10-1-2015

# Optimal Product Variety in Radio Markets

Steven T. Berry

Alon Eizenberg

Joel Waldfogel

# OPTIMAL PRODUCT VARIETY IN RADIO MARKETS

**By**

**Steven Berry, Alon Eizenberg, and Joel Waldfogel**

**September 2015**

**COWLES FOUNDATION DISCUSSION PAPER NO. 2023**

**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS**
**YALE UNIVERSITY**
**Box 208281**
**New Haven, Connecticut 06520-8281**

**http://cowles.yale.edu/**

# Optimal Product Variety in Radio Markets*

Steven Berry† Alon Eizenberg‡ Joel Waldfogel§

September 2015

**Abstract**

A vast theoretical literature shows that inefficient market structures may arise in free entry equilibria. The inefficiency may manifest itself in the number, variety, or quality of offered products. Previous empirical work demonstrated that excessive entry may obtain in local radio markets. Our paper extends that literature by relaxing the assumption that stations are symmetric, and allowing instead for endogenous station differentiation along both horizontal and vertical dimensions. Importantly, we allow station quality to be an unobserved station characteristic. We compute the optimal market structures in local radio markets and find that, in most broadcasting formats, a social planner who takes into account the welfare of market participants (stations and advertisers) would eliminate 50%-60% of the stations observed in equilibrium. This finding is robust to whether we consider horizontal differentiation only, or both horizontal and vertical differentiation. In 80%-94.9% of markets that have high quality stations in the observed equilibrium, welfare could be unambiguously improved by converting one such station into low quality broadcasting. In contrast, it is never unambiguously welfare-enhancing to convert an observed low quality station into a high quality one. This suggests local over-provision of quality in the observed equilibrium, in addition to the finding of excessive entry.

*Presentation slides for this paper have been publicly available since 2006. We are grateful to seminar and conference participants at Brown University, the Hebrew University, Tel Aviv University (economics & marketing), the University of Cyprus, the University of Pennsylvania (Wharton), the University of Zurich, Stanford University, Yale University, the Annual Meetings of the SEA, the Sixth Conference on the Economics of Advertising and Marketing, Tel Aviv, and the Media and Communications Conference, Chicago.

†Department of Economics, Yale University, and NBER. email: steven.berry@yale.edu

‡Department of Economics, Hebrew University of Jerusalem, and CEPR. email: alon.eizenberg@mail.huji.ac.il

§Carlson School, University of Minnesota, and NBER. email: jwaldfog@umn.edu

# 1 Introduction

A vast theoretical literature (e.g., Spence 1976) shows that free entry equilibria may result in inefficient market structures. The inefficiency may manifest itself in the number, variety, or quality of offered products. In the radio industry, various authors (Steiner 1952, Rogers and Woodbury 1996) have argued that inefficient content duplication and excessive station entry may be prevalent. Berry and Waldfogel (1999, hereafter BW99) demonstrated such excessive entry empirically. In their model, radio stations in each market were symmetrically differentiated.

In this paper, we extend the literature by introducing observed and unobserved product-level differentiation into the empirical study of excessive entry. To the degree that horizontal differentiation is important, prior estimates of excess entry may be overstated. Allowing for vertical differentiation is also important, as it allows us to empirically address questions regarding quality provision in an oligopoly equilibrium, an area in which unambiguous theoretical predictions are difficult to obtain. Our empirical treatment of vertical differentiation is novel, allowing it to be an *unobserved station characteristic*. From an econometric standpoint, we deal with product differentiation via a particularly simple application of recently popular "bounds" methods for treating fixed costs in the presence of multiple equilibria. The partial identification of fixed costs results in partial identification of the market's socially-optimal market structure, and we propose and implement an algorithm that, given the estimated model, computes bounds on the optimal market structure. By comparing these bounds to the observed market structure, we are able to place bounds on the extent of excessive entry into local radio markets.

Excessive entry may obtain if firms incur substantial fixed costs, and offer products that are close substitutes to one another (Mankiw and Whinston 1986). Firms continue to enter the market as long as their private gains exceed fixed costs, ignoring the negative externality associated with their entry, i.e., the reduction in rivals' output. In the context of local radio markets, if stations offer similar content, entrants would mostly "steal business" from other stations, while incurring additional fixed costs, resulting in excessive entry. On the other hand, if stations offer differentiated content, additional stations may help expand the market, creating positive externalities that may offset the additional fixed costs. The free-entry equilibrium may therefore result in excessive entry, insufficient entry, or an optimal amount of entry. Empirical work is required to determine which of these possibilities obtains in a given market.

We find that a social planner who maximizes the joint surplus of stations and advertisers would reduce the number of stations in most formats by about 50%-60%. This finding is robust to the dimension of differentiation considered, i.e., whether we allow for horizontal (format) differentiation only, or for both horizontal and vertical differentiation. Since listeners do not pay for radio content, quantifying their surplus in monetary terms is not possible. We provide, however, several exercises in which we convert listener welfare into monetary terms using reasonable assumptions. This allows us to argue that excessive entry is likely to be present in the data even when taking listeners' welfare into account.

While we find substantial excess entry in the market solution (roughly two times too many stations), the extent of excessive entry that we document is lower than that documented in BW99, where free entry was found to allow four times too many stations. The models differ in three major ways. First, our richer model allows for horizontal differentiation among stations in both the demand and entry models. Second, we allow for vertical differentiation among stations. Third, we employ a bounds approach to estimating fixed costs, and avoid making parametric assumptions on the distribution of fixed costs. This stands in contrast to the point-identification approach in BW99 which, in particular, assumed that fixed costs were normally distributed.

To explore which enrichment of the model explains the contrast between the previous and present results, we estimate a sequence of alternative models. First, we re-estimate the BW99 symmetric model with the new data, generating results nearly identical to the BW99 results: four times too many stations to maximize the surplus of market participants. Second, we estimate a model with only horizontal differentiation, not allowing for vertical differentiation as in our main model. Results from this model are much closer to the results of the main model of the paper: the socially optimal configurations include roughly twice the actual number of stations. We conclude both that relaxing the symmetric approach is important to estimates of the welfare costs of entry and that, even with a richer approach, entry remains excessive. As we discuss in detail in Section 5, pinning down the exact feature that differentiates the predictions of the symmetric vs. nonsymmetric models is difficult. Just the same, the analysis demonstrates that extending the analysis to allow for nonsymmetries, while technically more demanding, is important.

Our framework also allows us to shed light on equilibrium quality choices and their properties. The theories of quality choice are well developed for the monopoly case (Mussa and Rosen 1978, Maskin and Riley 1984), but much less so for the oligopoly case.[1] The difficulty of obtaining theoretical results has motivated empirical work on this issue, such as Mazzeo's (2002) analysis of quality choices in the motel industry. Compared to Mazzeo's motel setup, the quality of a radio station is more difficult to ascertain from observed data. We therefore pursue an approach that treats quality as an unobserved station characteristic, and we provide methods to identify and estimate a model with such unobserved vertical differentiation.

A striking result is that, in 80%-94.9% of markets in which we determine the presence of high-quality stations, welfare could be unambiguously improved by converting one such high quality station into low quality. In contrast, it is never unambiguously welfare-enhancing to convert an observed low quality station into a high quality one. This analysis suggests that over-provision of

[1] For an example of such work, see Rochet and Stole (2002).

quality, in a local sense, characterizes free-entry equilibria in radio markets.

**Methodology.** We base our analysis on a two-stage model. In the first stage, a large number of (ex-ante identical) potential entrants simultaneously decide whether to enter the market, and in which format (or format-quality combination) to operate. The market structure determined in this first stage is, therefore, a vector describing the numbers of stations operating in each format (or format-quality cell). The post-entry asymmetry implies that the market structure may not be uniquely determined in equilibrium. This contrasts with Bresnahan and Reiss (1991) and BW99, where the market structure is a scalar and equilibrium is unique.

In the second stage, entering stations pay fixed entry costs and garner revenues. Our model determines those revenues as follows: a discrete-choice model of listeners' preferences determines, given the market structure, how many listeners are captured by each station. Those listeners are then "sold" to advertisers at a price which is determined from a simple model of advertisers' demand for listeners. A station's revenue is, then, the product of the per-listener price paid by advertisers, and its total number of listeners.

Our discrete-choice model of listener preferences builds on the nested logit model and allows for systematic station differentiation along two dimensions: format (i.e., horizontal) differentiation, and unobserved quality (i.e., vertical) differentiation. We show how to estimate such a model, overcoming the following challenges: first, the fact that quality is unobserved requires us to assign it within the estimation procedure. Our estimation approach classifies stations into quality levels using observed market share variation within an observed market-format data cell. Naturally, the extent of such variation dictates the amount of vertical differentiation that can be incorporated into the model. This motivates us to restrict attention to two quality levels, and to allow for vertical differentiation in two popular broadcasting formats.

Second, quality choices are endogenous, an issue that we address with fixed effects that control for unobserved taste shocks at the market-format level. The presence of fixed effects complicates the estimation of the model substantially, and we propose and implement a two-step estimator that overcomes this issue.

