * AM dummy | −0.074
(0.072) | Sum of exp(fixed quality component) | 0.026
(0.009) | −0.007
(0.009) |
| | | Sum of Mean Revenue Measure of
Other Stations | −0.049
(0.028) | −0.020
(0.027) |
| Log-Likelihood: −2170.5, Observations 6,025 | | | | |

percentage). In the description of the estimation procedure, a particular state j is denoted as $\mathcal{M}_{j,o,t}$, where o indicates the firm of interest in the state and t denotes the initial period.

Variables Used in Approximating the Value Function. I assume that a firm's value function can be approximated by a linear parametric function of functions of the state variables. More variables and interactions are needed than in the counterfactual analysis to account for the fact that the same function holds across markets, which differ greatly in their size and demographic characteristics.

The approximating function for a state $\mathcal{M}_{j,o,t}$ includes the following variables:

- a market-quarter fixed effect (note that this takes the same value for each of the duplicates for the same market-quarter)
- the following measures of market demographics:
 - the proportion of blacks and the proportion of Hispanics
 - these proportions interacted with market population multiplied by the market-year price effect (γ_{my}), estimated from the revenue model, and interacted with the number of stations owned by firm o
- current firm no ξ revenues¹⁰
- the following variables on their own, interacted with the number of stations owned by firm o and the market-year price:
 - sum of $\overline{revenues}$ for firm o 's stations
 - the sum of the $\exp(\xi_{st})$ measures for firm o 's stations
 - the sum of the $\exp(X_{st}\gamma^S)$ measures for firm o 's stations (excluding the AM * format component)
 - the sum of the AM dummies for firm o 's stations
 - the sum of the AM dummies interacted with $\exp(X_{st}\gamma^S)$ (excluding the AM * format component) for firm o 's stations
 - the sum of the AM dummies interacted with $\exp(\xi_{st})$ for firm o 's stations
 - the sum of the $\exp(X_{st}\gamma^S)$ (excluding the AM * format component) interacted with $\exp(\xi_{st})$ for firm o 's stations
 - the sum of $\overline{revenue}$ interacted with $\exp(\xi_{st})$ for firm o 's stations
- for the largest competitor faced by the firm (call it firm x):¹¹
 - the number of stations owned by firm x
 - the sum of the $\overline{revenue}$ measures for firm x 's stations
 - the sum of the $\exp(\xi_{st})$ measures for firm x 's stations
 - the sum of interactions between the $\overline{revenue}$ measures and $\exp(\xi_{st})$ for firm x 's stations
 - the interactions of these four variables with market population multiplied by the market-year price

¹⁰See Appendix B (Counterfactual) for a description of alternative revenue measures.

¹¹The size of competitors is determined by the number of stations owned and, where this is equal, the sum of the $\overline{revenue}$ measures for the different firms.

- for the second largest competitor faced by the firm (call it firm y):
 - the number of stations owned by firm y
 - the sum of the $\overline{revenue}$ measures for firm y 's stations
 - the sum of the $\exp(\xi_{st})$ measures for firm y 's stations
 - the sum of interactions between the $\overline{revenue}$ measures and $\exp(\xi_{st})$ for firm y 's stations
 - the interactions of these four variables with market population multiplied by the market-year price
- for each active format (i.e., the coefficients can vary freely across formats):
 - number of rival stations in the format
 - number of rival stations in the format owned by firms that own more than one station
 - sum of rival stations $\exp(\xi_{st})$ s
 - sum of rival stations $\exp(X_{st}\gamma^S)$ (excluding the AM * format component)
 - sum of the $\overline{revenue}$ measures for rival stations
 - sum of the $\overline{revenue}$ measures for rival stations that are in the AM band
 - the interaction of these six variables with (i) the proportion black in the population, (ii) the proportion Hispanic in the population, (iii) the growth rate of the black population, (iv) the growth rate of the Hispanic population, and (v) the interaction of these 30 variables with market population multiplied by the market-year price (γ_{my}) effect from the estimated revenue equation.

In the following description of the estimation procedure, $\Phi_{j,k}(\mathcal{M}_{j,o,t})$ is the value of the k th approximating variable in state j and Φ is the matrix where these variables are stacked for the N states.