We also estimate advertisers' demand for listeners, modeled by a simple constant-elasticity specification in which the advertising price depends on the total "output" of listeners. This model is estimated via 2SLS and implies a downward-sloping demand curve with a constant elasticity of about $(-2)$, a similar value to that reported in BW99. Put together, the estimated listening equation and the advertisers' demand equation allow us to predict the revenue garnered by each station given any counterfactual market structure.

These revenue predictions allow us to estimate fixed costs, relying on necessary equilibrium conditions from the entry model described above. Suppose, for example, that three stations are observed to operate in the "Rock" format in a given local market. For this to be an equilibrium, it has to be the case that three stations are still profitable, while a fourth entrant would incur a loss. The first condition places an upper bound on the fixed cost of operating a Rock station in this market: the revenue of a Rock station in the observed market equilibrium. The second condition places a lower bound on this fixed cost: the revenue of a counterfactual, fourth entrant into the Rock format. Having estimated the listening equation and advertisers' demand as explained above, these revenue figures are easily computed and provide bounds on the relevant fixed cost. Allowing, in addition, for discrete vertical differentiation implies that stations enter format-quality cells and is similarly handled.

The extant literature typically proceeds by utilizing these bounds on the fixed costs to estimate the distribution of fixed costs across markets. In BW99, these bounds were utilized, along with a parametric assumption on the distribution of fixed cost, to generate an ordered probit estimator for the parameters of this distribution. This point-identification approach is not available here since the potential non-uniqueness of equilibria prohibits us from writing down the likelihood function. To address this issue, the literature offers the possibility of estimating bounds on the parameters of the distribution of fixed costs using moment inequalities.[2] These estimators are often technically demanding and sometimes rely on strong parametric assumptions.

In this paper we take a different approach: instead of using the bounds on fixed costs of operation in different markets to estimate the distribution of fixed costs across markets, we use the market-specific bounds directly in our welfare analysis. This has two benefits: first, we do not have to make parametric assumptions on the distribution of fixed costs across markets (nor do we have to assume that costs are independent across such markets). Second, by not estimating the distribution of fixed costs, we avoid the difficult task of doing inference on this distribution.[3]

Our paper offers some additional methodological contributions that may be applicable in many empirical studies of market structure. We develop an algorithm that, given bounds on fixed costs in format-quality cells, computes *a set of market structures* that cannot be ruled out as socially optimal. Our modeling of station unobserved quality, and the two-step estimation for the listening model that incorporates this quality differentiation, are also novel.

**Relationship to previous literature on the radio industry.** Several features make the radio industry an attractive arena to study product positioning in an oligopoly equilibrium. First, data is available from a large cross-section of local markets characterized by substantial variation in market population and listener demographics. Second, horizontal differentiation is easily observed since stations belong to well-defined broadcasting formats.

Two papers rely on reduced-form techniques to study the impact of mergers on broadcasting variety. Berry and Waldfogel (2001) document that the merger wave that followed from the 1996 Telecommunication Act reduced station entry but increased the variety of offered programming. Sweeting (2010) uses playlist data to study how station merger decisions affect positioning, and

[2]Some examples include Ho (2009), Crawford and Yurukoglu (2011), Ho and Pakes (2014), Eizenberg (2014).

[3]See, for example, Chernozhukov, Hong, and Tamer (2007), Andrews and Jia (2012), Pakes, Porter, Ho and Ishii (2015).

finds that, following a merger, owners of merging stations tend to push them apart (in characteristics space) to limit cannibalization, but at the same time reduce the extent of differentiation with respect to competitors.

Another line of research studies the market structure in the radio industry via the estimation of a dynamic oligopoly game (Jeziorski (2012, 2013a, 2013b), and Sweeting (2013)). Dynamic models are well suited to considering explicitly dynamic questions. For example, Sweeting's (2013) dynamic analysis emphasizes the estimation of repositioning costs of existing stations. On the other hand, the current state of estimation techniques for dynamic oligopoly requires many strong assumptions. We view the static modeling pursued in our paper as complementary to these dynamic models, especially when the questions at hand are not explicitly dynamic.

The static approach determines the cross-section of equilibrium market structures as a function of the local market's population and its socioeconomic and demographic makeup. This leads to a simple and transparent setup for studying the possibility of excessive entry, a question motivated both by theory, and by previous studies of the radio industry cited above. A static model has a couple of additional benefits in our setup: first, it makes our analysis conceptually comparable to BW99's analysis, allowing us to explore the impact of relaxing symmetry assumptions and admitting multiple equilibria. Second, it does not require a parametric assumption on our key primitive, the distribution of fixed costs.[4] Another relevant work is Goettler and Shachar (2001), who estimate a spatial location model of television program choice relying on an estimated discrete choice model that allows for latent product attributes. Unlike our work, that paper relies on panel data in devising its identification strategy, whereas our approach is suitable for the cross-sectional data utilized in this paper.

The remainder of this paper is organized as follows: Section 2 describes the data and station format classifications. Section 3 describes the various components of our model, and the estimation of its primitives. Section 4 uses the estimated model to analyze the discrepancy between the free-entry equilibrium and the optimal market structure. Section 5 performs a comparison to simpler models that eliminate one or both dimensions of station systematic differentiation. Section 6 discusses robustness, and Section 7 concludes.

## 2 Data

The data used in this study cover a cross-section of metropolitan radio markets in 2001. Market definitions follow those of Arbitron, a media marketing research firm that tracks activity and trends in the radio industry. While some

[4]Smith and O'Gorman (2008) study (independently from our work) the distribution of fixed costs in the radio industry relying on a static model and a partial identification approach. That paper pursues very different questions compared to our paper. In particular, we focus on questions that pertain to product variety, develop tools that compute optimal market structures, and address unobserved vertical differentiation issues.

of Arbitron's 286 radio markets coincide with Census MSA definitions, others do not.

Data regarding stations and listenership in these markets is obtained from the Spring 2001 edition of *American Radio*, by Duncan's American Radio. Rich information regarding individual stations is available from this source. We observe each station's "AQH-listeners," i.e., the number of listeners of age 12 and above who listened to the station during the average quarter-hour in Spring 2001.[5] These listenership figures are provided by Arbitron based on diaries retrieved from surveyed individuals in each market. We further observe the station's broadcasting format, which plays a key role in our analysis, and whether the station is considered "home to the market." A station's "home market" is determined by its city of FCC license, and it may additionally appear in the Arbitron data as an out-metro station in markets where it…"accumulated enough listening within the...metro to rate inclusion in the market report."[6] Finally, we observe technical information: whether the station is on the FM or AM band, its broadcasting wattage and antenna height.

At the market level, we observe the market's population of persons 12 or older, and the total number of diaries retrieved by Arbitron. The number of retrieved diaries is related to the accuracy of the listenership data. Our empirical model utilizes this information to formally take account of potential measurement error in reported market shares.

We compute the market share of each station by dividing the number of its AQH-listeners by total market 12+ population. The share of the "outside option" of not listening to commercial radio is computed as 1 minus the sum of stations' individual shares. Non-commercial stations (e.g., public radio), as well as commercial stations not listed by Arbitron (e.g., due to very low listening, or violation of Arbitron's rules), are included in this outside option.

Additional market-level data were obtained from *Duncan's Radio Market Guide*. From the 2002 edition, we obtained estimates of each market's total revenue in 2001 (i.e., the combined annual revenue of the market's stations). While some estimates are based on information provided by radio stations to their accounting firms (or directly to Duncan's Radio Market Guide), other estimates are based on Duncan's assessments. Similarly as in BW99, we compute the market's *ad price*, i.e., the average price paid by advertisers for an AQH-listener, by dividing total market revenue by the total number of listeners to in-metro stations.[7]

[5]This average is computed over all quarter-hours in the standard survey week: Monday-Sunday, 6AM - 12 Midnight.

[6]Source: Duncan's American Radio 2001. In some cases, the Arbitron data classifies a station's home market based on the station's choice, which may differ from its official city of license.

[7]This calculation assumes that all of the market's revenue is garnered by in-metro stations, i.e., stations that are home to the market. This assumption could be put into question in markets where substantial listenership is enjoyed by out-metro stations. For example, the extreme ad price of $2691 reported in Table 1 below occurs at Bridgeport, Connecticut, a market heavily served by out-metro stations. This assumption, however, may be justified even for such markets: it may be costly for a local advertiser to reach the local audience when the market is heavily served by out-metro stations, and the relatively high ad price may reflect that.

An alternative approach to using total market revenue data, pursued recently by Sweeting (2013) and Jeziorski (2012, 2013a, 2013b) utilizes station-specific data obtained from different sources than those we use here. These station-specific revenue data are computed using various assumptions based on methodology proprietary to the data provider. Our approach has both advantages and disadvantages: while it forces the ad price to be identical across different formats, it also avoids the potential measurement error stemming from the methodology used to assign station-specific revenue numbers. Ultimately, we employ the total market revenue figures, in part because of our interest in staying conceptually close to the methodology used in BW99.

The 2001 edition of Duncan's Radio Market Guide provides market-level demographic information for the year 2000. In particular, the market's percentage of Black and Hispanic population, average income, and percentage of college-educated is available. We have full data (including revenue and demographic information) for 163 of Arbitron's 286 markets, and we restrict our analysis to those 163 markets.[8] After dropping observations (stations) with reported zero listenership, the data we use cover 4,362 stations in the included markets. Finally, we classify markets into geographic regions (Northeast, Midwest, South, and West) based on Census definitions.

Summary statistics on some of the market-level variables are available in Table 1. The mean listenership share (i.e., the share not choosing an outside option) is about 12%. The average market has 19.6 in-metro stations, and 7.2 out-metro stations. The average ad price is 570 US$. Since this price was computed using annual revenue, it represents the average price paid for one listener over the course of one year.

Table 1: Description of Market-Level Data

| Variable | Units | Mean | Std. Deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| Share in-metro | % | 0.111 | 0.026 | 0.030 | 0.151 |
| Share Out-metro | % | 0.015 | 0.023 | 0.000 | 0.104 |
| N1 (in-metro) | integer | 19.64 | 7.565 | 4.000 | 45.00 |
| N2 (out-metro) | integer | 7.209 | 8.320 | 0.000 | 37.00 |
| Population | millions | 1.016 | 1.687 | 0.075 | 14.48 |
| Ad Price | $ | 570.5 | 237.7 | 258.2 | 2691.2 |
| Income | 10,000$ | 4.584 | 0.860 | 2.482 | 8.010 |
| College | % | 21.20 | 5.370 | 10.20 | 37.10 |

Notes: computed using the 163 markets with full data, see text.

**Format classification.** Stations' broadcasting formats represent horizon-

[8]Revenue figures are available for 200 markets, however the sample size is further restricted since we only collected the full demographic information for a subset of 163 such markets. Our restricted sample provides good coverage of the major metropolitan markets: it covers 29 of the top 30 markets (by population), and 89 of the top 100 markets. At the same time, it also covers some of the smallest markets (e.g., the smallest market covered is ranked 274 out of 286 in terms of its population). All in all, we believe that our data cover a representative sample of the population of markets of interest.

tal differentiation and, therefore, play an important role in our analysis of variety in radio markets. The number of different formats in the data is close to 70, motivating an aggregation into higher-level categories. We classify formats into ten such categories, based on intuition gained from a large number of sources about the nature of the formats. The ten format categories are described in Table 2.

Some idea on the performance of these format categories is provided in Table 3. The "Frequency" column describes the share of markets where a given category is represented by at least one station. Three format categories raise potential selection issues: "Religious" stations are present in about 80 percent of the markets, while "Urban" and "Spanish" are present in 74 and 40 percent, respectively. We discuss robustness to this issue in Section 6. Additional columns of Table 3 reveal that the most popular format (in terms of total format listening share, averaged across markets) is "Mainstream," followed by "Rock", "Country" and "News/Talk."

# 3 Model

The model has three components. The first component, described in section 3.1, is the listening equation which determines stations' market shares as a function of listeners' tastes, conditioning on a given market structure. The second component, described in section 3.2, models the other side of this media market: advertisers' demand for listeners. Together, the listening function and the advertisers' demand function determine stations' revenues given any fixed market structure. The market structure itself is determined by the third component of the model: the entry game, described in section 3.3.

## 3.1 The listening model

We introduce a listening model that departs from BW99 by incorporating two important dimensions of station differentiation. Horizontal (format) differentiation is accommodated via a nested-logit structure that treats the ten formats as nests, with an eleventh nest that contains the outside option of not listening to commercial radio. In addition, we admit unobserved vertical differentiation. We model quality as discrete, effectively creating a set of horizontal/vertical cells into which stations can enter. Although our arguments generalize to a larger number of discrete quality levels, we use only two levels for the unobserved discrete station quality of in-metro stations,"high" and "low." Practical considerations involving data variation motivate us to allow for quality differentiation in two dominant formats only, as we explain below.

How should one define and measure station quality? Our data offer, at best, some imperfect proxies for quality, such as the station's broadcasting wattage. A station's actual quality is likely to depend primarily on the quality of the content provided, a feature which is inherently difficult to quantify. As a consequence, we choose to model quality as an *unobserved* station characteristic

Table 2: Format classification

| Format Group | Formats Included | | | | |
|---|---|---|---|---|---|
| **"Mainstream"** | Adult Cont. | Hot AC | Modern AC | Soft AC | Adult Altern. |
| | Classic Hits | 80s Hits | | | |
| **CHR** | CHR | | | | |
| **Country** | Country | Classic Cntry. | Trad. Country | | |
| **Rock** | Rock | Active Rock | Modern Rock | Classic Rock | |
| **Oldies** | Oldies | | | | |
| **Religious** | Religious | Cont. Christ. | Black Gospel | Gospel | S. Gospel |
| **Urban** | Urban | Urban AC | Urban Oldies | Rhythmic Old. | |
| **Spanish** | Spanish | Span.-Oldies | Span.-Adult Alt | Span.-C. Christ | Span.-CHR |
| | Span.-Cl. Hits | Span.-EZ | Span.-Hits | Span.-NT | Span.-Relig. |
| | Span.-Talk | Tejano | Tropical | Reg'l Mex. | Span.-Stand. |
| | Ranchero | Romantica | | | |
| **News/Talk** | News/Talk | News | Talk | Hot Talk | Bus. News |
| | Sports | Farm | | | |
| **Other** | Variety | Bluegrass | Blues | cp-new | Americana |
| | Pre-teen | Ethnic | Silent | A22 | A26 |
| | A30 | N A | Jazz | Smooth Jazz | Dance |
| | Classical | Adult Stand. | Easy List. | | |

Table 3: Format category performance

| Format Group | Frequency | Mean stations | Max stations | Mean format share |
|---|---|---|---|---|
| **"Mainstream"** | 100.00% | 4.48 | 11 | 2.31% |
| **Rock** | 100.00% | 3.42 | 9 | 1.88% |
| **Country** | 99.39% | 2.99 | 9 | 1.85% |
| **News/Talk** | 100.00% | 4.31 | 13 | 1.55% |
| **Urban** | 73.62% | 2.10 | 6 | 1.24% |
| **CHR** | 93.25% | 1.66 | 6 | 1.16% |
| **Other** | 94.48% | 2.80 | 9 | 1.09% |
| **Oldies** | 98.16% | 1.48 | 5 | 0.79% |
| **Spanish** | 40.49% | 1.63 | 15 | 0.40% |
| **Religious** | 79.75% | 1.88 | 6 | 0.37% |

Notes: The first column describes the frequency with which a metro has at least one station (in- or out-metro) in format. Statistics computed over the 163 markets, both in- and out-metro taken into account.

that shifts listeners' mean utility. Stations' quality classifications are treated as discrete parameters to be estimated along with the other parameters of the model. An additional challenge is the endogeneity of horizontal and vertical differentiation: a station's quality and format choices may depend on the unobserved tastes for broadcasting formats in the relevant market. We address this by including market-format fixed effects in listeners' utility specification. As a consequence of those challenges, estimation becomes more complicated compared to the more standard nested logit model which is amenable to estimation via a linear equation, as in Berry (1994) and BW99.

Let $t = 1, ..., T$ denote our observed metropolitan radio markets, and $j = 1, ..., J_t$ index stations operating in market $t$. We partition the set of stations into $g = 0, ...10$ nests such that $g = 0$ corresponds to the outside option and the remaining nests are the ten broadcasting formats. The utility for listener $i$ from listening to station $j$ in format $g$, in market $t$ is defined by:

$$u_{i,j\in g,t} = \delta_{jt} + \nu_{igt}(\sigma) + (1-\sigma)\epsilon_{ijt}, \tag{1}$$

where $\delta_{jt}$ is the mean utility common to all listeners from listening to station $j$, and $\nu_{igt}(\sigma) + (1-\sigma)\epsilon_{ijt}$ captures the listener's idiosyncratic deviation from the mean utility. The term $\nu_{igt}(\sigma)$ captures an idiosyncratic taste of listener $i$ toward format $g$, and has a unique distribution derived by Cardell (1997), which depends on the parameter $\sigma \in [0, 1)$. The shock $\epsilon_{ijt}$ is an idiosyncratic taste of listener $i$ toward station $j$ in market $t$, assumed to follow a Type-I Extreme Value distribution. These shocks are independently and identically distributed across listeners, stations and markets.

The extent to which individual stations within the format are allowed to deliver unique benefits via the $\epsilon_{ijt}$ term is determined by the estimated parameter $\sigma$, which captures the degree of within-nest correlation in unobserved individual tastes. As $\sigma$ approaches 1, the unobserved tastes of any individual listener toward stations within the same format become near-perfectly correlated, leading to strong "business stealing" within the format. In contrast, Cardell's unique distribution guarantees that, as $\sigma$ approaches 0, $\nu_{igt}$ approaches zero as well, implying no correlation in unobserved tastes, corresponding to maximal diversity in the content provided by stations within the format. Estimating $\sigma$, therefore, is a key task for our empirical framework: the value of $\sigma$ informs us about the scope of business stealing and potential excessive entry. The transparent association of the $\sigma$ parameter with "business stealing" is an advantage of the nested logit framework in this context.