Initial Choice Probabilities. Initial values of the CCPs are required to start the estimation procedure. For the modified procedures, the choice probabilities of *other* firms are held constant at these initial values during estimation. In both cases, I use the first-stage estimates from the multinomial logit choice model. This involves extrapolation using the parametric form of this model for the states that are not observed in the data.

C.5.1. Modified Pseudo-Likelihood and Moment-Based Procedures

I begin by describing the procedures used to produce the estimates in the first two columns of Table VI in the text. In these estimators, the choice probabilities of other firms (P_{-o}) are held fixed at their initial (first-stage multinomial logit) estimates. The logic of these procedures follows Aguirregabiria and Mira (2010) (discussed in Aguirregabiria and Nevo (2012)), although the moment-based version is also inspired by the discussion and Monte Carlo results in Pakes, Berry, and Ostrovsky (2007, POB). One could also implement the procedures using an MPEC-based method, combining value function approximation with the procedure proposed in Egedal, Lai, and Su (2012). In some settings, MPEC-based methods have been shown to have su-

rior numerical properties to the type of iterative procedure used here (Su (2012)).

Estimation is based on an iterative procedure with the following steps in iteration i . In state $\mathcal{M}_{j,o,t}$, $P_o^i(a|\mathcal{M}_{j,o,t})$ is the iteration i guess of the probability that firm o chooses action a , and $P_{-o}(\mathcal{M}_{j,o,t})$ are the (fixed) choice probabilities of o 's competitors in that state.

Step 1. For each of the N states, $\tilde{\pi}(P_o^i(\mathcal{M}_{j,o,t}), \theta^i)$ is calculated as

$$\begin{aligned}
 \text{(A-18)} \quad & \tilde{\pi}(P_o^i(\mathcal{M}_{j,o,t}), \theta^i) \\
 &= \sum_{s \in S^o} R_s(\mathcal{M}_{j,o,t} | \gamma) \\
 &+ \sum_{a \in A_o(\mathcal{M}_{j,o,t})} P_o^i(a|\mathcal{M}_{j,o,t}) (\beta C_o(a) \theta^{C,i} \\
 &- W_o(a) \theta^{W,i} + \theta^{\varepsilon,i} (\varkappa - \log(P_o^i(a|\mathcal{M}_{j,o,t}))))),
 \end{aligned}$$

where θ^i denotes the current guess of the structural parameters, and \varkappa is Euler's constant, $C_o(a)$ is the number of stations that the firm will have operating in the same format as one of its other stations in the next period if it chooses action a , and $A_o(\mathcal{M}_{j,o,t})$ is o 's choice set. $\tilde{\pi}(P_o^i(\mathcal{M}_{j,o,t}), \theta^i)$ is a function only of o 's choice probabilities. β is the discount factor. $\tilde{\pi}(P^i, \theta^i)$ is the vector that stacks these values for the N states.¹²

For each of the N states, the choice probabilities of all firms are used to calculate $E_{P_o^i} \Phi$, a vector that contains the expected value of each of the approximating variables given strategies. For variable k ,

$$\begin{aligned}
 \text{(A-19)} \quad E_{P_o^i} \phi_{j,k} &= \int \phi_{h,k}(\mathcal{M}_{h,o,t+1}) \\
 &\times g(\mathcal{M}_{h,o,t+1} | P_o^i(\mathcal{M}_{j,o,t}), P_{-o}(\mathcal{M}_{j,o,t}), \mathcal{M}_{j,o,t}) d\mathcal{M}_{h,o,t+1},
 \end{aligned}$$

where g is the transition density. This integral is approximated by reweighting variables for a prespecified sample of $\mathcal{M}_{h,o,t+1}$ states, as calculating $\phi_{h,k}(\mathcal{M}_{h,o,t+1})$ requires solving a random coefficients demand model. Specifically, for a given state $\mathcal{M}_{j,o,t}$, I consider a set of H states $\mathcal{M}_{h,o,t+1}$ which is equal to the set of states that can be reached by any move by o , a set of S^ξ draws for innovations in ξ , and $S^{-o,m}$ moves by other firms in the same local