The mean utility $\delta_{jt}$ is specified as follows:

$$\delta_{jt} = \gamma^q \cdot q_{jt} + \gamma^h \cdot h_{jt} + \psi_{gt}. \tag{2}$$

where the dummy variable $q_{jt}$ takes the value 1 for high quality stations, and zero otherwise. The variable $h_{jt}$ is a "home" dummy variable, taking the value 1 if $j$ is an in-metro station, while $(\gamma^q, \gamma^h)$ are parameters to be estimated. Since quality is unobserved, the values of the high-quality dummies $q_{jt}$ are also treated as parameters to be estimated. The term $\psi_{gt}$ is a format-market fixed effect, capturing the mean taste for format $g$ in market $t$. This fixed effect is specified to depend, in turn, on both observed and unobserved variables,

$$\psi_{gt} = d_{gt}\lambda + \xi_{gt}, \tag{3}$$

where $d_{gt}$ is a vector of observed variables including market-level variables such as average income and college education, format dummy variables, and natural interaction terms (e.g. the percentage of Hispanic population interacted with the Spanish format dummy). The parameter vector $\lambda$ captures the effect of those shifters on the mean utility, while $\xi_{gt}$ is an unobserved taste for format $g$ in market $t$.

For simplicity, we impose the restriction that quality differentiation only applies to in-metro stations in the market, so that $q_{jt} = 0$ whenever $h_{jt} = 0$. We thus have three cells of stations within each format-market pair: out-metro, in-metro low-quality, and in-metro high-quality stations, with respective quality levels of 0, $\gamma^h$ and $(\gamma^h + \gamma^q)$.

As indicated above, out-metro stations play a marginal role in most markets, and they systematically garner lower market shares than stations that are home to the metro. We assume throughout our analysis that their presence as "out-metro" stations in a given market is exogenous (in line with the assumptions in BW99). This presence stems from the fact that their signal is strong enough to be captured in that market and garner sufficient listenership to be included in the Arbitron report. We treat as endogenous the decision to be licensed in a given market as a "home" station, and this assumption is justified by the fact that most (if not all) the station's revenue is likely to arise from the audience in its city of license. Our choice not to classify out-metro stations into "high" and "low" quality stems mostly from tractability: as will become clear below, our approach to classifying stations into quality levels relies on observing a significant number of stations in the relevant data cell, with sufficient market share variation among them.

Mean utilities, and hence expected market shares, are predicted to be identical within each of the three cells. The restriction that out-metro stations offer low quality has empirical implications: it orders predicted market shares such that, as long as $\gamma^h$ and $\gamma^q$ are positive, out-metro stations have lower predicted shares than in-metro low-quality stations, which in turn have lower shares than in-metro high-quality stations. Since out-metro stations do typically have lower shares than in-metro stations, we view this as a reasonable and useful simplification.

#### 3.1.1 Identification of the listening model

We begin by considering identification when the expected market shares are perfectly observed – that is, there is no sampling error due to the Arbitron listener diaries. We can think of this as an approximation to the case where there are very many sampled listeners in every market. We assume throughout that $\gamma^h$ and $\gamma^q$ are positive and so the shares of high quality in-metro stations are higher than those of low quality in-metro stations. Thus, in any market where we see two distinct market share values for in-metro stations, the higher market share implies high quality and the lower share implies low quality.[9]

When market shares are equal for all in-metro stations, identification is harder. The reason is the confounding effect of the market-format taste terms $\psi_{gt}$. Consider a market where all the in-metro stations have the same market share within some format. For any guess at the quality level of stations in the format, there is a value of $\psi_{gt}$ that explains the observed common level of shares. Because the market-format taste does not affect the within format shares, we can avoid this potential problem of non-identification if we focus on the within format shares. This is similar to the idea of "differencing out" a fixed effect to deal with endogeneity. To proceed, let

$$\begin{aligned} \kappa_1 &\equiv \gamma^q/(1-\sigma), \\ \kappa_2 &\equiv \gamma^h/(1-\sigma), \end{aligned} \tag{4}$$

and let the vector $\kappa \equiv (\kappa_1, \kappa_2)$. The nested logit then implies that conditional on choosing format $g$ the expected probability of choosing station $j$ in market $t$ (the "within format share") is given by

$$p_{j/gt}(\kappa, q) = \frac{exp(\kappa_1 \cdot q_{jt} + \kappa_2 \cdot h_{jt})}{\sum_{\ell \in g} exp(\kappa_1 \cdot q_{\ell t} + \kappa_2 \cdot h_{\ell t})}, \tag{5}$$

where $q$ is notation for the long vector of quality levels for all markets' stations. Of course, the expression in (5) depends only on the quality levels in the given market-format.

This expression for within format shares allows us, first, to identify $\kappa$ using data on markets where differences in shares identify quality levels. We can then

[9] If we could perfectly observe in-metro expected market shares and we found these took on more than two levels, then we could reject the model with only two qualities for in-metro stations.

use $\kappa$, together with out-metro shares, to identify quality levels in additional markets. In particular, we can identify $\kappa$ from a single market $t$ with two in-metro stations ($j$,$k$) in format $g$ such that $s_{jt} > s_{kt}$. Because the shares are different, we know that $q_{jt} = 1$ and $q_{\ell t} = 0$. A simple manipulation of the nested logit share equation then identifies $\kappa_1$ as

$$\kappa_1 = ln(s_{jt}) - ln(s_{kt}).$$

Performing a similar exercise with two stations such that $q_{jt} = 0, h_{jt} = 1, h_{\ell t} = 0$ (i.e., $j$ is known to be a low quality in-metro station, while $\ell$ is an out-metro station), identifies $\kappa_2$:

$$\kappa_2 = ln(s_{jt}) - ln(s_{\ell t}).$$

Given $\kappa$, we can then identify the quality level for in-metro station $j$ in format $g$ in any market $t$ that has an out-metro station, denoted $\ell$, in that same format. As usual, the nested logit structure delivers the following equation for station $j$:

$$ln(s_{jt}) - ln(s_{0t}) = \gamma^q q_{jt} + \gamma^h h_{jt} + \psi_{gt} + \sigma ln(s_{j/g,t}).$$

Writing the same expression for station $\ell$ and subtracting one from the other leads to:

$$q_{jt} = \frac{1}{\kappa_1} ln\big(s_{jt}/s_{\ell t}\big) - \frac{\kappa_2}{\kappa_1}.$$

This gives us identification of quality whenever there is an out-metro station. Since we treat the presence of out-metro stations as exogenous, we have an exogenously chosen sample of markets where station-level quality is identified.

Since we trivially have identification of quality whenever in-metro stations have different shares, this leaves us with one remaining case of partial identification: market/formats with no out-metro station and identical shares for in-metro stations.[10] In these market/formats, we know that either [i] all stations are high quality or [ii] all stations are low quality. In the sections on estimation and counterfactual simulation below, we discuss how we approach this issue of partial identification in some market/format pairs.

Note that having identified qualities, we move back to the usual case of Berry (1994) with a nested logit where all characteristics are observed (although recall that the unobservable taste variable $\xi_{gt}$ is at the level of the format, not station.) Thus, the remainder of the demand identification problem is standard and we will need an instrument variables approach to identify $(\gamma^q, \gamma^h, \sigma)$ separately, as opposed to the composite parameters $(\kappa_1, \kappa_2)$ that are identified from the within format choice problem. The procedure that implements this approach in practical estimation is reviewed in the next subsection.

[10]The case of "identical shares" obviously includes the case where there is only one in-metro station in the market/format.

### 3.1.2 Estimating the listening model

Analogous to the identification argument, we consider estimation in two steps. First, we estimate $(\kappa, q)$ from the within group shares. Second, we use a more classic IV method to estimate the remaining parameters.

**Step 1: estimating quality levels.** Moving from identification to estimation, we face the problem that we do not observe expected market shares, but only sampled shares computed from the Arbitron diaries. The sampling error means that we cannot directly observe whether expected shares are equal or not. Indeed, even when expected shares are exactly equal, we are exceedingly unlikely to observe $s_{jt} = s_{\ell t}$ for two stations $j$ and $\ell$. However, the observed Arbitron shares are a draw from a multinomial distribution with known properties, so for estimation we can employ a maximum likelihood approach. In particular, to estimate quality we consider maximum likelihood estimation based on the within format shares where, since $\psi_{gt}$ drops out, the only sampling error is from the Arbitron diaries, and the endogeneity problem of potential correlation between quality and unobserved taste is not present.