¹²An initial guess of the structural parameters is required. I assume a common switching cost of \$2 m. for all switches between active formats, a cost of \$1 m. for a switch from Dark, a cost of \$4 m. for a switch to Dark, $\theta^C = 0.1$ and $\theta^\varepsilon = 0.5$.

market. The integral is approximated by

$$(A-20) \quad E_{P_o^i} \phi_{j,k}^i = \sum_{h=1}^H \phi_{h,k}(\mathcal{M}_{h,o,t+1}) \\ \times \frac{g(\mathcal{M}_{h,o,t+1} | P_o^i(\mathcal{M}_{j,o,t}), P_{-o}(\mathcal{M}_{j,o,t}), \mathcal{M}_{j,o,t})}{\sum_{h'=1}^H g(\mathcal{M}_{h',o,t+1} | P_o^i(\mathcal{M}_{j,o,t}), P_{-o}(\mathcal{M}_{j,o,t}), \mathcal{M}_{j,o,t})}$$

$S^\xi = 10$ and $S^{-o,m} = 500$. To be accurate, the integration procedure requires the $S^{-o,m}$ moves to include those that are most likely to be made. I choose the ones that are most likely to be made based on the first-stage estimates of the conditional choice probabilities.

Step 2. Create matrices $(\Phi - \beta E\Phi^i)$ and, as the parameters λ are overidentified when $N > K$, use an OLS regression to calculate the coefficients $\widehat{\lambda}^i$:

$$(A-21) \quad \widehat{\lambda}^i(\theta^i, P_o^i) = ((\Phi - \beta E_{P_o^i} \Phi^i)' (\Phi - \beta E_{P_o^i} \Phi^i))^{-1} (\Phi - \beta E_{P_o^i} \Phi^i)' \widetilde{\pi}(P_o^i, \theta^i).$$

Step 3. Use $\widehat{\lambda}^i$ to calculate the future value of each firm when it makes choice a (note that this is not quite the same as the choice-specific value function as defined in the text, as that also includes current revenues and repositioning costs associated with a):

$$(A-22) \quad FV(a, \mathcal{M}_{j,o,t}, P_{-o}(\mathcal{M}_{j,o,t})) \\ = \sum_{h=1}^H \sum_{k=1}^K \phi_{h,k}(\mathcal{M}_{h,o,t+1}) \\ \times \left\{ \frac{g(\mathcal{M}_{h,o,t+1} | a, P_{-o}(\mathcal{M}_{j,o,t}), \mathcal{M}_{j,o,t})}{\sum_{h'=1}^H g(\mathcal{M}_{h',o,t+1} | a, P_{-o}(\mathcal{M}_{j,o,t}), \mathcal{M}_{j,o,t})} \right\} \lambda_k^i(\widehat{\theta}^i, P_o^i).$$

Step 4. Estimate the structural parameters θ' using firms' observed choices. The probability that a is chosen is

$$(A-23) \quad P_o(a | \mathcal{M}_{j,o,t}) \\ = \exp\left(\frac{FV(a, \mathcal{M}_{j,o,t}, P_{-o}(\mathcal{M}_{j,o,t})) - W_o(a)\theta^W + \beta C_o(a)\theta^C}{\theta^\varepsilon}\right) \\ / \sum_{a' \in A_o(\mathcal{M}_{j,o,t})} \exp((FV(a', \mathcal{M}_{j,o,t}, P_{-o}(\mathcal{M}_{j,o,t})) \\ - W_o(a')\theta^W + \beta C_o(a')\theta^C) / \theta^\varepsilon).$$

Current revenues drop out because they are common across choices. The pseudo-likelihood and moment-based estimators differ in how these probabilities are used.

For the *pseudo-likelihood estimator*, the probabilities are used in what is similar to a standard multinomial logit estimation, except that the scale parameter differs across markets. Observations for firms moving more than one station are excluded from the calculation of the pseudo-likelihood. One advantage of this estimator is that the log-likelihood objective function is well-behaved and easy to maximize.