Denote by $n_{jt}$ the number of Arbitron diaries reporting listenership to a given station.[11] The log-likelihood for the within group choices, conditional on choice of format, is then

$$log\mathcal{L}(\kappa, \kappa, q) = \frac{1}{\overline{N}} \sum_{t} \sum_{g} \sum_{j \in gt} n_{jt} \cdot log\big[p_{j/gt}(\kappa, q)\big] \tag{6}$$

To derive the asymptotic behavior of the ML estimates, we take the total number of Arbitron diaries,

$$\overline{N} = \sum_{t} \sum_{j=0}^{J_t} n_{jt}$$

to infinity, holding fixed the relative sample sizes in each market.

The discrete quality parameters $q$ raise issues of both estimation and computation. There is one quality parameter per station, but this is not a problem since there are a large number of diary responses per station. Our estimates are consistent, by usual arguments, as the number of sampled diaries goes off to infinity. Indeed, since each quality parameter is discrete, taking on only two values, by usual arguments the estimate of quality is super-efficient – it converges faster than rate $\sqrt{\overline{N}}$. In contrast, $\kappa$ converges at the usual rate.

There is also the computational issue of maximizing the likelihood over the large number of possible combinatoric assignments of quality. First note that, conditional on $\kappa$, the quality assignment breaks up across market/formats—the assignment of quality in one market/format does not affect the likelihood contribution of other market/formats. Second, as long as $\gamma^q > 0$, in any market/format a high-quality station has a higher predicted market share than a

[11] Strictly speaking, this is the number of diaries reporting listenership to the station in the average quarter-hour, see Section 2.

low-quality station. It is easy to show that if the maximum likelihood estimate of $q$ assigns a high quality to given station $j$, it also assigns high quality to all stations with observed sample shares larger than $j$. Thus, the problem of estimating the quality vector for any market/format reduces to the problem of choosing the threshold station: the station with the largest observed share that still corresponds to a low quality station. We denote the index of this threshold station by the discrete parameter $j'_{gt}$. If $j'_{gt} = 0$, all in-metro stations in the market-format pair offer high quality. If this index is equal to the number of such in-metro stations, all those stations offer low quality.[12] If there are $J_{gt}$ stations in market/format $(g, t)$ then conditional on a value of $\kappa$ we have to compute the likelihood only $J_{gt} + 1$ times to choose the best value for the threshold $j'_{gt}$.

The analog of the partial identification problem discussed above arises in the estimation context for market/format pairs that have no out-metro stations. In all of these market/formats, and for each value of $\kappa$, setting $j'_{gt}$ equal to either zero (implying that all stations offer high-quality) or to the number of in-metro stations (implying that all stations offer low-quality) yields the same value for the log-likelihood contribution of the market-format. If that is also the value that maximizes the likelihood, then we have a set estimate of the qualities of stations in this market-format pair: the maximized likelihood is generated by the case where the stations are all of high quality and by the case where all are of low quality.

Importantly, because the ML objective function obtains the same value whether we assign the unclassified stations to be of high quality, or of low quality, our ML estimates for $\kappa$ are unaffected by the set estimates of quality. However, the "IV" estimates of $(\gamma^q, \gamma^h, \sigma, \lambda)$, discussed below, will be affected by the allocation of all stations to either high or low quality. We discuss strategies to deal with this below.

**Step 2: estimating the remaining parameters via a restriction on the distribution of $\xi$.** Having obtained estimates of $(\kappa, q)$, we now hold those fixed and proceed with estimating the remaining parameters of interest. Given $\kappa$ and $\sigma$ we can solve for $(\gamma^q, \gamma^h)$ from (4) so we can treat the "business stealing" parameter $\sigma$, together with the format taste parameters $\lambda$ in (3), as the only remaining unknowns. To ensure that we have point-estimates of quality for all market/formats, our base case estimation allows for quality differentiation in only a subset of formats and, for those formats, uses only the exogenously selected sample of market/formats with out-metro stations.

Our identifying assumption, which follows much of the literature on estimating differentiated-product demand models, states that the unobserved format-level taste shifter in equation (3) is mean-independent of a set of instruments,

$$E[\xi_{gt} | Z_{gt}] = 0. \tag{7}$$

In the empirical application, we let $Z_{gt}$ contain the market's population, the

[12] Note the slight abuse of notation, made for convenience: here we use $j$ to index stations within the market/format pair, whereas elsewhere we use it to index stations in the entire market.

number of the market's out-metro stations, and the number of out-metro stations in the same format, as well as the $d$ covariates. Treating population and the presence of out-metro stations as exogenous follows BW99, and we believe that these are reasonable restrictions. Since stations garner most, if not all their revenue from advertisers in their home market, their presence in other markets as "out-metro" stations can be reasonably viewed as independent of the taste shocks in those other markets. As for population, our entry model (see below) implies that it is an effective shifter of variable profits and, therefore, entry. As a consequence, it is an effective instrument for within-format shares in the listening equation.

To be a valid instrument, however, population should also be uncorrelated with the unobserved taste shifter $\xi_{gt}$, and one may be worried that more populous markets may systematically display a strong taste to particular formats such as Urban or Hit radio. Our view is that, by controlling for the market-format fixed effect $\psi_{gt}$, and requiring only the error in that fixed effect $\xi_{gt}$ to be mean-independent of population, our approach is robust to this concern. Just the same, this concern may be valid for our simpler models, discussed in Section 5 below, that do not control for such fixed effects. Even there, however, we believe that by controlling for features of the market such as the percentage of population that belongs in particular demographics (Black, Hispanic, college-educated) we are able to address this concern.

A crucial step is to note that, given $(\kappa, q, \sigma)$, there is a unique vector of fixed effects $\psi$ that maximizes the overall multinomial log-likelihood of the observed shares. Further, there is a closed form solution for $\psi$. Appendix D shows that

$$\psi_{gt}(\kappa, q, \sigma) = log(s_{gt}) - log(s_{0t}) - (1-\sigma)log[\sum_{j\in g} e^{(\kappa_1 q_{jt} + \kappa_2 h_{jt})}]. \tag{8}$$

This further suggests that, given candidate values for $(\sigma, \lambda)$, and the fixed estimates $(\hat{\kappa}, \hat{q})$ obtained in step 1, we can solve for the unobserved taste shifter $\xi_{gt}$ as

$$\xi_{gt}(\hat{\kappa}, \hat{q}, \sigma, \lambda) = \psi_{gt}(\hat{\kappa}, \hat{q}, \sigma) - d_{gt}\lambda$$

The mean-independence condition (7) now motivates estimating $(\sigma, \lambda)$ by minimizing the following classic GMM objective function:

$$J(\sigma, \lambda; \hat{\kappa}, \hat{q}) = \left[\psi(\sigma, \hat{\kappa}, \hat{q}) - d\lambda\right]' Z\Phi Z' \left[\psi(\sigma, \hat{\kappa}, \hat{q}) - d\lambda\right].$$

where $\hat{\kappa}$ is the vector of the first-stage estimates which we hold fixed in this GMM estimation procedure, $d$ is a matrix whose rows are the $d_{gt}$ covariates, $Z$ is the instrument matrix, and $\Phi = (Z'Z)^{-1}$ is a weighting matrix.

This objective function can be further simplified by noting that, conditional on $\sigma$, it is possible to "concentrate out" the $\lambda$ parameters[13], leading to a GMM objective that can be maximized by searching over values of the scalar parameter $\sigma$ only:

$$J(\sigma; \hat{\kappa}, \hat{q}) = \left[\psi(\sigma; \hat{\kappa}, \hat{q}) - d\lambda(\sigma; \hat{\kappa}, \hat{q})\right]' Z\Phi Z' \left[\psi(\sigma; \hat{\kappa}, \hat{q}) - d\lambda(\sigma; \hat{\kappa}, \hat{q})\right] \tag{9}$$

[13]See Berry, Levinsohn and Pakes (1995), Nevo (2000).

This second-step estimation yields the GMM estimates $(\hat{\sigma}, \hat{\lambda})$. Finally, estimates for the home and quality effects are now easily computed by $\hat{\gamma}^h = \hat{\kappa_2} \cdot (1 - \hat{\sigma})$ and $\hat{\gamma}^q = \hat{\kappa_1} \cdot (1 - \hat{\sigma})$, respectively.

**Dealing with set-estimates of qualities.** A remaining issue is how to deal with market/formats in which two different vectors of quality levels are equally consistent with the data. Note that there are many formats that feature a small number of stations, including a small number of out-metro stations. In these formats it is often not possible to point-estimate quality levels. Importantly, then, we restrict our endogenous unobserved quality differentiation to apply only to the main music format, Mainstream, and to the News/Talk format. Mainstream has the highest listening share among all ten formats while News/Talk is the leading non-music format. In these formats the quality assignment procedure seems to work robustly well. Stations in the other eight formats are assumed to offer a single quality level.

Even in these two formats, there are still market-format pairs where in-metro stations could not be assigned to a single quality level in step 1 of our procedure. Quality in the Mainstream format was undetermined in 44 out of the 163 markets (27%). In the News/Talk format, it was undetermined only in 17 markets (10.4%). It may be that quality differentiation is particularly pronounced in News/Talk. In market/formats with set-estimates of quality, we cannot compute the solution for $\psi_{gt}$ which is necessary for the second estimation step.