For the *moment-based estimator*, the probabilities are used to match a number of informative format-switching rates in the data, where the rates are formed by averaging across states. Specifically, for the three market size groups (population less than 0.25 m., 0.25 m.–1 m., 1 m.+), the estimator matches: (i) the (average-across-states) probability that a station is moved from one active format to another active format, (ii) the probability that a station is switched to Dark, (iii) the probability that a Dark station is moved to an active format; (iv) the probability that a station that switched formats in the previous period makes a further switch, the probability that a non-Urban station switches to Urban in markets with (v) a below median proportion of blacks and (vi) an above median proportion of blacks, the probability that a non-Spanish station switches to Spanish in markets with (vii) a below median proportion of Hispanics and (viii) an above median proportion of Hispanics, (ix) the probability that a non-News AM station is moved to News, and (x) the probability that a multi-station firm chooses a move that increases the number of stations that it operates in the same format. One more moment is provided for each market group by matching the average revenue of a switching station. Observations for firms moving more than one station are excluded when calculating both the data and predicted moments. The identity matrix is used to weight the moments. One disadvantage of this estimator is that the objective function can have multiple local minima. The estimation routine therefore uses both Nelder–Mead and derivative-based optimization routines from different starting points to search for the global minimum.

Step 5. Use θ' to compute

$$\begin{aligned}
 \text{(A-24)} \quad P'_o(a|\mathcal{M}_{j,o,t}) &= \exp\left(\frac{\text{FV}(a, \mathcal{M}_{j,o,t}, P_{-o}(\mathcal{M}_{j,o,t})) - W_o(a)\theta^W + \beta C_o(a)\theta^C}{\theta'^\varepsilon}\right) \\
 &\quad / \sum_{a' \in A_o(\mathcal{M}_{j,o,t})} \exp((\text{FV}(a', \mathcal{M}_{j,o,t}, P_{-o}(\mathcal{M}_{j,o,t})) \\
 &\quad - W_o(a')\theta^W + \beta C_o(a')\theta^C)/\theta'^\varepsilon).
 \end{aligned}$$

Step 6. If the maximum absolute difference between P'_o and P_o^i is less than $1e-6$ and the maximum absolute difference between θ^i and θ' is also less than

$1e-4$, the procedure stops. Otherwise, P_o is updated as a weighted combination of P_o^i and P' :

$$P_o^{i+1} = \psi P_o' + (1 - \psi) P_o^i,$$

and θ is updated as

$$\theta_o^{i+1} = \psi \theta' + (1 - \psi) \theta^i,$$

where $\psi = 0.1$, and the procedure returns to Step 1 for iteration $i + 1$.

C.5.2. Iterated Pseudo-Likelihood Procedure

The iterated pseudo-likelihood procedure follows [Aguirregabiria and Mira \(2010\)](#) in that the choice probabilities of other firms are also updated during estimation. Specifically, in the description set out above, P_{-o} should be replaced by P_{-o}^i , and in Step 6 the choice probabilities of all players are also updated.

APPENDIX D: ESTIMATION OF THE DYNAMIC MODEL USING FORWARD SIMULATION

An alternative approach to estimating dynamic games, and approximating value functions, involves the use of forward simulation, an approach most closely associated with [Bajari, Benkard, and Levin \(2007, BBL\)](#). I implement two estimators that use forward simulation and inequalities: one based on the objective function proposed by BBL, and one that is based on moment inequalities following [Pakes, Porter, Ho, and Ishii \(2011, PPHI\)](#), the difference being that the latter averages across states when forming the inequalities.

As explained in the text, it can be difficult to estimate a large number of parameters using these methods, so I only consider a model with three parameters (a cost of switching to an active format, θ^W , the economy of scope from operating multiple stations θ^C in the same format, and the scale parameter of the i.i.d. payoff shocks to each format choice, θ^ϵ) and estimate the model separately for each of the three market size groups.¹³ I begin by describing the BBL estimator, and then explain the changes made to implement the PPHI estimator.

BBL. The BBL estimator uses the equilibrium assumption that, given the strategies of other firms, a firm's observed policy should give it higher expected

¹³In simplifying the model, I assume that the cost of moving from Dark to an active format is the same as moving between a pair of active formats, and that there is no cost to moving to Dark. I have estimated specifications with separate coefficients for these costs, but without the imposition of arbitrary constraints, found that the estimates produced were often completely implausible (e.g., a cost of \$100 million for switching to Dark).

payoffs than any alternative policy. Given the linear form of the payoff function, a firm's value when it uses strategy Γ_o and other firms use strategies Γ_{-o}^* can be expressed as

$$(A-25) \quad V(\mathcal{M}_{j,o,t}|\Gamma_o, \Gamma_{-o}^*, \theta) \\ = \mathbf{V}_{o,\Gamma_o,\Gamma_{-o}^*} \theta = \mathbf{R}_{o,\Gamma_o,\Gamma_{-o}^*} - \theta^W \mathbf{W}_{o,\Gamma_o,\Gamma_{-o}^*} + \theta^C \mathbf{C}_{o,\Gamma_o,\Gamma_{-o}^*} + \theta^\varepsilon \boldsymbol{\varepsilon}_{o,\Gamma_o,\Gamma_{-o}^*}^F,$$

where $\mathbf{R}_{o,\Gamma_o,\Gamma_{-o}^*} = E_{o,\Gamma_o,\Gamma_{-o}^*} \sum_{t'=0}^{\infty} \beta^{t'} \sum_{s \in S^o} R_s(\mathcal{M}_{o,t+t'}|\gamma)$,

$$(A-26) \quad \mathbf{W}_{o,\Gamma_o,\Gamma_{-o}^*} = E_{o,\Gamma_o,\Gamma_{-o}^*} \sum_{t'=0}^{\infty} \beta^{t'} \sum_{s \in S^o} I(f_{st+t'} \neq f_{st+t'+1}, f_{st+t'+1} \neq \text{DARK}), \\ \mathbf{C}_{o,\Gamma_o,\Gamma_{-o}^*} = E_{o,\Gamma_o,\Gamma_{-o}^*} \sum_{t'=1}^{\infty} \beta^{t'} C_o(\mathcal{M}_{o,t+t'}), \\ \boldsymbol{\varepsilon}_{o,\Gamma_o,\Gamma_{-o}^*}^F = E_{o,\Gamma_o,\Gamma_{-o}^*} \sum_{t'=0}^{\infty} \beta^{t'} \boldsymbol{\varepsilon}_{ot+t'}^F(a_{ot}),$$

where I use $\mathcal{M}_{o,t+t'}$ to denote whatever state firm o is in period $t+t'$. The necessary equilibrium condition is that

$$(A-27) \quad V(\mathcal{M}_{j,o,t}|\Gamma_o^*, \Gamma_{-o}^*, \theta) - V(\mathcal{M}_{j,o,t}|\Gamma_o^a, \Gamma_{-o}^*, \theta) \geq 0 \quad \forall \Gamma_o^a, \mathcal{M}_{j,o,t},$$

where Γ_o^* is firm o 's observed equilibrium strategy and Γ_o^a are alternative strategies.

$\mathbf{R}_{o,\Gamma_o^*,\Gamma_{-o}^*}$, $\mathbf{W}_{o,\Gamma_o^*,\Gamma_{-o}^*}$, $\mathbf{C}_{o,\Gamma_o^*,\Gamma_{-o}^*}$, and $\boldsymbol{\varepsilon}_{o,\Gamma_o^*,\Gamma_{-o}^*}^F$ can be approximated using forward simulation of the model and initial estimates of firms' conditional choice probabilities. $\mathbf{R}_{o,\Gamma_o^a,\Gamma_{-o}^*}$ (and the equivalent for the other terms) can be approximated by using a different set of choice probabilities. BBL proposed finding the parameters that make these inequalities hold in the data for a finite set of alternatives using an objective function

$$\widehat{\theta}^{\text{BBL}} = \arg \min_{\theta} \sum_o \sum_{\forall a} \max\{(\mathbf{V}_{o,\Gamma_o^a,\Gamma_{-o}^*} - \mathbf{V}_{o,\Gamma_o^*,\Gamma_{-o}^*})\theta, 0\}^2,$$

where the estimates will be a set if there are parameters that satisfy all of the inequalities. The estimator has a manageable computational burden because it is not necessary to recalculate \mathbf{R} , \mathbf{W} , \mathbf{C} , and $\boldsymbol{\varepsilon}$ as the parameters change. It is straightforward to add the additional parameter restriction that $\theta^\varepsilon \geq 0$ (scale of the payoff shocks must be nonnegative).