Our solution is a version of an "exogenous selection" procedure often used to overcome sample selection problems. Recall that, in the presence of an out-metro station, quality is assigned with probability 1, and that by our assumptions, the presence of out-metro stations is considered exogenous to taste shocks at the market-format level. Eliminating from the GMM objective function market-format pairs that have no out-metro stations, therefore, leaves us with observations in which quality is always assigned. This approach leads to estimators that are robust to selection bias.[14] We therefore pursue this strategy as our leading specification. A total number of 180 out of 1,433 market-format pairs are dropped in practice, leaving us with 1,253 observations. We report below several robustness checks for this approach.

Standard errors for the first-step ML estimator of $(\hat{\kappa}, \hat{q})$ were obtained using the usual ML formulae. In practice, we only computed standard errors for $\hat{\kappa}$, and not for the many threshold quality parameters $\hat{j}'$, which converge at a faster rate. Standard errors for $(\sigma, \lambda)$, estimated in the second step, were corrected for the error stemming from the first-step estimation of $(\hat{\kappa}, \hat{q})$ using results for two-step estimation models (see Newey and McFadden (1986)). Details are available in Appendix E.[15]

[14] An alternative approach, which would remove only the market-format pairs where quality was actually unassigned, would not be robust to selection bias.

[15] We are grateful to Donald Andrews for his feedback on some aspects of these calculations. For simplicity, the computations were performed using all observations, rather than excluding observations from the 180 market-format pairs discussed above. Given that estimation results were reasonably robust to this exclusion, this issue is not likely to have a major impact on the estimated

**Estimation results.** Table 4 presents estimation results for the listening model that allows for both horizontal and vertical differentiation. Panel A of this table presents the Maximum Likelihood estimation results for the parameters $\kappa$. This procedure has 4,362 observations, representing individual stations in the sample.

Panel B presents results obtained from the second step of our procedure, which holds fixed the estimated values of $(\kappa_1, \kappa_2, j')$ and estimates the parameters $(\sigma, \lambda)$ via GMM. An estimate of 0.589 is obtained for the correlation parameter $\sigma$. As expected, both $\gamma^h$ (the effect of in-metro status) and $\gamma^q$ (the effect of quality) are positively signed. The $\lambda$ coefficients are highly intuitive, with popular formats (e.g. Mainstream, Country, Rock) obtaining larger estimated coefficients than less popular formats. Also apparent is the important role played by interaction terms.

To examine robustness to our handling of the missing quality assignments, we consider three alternatives to the exogenous selection approach which led to the elimination of 180 market-format pairs that did not have an out-metro station. The first robustness check utilizes all 1,433 market-format pairs and sets all undetermined qualities to "low." This yields a value for $\sigma$ of 0.569, i.e., very close to the 0.589 from our leading specification. A second robustness check also keeps all market-format pairs, but sets all undetermined quality to "high." This yields a somewhat higher estimate for $\sigma$: 0.702. Finally, the third robustness check eliminates all observations pertaining to the Mainstream format. This leaves us with eight formats in which quality assignments are assumed to be fixed, and one format—News/Talk—in which quality assignment succeeds in close to 90% of markets. We set the unassigned cases to "low" quality. This yields an estimate for $\sigma$ of 0.503.

These robustness checks suggest that the estimates obtained from our baseline specification are reasonable, in addition to being theoretically justified by the exogenous selection approach. It is these estimates, therefore, that we carry forward to the remainder of the empirical analysis.

### 3.2 Advertisers demand for listeners

Having estimated listeners' demand for programming, we now describe the second component of our framework: a model for advertisers' demand for listeners. Here, we face similar data issues as in BW99 and we closely follow the approach taken there. The model relates the price of advertising to the share of the population listening to in-metro stations, as well as to market characteristics such as demographic and regional effects. We assume that a station's revenue is proportional to the number of its AQH-listeners. Stations "produce" listeners and sell them to advertisers at a price which is determined from advertisers' inverse demand curve, reflecting advertisers' willingness to pay for listeners. Market $t$'s inverse demand curve is given by the following

standard errors.

Table 4: Listening equation estimates

| Parameter | Estimate | SE |
|---|---|---|
| **A. "First step" estimates (4,362 observations)** | | |
| $\kappa_1$ | 1.472 | 0.006 |
| $\kappa_2$ | 1.134 | 0.008 |
| **B. "Second step" estimates $(\sigma, \lambda)$ (1,253 observations)** | | |
| $\sigma$ | 0.589 | 0.017 |
| constant | -5.143 | 0.007 |
| northeast | 0.097 | 0.008 |
| midwest | 0.067 | 0.010 |
| south | 0.088 | 0.011 |
| mainstream | 0.450 | 0.007 |
| chr | 0.438 | 0.007 |
| country | 0.617 | 0.007 |
| rock | 0.642 | 0.011 |
| oldies | 0.067 | 0.006 |
| religious | -0.954 | 0.004 |
| urban | -0.473 | 0.006 |
| spanish | -1.235 | 0.007 |
| nt | 0.189 | 0.004 |
| income/10 | -0.092 | 0.003 |
| college/10 | -0.656 | 0.001 |
| black/10 | -0.712 | 0.002 |
| hisp/10 | -0.370 | 0.003 |
| blackXurban/10 | 5.555 | 0.001 |
| hispXspan /10 | 3.962 | 0.002 |
| **C. Quality and home effects $(\gamma^q, \gamma^h)$*** | | |
| $\gamma^q$ | 0.604 | |
| $\gamma^h$ | 0.466 | |

Notes: *computed by $\kappa_1(1-\sigma), \kappa_2(1-\sigma)$. See text.

constant-elasticity specification:

$$\mathcal{P}_t = \alpha_t \times (S_t^1)^{-\eta} \tag{10}$$

where $\alpha_t$ is a market-specific constant, and $S_t^1$ is the total listening share to in-metro stations in market $t$. We further parameterize the log of $\alpha_t$ by $ln(\alpha_t) \equiv k_t\nu + \omega_t$, where $k_t$ is a vector of market characteristics, and $\omega_t$ is an additive error term. Taking logs, and replacing the model's predicted ad price $\mathcal{P}_t$ by its empirical counterpart, $p_t$ (computed from data as explained in section 2), we obtain the following estimation equation:

$$ln(p_t) = k_t\nu - \eta ln(S_t^1) + \omega_t \tag{11}$$

Estimation of this equation must take into account the endogeneity of the total in-metro share $S_t^1$: a high value for $\omega_t$ induces entry, which in turn increases this share. We instrument for this share using the market's population and its number of out-metro stations. Table 5 provides the results of estimating the model in (11) via 2SLS. Since the elasticity of demand is $-(1/\eta)$, the estimate of $\eta$ implies an elasticity of about (-2), a similar result to that in BW99. As can be expected, the ad price is positively correlated with higher metro income and education levels, implying that advertisers are willing to pay more for more affluent listeners.

Ideally, one would like to allow the ad price to vary not only across markets, but across listening formats as well. This would make sense since advertisers (and stations) are likely to internalize the fact that different formats target different consumer types. As discussed above, data limitations prohibit us from pursuing such an approach.

### 3.3 The entry game and estimation of fixed costs

The discussion of the listener's utility model (subsection 3.1) and of advertisers' demand for listeners (subsection 3.2) was conditioned on a given market structure, that is: given numbers of stations operating in each market-format-quality data cell. We now turn to describing the third and final component of our framework: an entry game in which this market structure is endogenously determined.

We assume that a large number of (ex-ante identical) potential entrants contemplate entry into each local market. They engage in a two-stage game:

1. Potential entrants simultaneously choose whether to enter the market as an in-metro station and, if so, in which format-quality category to operate. Entering stations incur fixed costs that are specific to the market-format-quality cell.

2. Entering stations produce listeners as described by the listening model, and sell them to advertisers at a price which is determined from the inverse demand curve for listeners.

Table 5: Advertiser's demand for listeners

| | OLS | IV |
|---|---|---|
| northeast | -0.0746 | -0.0739 |
| | (0.064) | (0.063) |
| midwest | 0.0835 | 0.0799 |
| | (0.061) | (0.059) |
| south | 0.0148 | 0.0132 |
| | (0.060) | (0.059) |
| income | 0.0567* | 0.0606** |
| | (0.030) | (0.029) |
| college | 0.167*** | 0.164*** |
| | (0.043) | (0.042) |
| black | -0.0231 | -0.0242 |
| | (0.021) | (0.020) |
| hisp | -0.0120 | -0.0124 |
| | (0.014) | (0.013) |
| $-\eta$ | -0.541*** | -0.510*** |
| | (0.062) | (0.072) |
| Constant | 4.492*** | 4.554*** |
| | (0.17) | (0.18) |
| Observations | 163 | 163 |
| R-squared | 0.52 | 0.52 |

Notes: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Notice that this is a static, complete-information game in which firms observe everything (including the realizations of taste shocks $\xi_{gt}$) before making their entry decisions. The only active decision modeled here is entry: once in the market, stations' market shares are determined by the listening equation, while the ad price charged to advertisers is determined from the inverse demand curve (10). The solution concept employed is complete information Nash equilibrium. Importantly, this is a "once-and-for-all" model in which stations make the correct decision of being in or out of the market (and, when in the market, in which format-quality to operate). The entry decisions determine market $t$'s structure, $N_t$. This could potentially be a 20-vector (ten formats times two quality levels), but since we only allow quality differentiation in two formats, this is a 12-vector instead.[16]

A key feature of many entry models is that uniqueness of equilibrium is not guaranteed. Intuitively, the market may have one equilibrium in which two stations operate in format A and a single station operates in format B, and another equilibrium in which these numbers are reversed. The non-uniqueness implies that fixed costs are only partially identified. We next explain how necessary equilibrium conditions provide such partially-identifying information.