The iterative forward simulation procedure is straightforward. Suppose that we want to simulate the values of \mathbf{R} , \mathbf{W} , \mathbf{C} , and $\boldsymbol{\varepsilon}$ for a particular firm o using observed policies Γ_o^* . For a given simulation sim , we start from an initial state,

setting $\mathbf{R}_{sim,o,\Gamma_o^*,\Gamma_{-o}^*}$, $\mathbf{W}_{sim,o,\Gamma_o^*,\Gamma_{-o}^*}$, $\mathbf{C}_{sim,o,\Gamma_o^*,\Gamma_{-o}^*}$, and $\boldsymbol{\epsilon}_{sim,o,\Gamma_o^*,\Gamma_{-o}^*}$ to zero, and then iterate the following steps:

Step 1. Given the state, calculate o 's revenues by solving the random coefficients model of listener demand and then using the estimated revenue model to calculate the total revenues each station receives. For o , increase the value of $\mathbf{R}_{sim,o,\Gamma_o^*,\Gamma_{-o}^*}$ and $\mathbf{C}_{sim,o,\Gamma_o^*,\Gamma_{-o}^*}$ based on its revenues and the format configuration of its stations.

Step 2. Given the state, form a matrix that contains the same explanatory variables used for the first-stage multinomial logit estimates of conditional choice probabilities. Then use the estimated coefficients of this model to calculate the CCPs for all firms in the market.

Step 3. Using these CCPs, choose an action for each firm and update all station formats. For o , update $\mathbf{W}_{sim,o,\Gamma_o^*,\Gamma_{-o}^*}$ if it changes the format of one of its stations. Update $\boldsymbol{\epsilon}_{sim,o,\Gamma_o^*,\Gamma_{-o}^*}$ by adding $(\varkappa - \log(P_o^*(a)))$, where \varkappa is Euler's constant and a is the action that firm o chooses.

Step 4. Use the estimated transition processes for ξ and market demographics (specifically, the growth rates for each ethnic/racial group) to update these variables. I impose some constraints on the level of the ξ 's and growth rates to keep them within the approximate ranges that I observe in the data.

Step 5. Repeat Steps 1–5 for the next period, and continue until the model has been simulated forward 50 periods.

Given that a market can evolve in many ways, it is necessary to average across many simulations so as to reduce simulation error, although this increases the computational burden, especially as it is necessary to solve a random coefficients demand model in each period for each simulation and policy. I use 500 simulations and construct inequalities based on all of the observed states in the data. I experimented using 2,000 simulations for small markets, where the BBL estimates of θ^W and θ^ε are larger than the parametric approximation estimates and lie outside the PPHI bounds that I describe below. The BBL estimates were slightly further away from the other estimates and the PPHI bounds in this case, while the PPHI bounds were similar, suggesting that the number of simulations does not explain the results. However, one advantage of the PPHI formulation discussed below is that it may reduce the effects of simulation error in the estimates of the components of the value function.

Any deviation from Γ_o^* provides a possible alternative policy that can be used for estimation. My experience from estimating this model is that the choice of alternative policies can significantly affect the results, especially with the BBL objective function.¹⁴ Out of the alternatives that I tried (and I tried many), the ones described below provided the BBL and PPHI estimates that were most similar to each other and which seemed intuitively plausible. They also appeared to be among the most robust to varying the set of states used in

¹⁴Srisuma (2010) discussed an example where a commonly used type of alternative policy, which involves simply adding noise to the choice probabilities, cannot identify the parameters.

calculating the objective function. There is also an intuitive reason why each of the alternatives used should help to identify the parameters. In forming them, I use information on what the model predicts about a firm's revenues in the following period for each of its possible actions, assuming that the formats of other stations are held fixed and the unobserved qualities of all stations are set equal to 0.¹⁵ It is these revenues to which I am referring when I talk about "next period revenues" in the following descriptions.

Alternative Policy 1: More format switching: if a format choice, involving a switch, gives a higher next period revenue than maintaining the same format configuration but, under the estimated actual policies (i.e., estimated CCPs), it would be chosen with lower probability,¹⁶ make the probabilities of making this switch and maintaining the same configuration equal to each other. Intuitively, this alternative policy should tend to increase a firm's expected future revenues, but also increase the amount of switching that it does, and the fact that this policy is not optimal should identify a lower bound on repositioning costs.

Alternative Policy 2: Less format switching: reduce the probability that a firm makes each choice involving moving a station to another active format by 90%, increasing the probability that the firm chooses to maintain its existing formats. Assuming that moves that increase revenues are more likely to be chosen, this change will reduce switching and expected future revenues, and the fact that this policy is not optimal should identify an upper bound on repositioning costs.

Alternative Policy 3: Higher probability of making format choices that increase clustering of stations: identify format choices that would increase the number of stations in the same active format relative to the choice of no move (call these the "increase options"), and those that would reduce it (the "reduce options"). Reduce the probability of choosing each of the reduce options by two-thirds, and proportionally increase the probability of choosing each of the increase options. As clustering of stations will result in cannibalization, it will tend to reduce expected future revenues. Intuitively, the fact that this policy is not optimal should identify an upper bound on the value of economies of scope.

Alternative Policy 4: Lower probability of making format choices that will increase station clustering: this is simply the reverse of Alternative Policy 3 (i.e., the probability of the reduce options is increased, and the probability of increase options is reduced). As spreading out stations will tend to increase expected future revenues because it reduces cannibalization, intuitively the fact

¹⁵Results are similar if instead the ξ 's are assumed to stay fixed at their current values, but experimentation indicated that using the $\xi = 0$ revenues gives a slightly better prediction about which switches will increase a firm's revenues in the long run, consistent with the fact that the ξ 's are transitory.

¹⁶If multiple moves would produce higher expected revenues, I use the one that has the highest expected revenue.

that this policy is not optimal should identify a lower bound on the value of economies of scope.

Alternative Policy 5: More random switching: identify all of the format choices that will raise a firm's next period revenues relative to keeping its current format configuration, and set the probability of all of these choices equal to each other. Conditional on one of these choices being made, setting the probabilities to be equal to each other maximizes the expected value of the payoff shock associated with the choice. Intuitively, this alternative policy should also reduce expected future revenues (because the probability of the best of these options will have fallen) or reduce economies of scope, so the fact that this policy is not optimal should identify an upper bound on θ^ε (the scale of the payoff shocks).

Alternative Policy 6: Less random switching: identify all of the format choices that will raise the firm's next period revenues relative to keeping its current format configuration and set the probability of all of these choices except the one that maximizes next period revenues equal to zero, attributing these probabilities to the choice that does maximize next period revenues. Intuitively, this alteration should reduce the expected value of ε 's while increasing expected future revenues, and the fact that this policy is not used should identify a lower bound on θ^ε .

For each of these alternative policies, the forward simulation calculations are repeated. Other firms continue to use the conditional choice probabilities implied by Γ_{-o}^* , and the draws of demographics and innovations in ξ are the same as in the simulations for Γ_o^* . Applied to my data, the BBL estimator always produces point estimates because there are no parameters that satisfy the inequalities for all states.

Moment Inequalities (PPHI). The BBL estimator uses the fact that the inequality (A-27) should hold for any state and any alternative policy, but, in practice, the simulated inequalities do not hold at the estimated parameters for a significant number of states and alternative policies, possibly because of inaccurate estimates of the first-stage conditional choice probabilities or simulation error in \mathbf{R} , \mathbf{W} , \mathbf{C} , and ε .