For notational convenience, let $g$ index format-quality cells (rather than format cells as before). Our goal is to compute upper and lower bounds on $f_{gt}$, the fixed cost of operating a station in market $t$ and format-quality $g$. We do not specify an equilibrium selection mechanism. We do assume, however, that the observed market structure constitutes *some* equilibrium outcome of the game described above. As a consequence, the following necessary conditions must hold: (i) entrants' variable profits must not be lower than their fixed operating costs, and (ii) no additional entrant could garner variable profits in excess of the operating fixed costs.

Note that the variable profit predicted by the model for a station operating in format-quality $g$ and market $t$ is given by:

$$V_{gt}(N_t, d_t, \theta^0) = \mathcal{S}_{gt}(N_t, d_t, \theta_\ell^0) \times pop_t \times \mathcal{P}_t(N_t, d_t, \theta_d^0)$$

where $\theta = (\theta_\ell, \theta_d)$ is a vector containing all parameters from the listening model (section 3.1) and from the demand for listeners model (section 3.2), and $\theta^0$ denotes their true value. The expression $\mathcal{S}_{gt}(N_t, d_t, \theta_\ell^0)$ is the market share function which determines the share of a station in format-quality $g$, market $t$ as a function of the market structure vector $N_t$, and of the market level variables $d_t$ (such as income, education and other demographics). This function is given by the nested-logit market share formula. The market's population is given by $pop_t$. The market's ad price, predicted from equation (10), is denoted $\mathcal{P}_t(\cdot)$.

We now compute bounds on $f_{gt}$ using necessary equilibrium conditions. Condition (i) implies the following upper bound:

$$f_{gt} \leq s_{gt} \times pop_t \times p_t \equiv \overline{f}_{gt} \tag{12}$$

[16] While not reflected by this notation, recall that the market structure also includes the numbers of out-metro stations in all format-quality cells, which are taken to be fixed and exogenous.

where $s_{gt}$ is the *observed share* of an in-metro station operating in format-quality $g$, market $t$, and $p_t$ is the *observed ad price*. A lower bound can be computed from the necessary condition (ii):

$$f_{gt} \geq \mathcal{S}_{gt}(N_t + e^g, d_t, \theta_0) \times pop_t \times \mathcal{P}_t(N_t + e^g, d_t, \theta_0) \equiv \underline{f}_{gt} \tag{13}$$

where $e^g$ is a twelve-vector with zeros everywhere and the $g^{th}$ entry equal to 1. This necessary condition implies that an additional in-metro entrant into format-quality $g$ would not be able to recover its fixed costs of operation. Notice that computing this bound requires predictions for both the counterfactual market share enjoyed by such a potential entrant, $\mathcal{S}_{gt}(\cdot)$, and for the counterfactual market ad price $\mathcal{P}_t(\cdot)$. The latter prediction requires computation of the counterfactual total market share of in-metro stations, and application of the inverse demand curve in (10).[17]

Several practical issues arise in the computation of these bounds. The first issue concerns the computation of the market's observed ad price $p_t$ in (12). In principle, this price can be computed from data by dividing the market's observed total revenue by the observed total number of listeners to in-metro stations. In our framework, however, the predicted total listenership to in-metro stations does not match the data, a fact rationalized by measurement error. Since upper bounds are computed from observed revenues, whereas lower bounds are computed from counterfactual (predicted) revenues, this issue can create an artificial wedge between the two. To overcome this issue, we compute the "observed" ad price by dividing total revenue by the number of listeners to in-metro stations that is predicted by the model given the observed market structure, rather than by the number of listeners observed in the data. If the observed listening data are indeed subject to measurement error, this approach is appropriate, and allows for a consistent analysis.[18]

A second issue concerns cases where no stations are observed in the $(g, t)$ market-format-quality cell. In such cases, one cannot compute these bounds. Clearly, there is no information that provides an upper bound on fixed costs. Moreover, we do not have an estimate of $\delta_{gt}$, the mean utility level associated with such stations, so computation of a lower bound is also infeasible: one can compute the systematic portion of this utility level but, without an actual observation, no estimate is available for the unobserved taste shifter $\xi_{gt}$. As a consequence, we set $\underline{f}_{gt} = 0$, $\overline{f}_{gt} = \infty$ in this case.

Alternative strategies are possible for the "missing markets" problem: Pakes, Porter, Ho and Ishii (2015, PPHI), for example, consider an ordered choice ex-

[17] Our approach could be challenged if, in some markets, all available frequencies are being used by stations. In such cases, only the upper bounds on the fixed cost are valid, while the lower bounds, computed from the hypothetical revenues of potential additional entrants, would not be valid. The limited frequency issue is likely to be present in the densely-populated Northeast region. Excluding the Northeast yields near-identical excessive entry rates in all format/quality cells as those reported in Section 4 below. The limited frequencies issue, therefore, does not drive our main findings.

[18] We also note that, for simplicity, the "advertisers demand for listeners" model of Section 3.2 was estimated using the ad price computed directly from data, as described in Section 2, ignoring the measurement error issue.

ample and handle boundary outcomes (specifically, the case of no entry) via a symmetry assumption on a structural error in the fixed cost specification.[19] Our framework takes a different (and simplified) approach to the bounds issue: we avoid making assumptions on such structural errors altogether since we do not estimate the distribution (or features of the distribution) of fixed costs across markets, and instead base our welfare analysis on estimated bounds at the market-format-quality level. We therefore do not follow PPHI's approach here. Yet another possible approach would be to assign a value for $\xi_{gt}$ by considering the lowest value of these fitted errors in the sample (thereby taking into account that missing markets are systematically associated with negative taste shocks). For simplicity, we do not pursue this approach.

A third and final issue concerns the fact that some stations could not be assigned a quality level within our estimation procedure, as discussed in Section 3.2.2. In such cases, we compute the bounds under each possible assignment. Specifically, in such cases we do know that all stations offer the same quality level, but we do not know what it is. For example, if quality in the Mainstream format was undetermined (while that in the News/Talk format was determined), we compute the vectors of upper and lower bounds on fixed costs *for all formats* twice: once assuming that the observed Mainstream stations offer low quality, and the second time, assuming that those stations offer high quality.[20] In market-format cells where quality is undetermined in both formats, a total of four sets of fixed cost bounds are estimated. We return to this issue in Section 4, where we explain how we take it into account when computing bounds on optimal market structures.

**Descriptive evidence on the estimated fixed cost bounds.** Table 6 displays means of lower and upper bounds on fixed costs across the twelve format-quality cells. These means are computed over non-empty data cells since, as explained above, we do not compute meaningful bounds for empty cells. In markets where several sets (two or four) of bounds are computed due to the partial identification issue, all these sets are included in the computed mean (each getting the same weight as any other set of fixed cost estimates). The evidence in the table is consistent with some degree of heterogeneity in fixed costs across formats. As discussed below, there are some reasons why fixed cost distributions could differ across formats, which is why we do not want to impose that they are identical. In any event, what we end up using in our welfare analysis is not format-specific distributions but the market-format-quality specific bounds.

Another way to gain some descriptive insight into the estimated fixed costs is to use the cross-section of estimated intervals $[\underline{f}_{gt}, \overline{f}_{gt}]$ across markets to obtain upper and lower bounds on the Empirical Distribution Function (EDF) of format-quality $g$'s fixed costs across markets. Denoting the total number of

[19]We are grateful to a referee for pointing out this possibility.

[20]Interestingly, since the fixed effects $\psi$ adjust to perfectly offset the effect of shifting *all* stations in the relevant format from one quality level to another, the market shares and revenues of stations in other formats are actually not affected. Only fixed costs in the particular format in which quality was unassigned, therefore, can be affected by the quality assignment.