An alternative estimation approach makes the weaker assumption that the same set of inequalities should hold when an average is taken across O states (in practice, all states in the observed data for a given market size group), producing an estimating moment inequality of the form

$$\frac{1}{O} \sum (\mathbf{V}_{o,\Gamma_o^*,\Gamma_{-o}^*} - \mathbf{V}_{o,\Gamma_o^a,\Gamma_{-o}^*}) \theta \geq 0 \quad \text{for an alternative policy } \Gamma^a.$$

Arguments for why averaging might be helpful in the presence of first-stage bias or simulation error are analogous to the arguments presented in Pakes et al. (2007) for why a moment-based estimator using switching rates may be

more reliable.¹⁷ For each of the market size groups, I construct one of these linear moment inequalities for each of the alternative policies considered above (using exactly the same simulations used for the BBL estimator), and find the set of parameters that satisfy all of these inequalities. In my data, and for these alternative policies, this approach always generates a set.

REFERENCES

- AGUIRREGABIRIA, V., AND P. MIRA (2010): “Stability of Equilibria and Properties of Sequential Estimators: Problems and Solutions,” Report, University of Toronto. [19,23]
- AGUIRREGABIRIA, V., AND A. NEVO (2012): “Recent Developments in Empirical IO: Dynamic Demand and Dynamic Games,” Report, University of Toronto. [19]
- ARBITRON COMPANY (various years): *Radio Today*. New York: Arbitron Company.
- BAJARI, P., L. BENKARD, AND J. LEVIN (2007): “Estimating Dynamic Models of Imperfect Competition,” *Econometrica*, 75 (5), 1331–1370. [23]
- BERRY, S. T. (1994): “Estimating Discrete-Choice Models of Product Differentiation,” *RAND Journal of Economics*, 25 (2), 242–262. [9]
- BERRY, S. T., AND J. WALDFOGEL (1999): “Free Entry and Social Inefficiency in Radio Broadcasting,” *RAND Journal of Economics*, 30 (2), 397–420. [10]
- BERRY, S. T., J. LEVINSOHN, AND A. PAKES (1995): “Automobile Prices in Market Equilibrium,” *Econometrica*, 63 (4), 841–890. [9,10]
- DUBE, J.-P. H., J. T. FOX, AND C.-L. SU (2012): “Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients,” *Econometrica*, 80 (5), 2231–2267. [11]
- EGESDAL, M., Z. LAI, AND C.-L. SU (2012): “Estimating Dynamic Discrete-Choice Games of Incomplete Information,” Report, Harvard University. [19]
- NEVO, A. (2000): “A Practitioner’s Guide to Estimation of Random-Coefficients Logit Models of Demand,” *Journal of Economics and Management Strategy*, 9 (4), 513–548. [9,10]
- (2001): “Measuring Market Power in the Ready-to-Eat Cereal Industry,” *Econometrica*, 69 (2), 307–342. [9]
- PAKES, A., S. T. BERRY, AND M. OSTROVSKY (2007): “Simple Estimators for the Parameters of Discrete Dynamic Games (With Entry/Exit Examples),” *RAND Journal of Economics*, 38 (2), 373–399. [19,27]
- PAKES, A., J. PORTER, K. HO, AND J. ISHII (2011): “Moment Inequalities and Their Application,” Report, Harvard University. [23]
- PETRIN, A. (2002): “Quantifying the Benefits of New Products: The Case of the Minivan,” *Journal of Political Economy*, 110 (4), 705–729. [9]
- SRISUMA, S. (2010): “Estimation of Structural Optimization Models: A Note on Identification,” Report, London School of Economics. [25]
- SU, C.-L. (2012): “Estimating Discrete-Choice Games of Incomplete Information: A Simple Static Example,” Report, University of Chicago. [20]

¹⁷Of course, the disadvantage of averaging is that information about changes in payoffs in individual states when alternative policies are used is lost. If the only effect of averaging across states was to lose information, then we would expect the BBL estimates to be within the (possibly wide) sets generated by the PPHI estimator, but, for at least two of the three market size groups, this is not the case, suggesting that averaging across states has some other effect, such as reducing problems created by using inaccurate estimates of the choice probabilities or forward simulation error.

SWEETING, A. T. (2011): “Dynamic Product Positioning in Differentiated Product Markets: The Effect of Fees for Musical Performance Rights on the Commercial Radio Industry,” Report, Duke University. [3]

*Dept. of Economics, University of Maryland, College Park, MD 20742, U.S.A.,
Duke University, and NBER; sweeting@econ.umd.edu.*

Manuscript received October, 2007; final revision received July, 2012.