Table 6: Format-specific mean bounds on fixed costs (in $ million)

| Format | Mean lower bound | Mean upper bound |
|---|---|---|
| Mainstream low quality | 1.57 | 1.82 |
| Mainstream high quality | 3.39 | 4.22 |
| CHR | 2.82 | 4.07 |
| Country | 2.33 | 3.06 |
| Rock | 2.46 | 3.02 |
| Oldies | 2.16 | 3.19 |
| Religious | 0.69 | 0.87 |
| Urban | 3.43 | 4.21 |
| Spanish | 1.73 | 1.96 |
| News/Talk low quality | 1.05 | 1.16 |
| News/Talk high quality | 3.48 | 4.36 |
| Other | 1.77 | 2.15 |

Notes: the left column provides the mean (across the 163 formats) lower bound on fixed costs in each format, while the right column provides the mean upper bounds (see text).

markets by $N_m = 163$, we have, for each constant $c > 0$:

$$\frac{1}{N_m}\sum_{t=1}^{N_m} I\{\overline{f}_{gt} \leq c\} \leq \frac{1}{N_m}\sum_{t=1}^{N_m} I\{f_{gt} \leq c\} \leq \frac{1}{N_m}\sum_{t=1}^{N_m} I\{\underline{f}_{gt} \leq c\} \qquad (14)$$

That is, the EDF of the lower (upper) bounds on fixed costs in format $g$ is an upper (lower) bound on the EDF of these costs. If one assumes, in addition, that $f_{gt}$ are independent (over markets) draws from the true CDF of fixed costs in this format, then the EDFs of the bounds in (14) converge to lower and upper bounds on this true CDF as $N_m \to \infty$. Importantly, however, we will not rely on such an IID assumption, and will not use the bounds from (14) in our analysis. We shall merely present graphs of the estimated EDFs for illustrative purposes. What we do employ in our welfare analysis are the intervals $[\underline{f}_{gt}, \overline{f}_{gt}]$ for each market-format-quality cell.

Figure 1 shows bounds on the EDFs of fixed costs in the Mainstream and News/Talk formats, for high-quality stations and for low-quality stations separately. The costs of operating a high-quality station appear to be higher, in a distributional sense. This is not surprising: given that our estimate of $\gamma^q$ is positive, high-quality stations are predicted to enjoy higher mean utilities, and so higher revenues, than low quality stations. This implies that both the lower bound and the upper bound on fixed costs should be higher for high-quality stations. The somewhat-wide bounds on the EDFs are a consequence of the fact that bounds are non-informative in empty market-format-quality cells. The bounds *on fixed costs* in the non-empty cells are tight, and those are the ones used in the welfare analysis below.

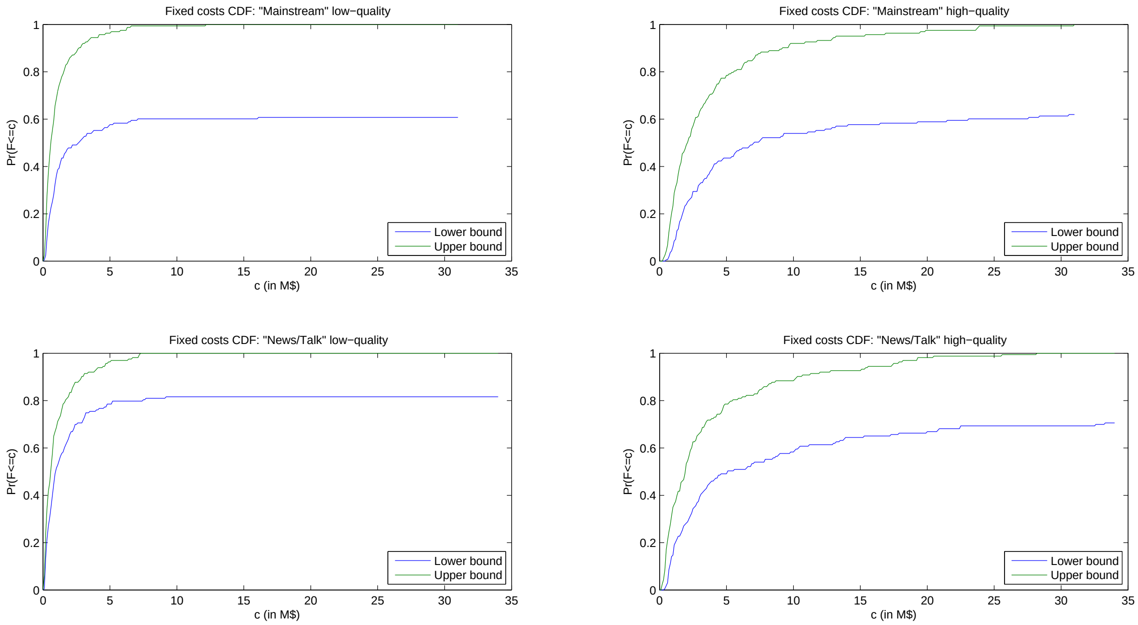


Figure 1: Estimated bounds on the CDF of fixed costs in Mainstream and News/Talk

**Discussion: sources of fixed costs.** Our model allows for different distributions of fixed costs for different formats. It is worth discussing what might drive such heterogeneity. Our approach assumes away marginal costs, as the non-rival nature of radio signals makes it seem inappropriate to model the cost of serving a marginal listener. All the costs of operating a radio station, therefore, are considered here to be fixed. These include the cost of equipment, employee salary, licensing fees and royalties paid for content.

While equipment costs need not, a-priori, diverge across formats, the cost of the content provided can vary substantially. A station that operates in a niche segment such as Jazz may need to physically possess thousands of records or music CDs, while a "big hits" station need not incur such costs.[21] Another example is that some formats (most notably, News/Talk) may hire radio "personalities" while other formats would spend mostly on music content. In light of the above, we allow the distributions of fixed costs to diverge across formats, leaving it to the data to inform us about such potential divergence.

# 4 Socially-optimal market structures

In this section, we use the estimated model to investigate what market structures would have been optimally chosen by a social planner who maximizes the joint surplus available to radio stations and advertisers, less the fixed costs of station operation. Similarly as in BW99 we cannot take into account listeners'

[21]See, for example http://www.ehow.com/how_2316008_calculate-startup-costs-radio-station.html.

surplus, since listeners do not pay for tuning in to a radio broadcast. As a consequence, it is not possible to evaluate their willingness to pay for radio broadcasting, or their surplus. It is, however, possible to gain some insight on this issue. In Section 6, we provide some suggestive evidence that the losses to listeners are not likely to outweigh the gains to market participants from station elimination, indicating that at least some station elimination is socially beneficial.

A market structure is a vector $N$, describing the numbers of in-metro stations in each of the format/quality cells. Let $G$ denote the dimension of this vector. In our case, it is equal to twelve, as explained above. Conditional on the market structure, total welfare is given by:

$$W(N) = pop \int_{0}^{S_1(N)} p(x)dx - \sum_{g=1}^{G} N_g \times f_g \tag{15}$$

where *pop* is market population, $S_1(N)$ is the total listening share to in-metro stations, $N_g$ is the $g^{th}$ component of $N$, and $f_g$ is the fixed cost associated with operating an in-metro station in format-quality $g$ in the given market.[22] Advertisers' inverse demand function is given by $p(\cdot)$.

The welfare measure in (15) merits discussion. Notice that its first term integrates under the advertiser's demand curve and captures the sum of stations' variable profits and advertisers' surplus. The second term subtracts the total fixed costs. Conceptually, this expression reflects the static nature of our model. In the static equilibrium, stations that operate in the market should be profitable, while additional stations should not. The empirical adaptation interprets a station's fixed cost as the cost of operating the station in a single year (rather than as a dynamic "entry cost.") For the station to be profitable, this cost must not exceed its annual revenue. Stations play this game every year (including our sample year: 2001), in a static fashion: that is, each year, they decide whether to operate in the market based on profitability considerations for that year, abstracting from dynamic considerations (e.g., this modeling approach abstracts from a possible effect of participation in the market in the current year on next year's costs or revenues). The corresponding static welfare expression sums over the annual benefits to stations and advertisers, and subtracts the annual fixed costs, and is consistent with the relevant theory of static entry games (e.g., see Mankiw and Whinston 1986).

Had fixed costs been point-identified, searching for an optimal market structure would have involved solving a $G$-dimensional discrete (in the sense that numbers of stations must be integers) problem in each market. The problem is further complicated, however, by the fact that we do not have a point estimate of the fixed costs $f_{gt}$ for stations in market $t$, format-quality $g$, but rather an estimated interval $[\underline{f}_{gt}, \overline{f}_{gt}]$. As a consequence of the partial identification of fixed costs, the optimal market structure is also partially-identified.

[22] Once again, it is understood that market shares also depend on the presence of out-metro stations.

To address this issue, we develop an algorithm to compute, in each market, a set of market structure vectors *that are not ruled out as socially-optimal*. That is, we compute a set of vectors that, given the set estimates of fixed costs and the point-estimated parameters of the listening model and the advertising demand model, cannot be unambiguously ruled out as maximizers of the welfare function in (15).[23]

A simple intuition for our approach is the following: a candidate market structure vector $N$ can be unambiguously ruled out as socially optimal if, for example, adding a station to some given format-quality cell $g$ contributes to the surplus term $pop \int_0^{S_1(N)} p(x)dx$ an amount that exceeds the upper bound $\overline{f}_g$. The market structure can also be ruled out as optimal if removing a station from some format-quality cell $g$ reduces the surplus term by an amount that falls short of the lower bound $\underline{f}_g$. The algorithm enumerates candidate vectors and considers such deviations from them, resulting in a final set of vectors that were not ruled out. This set must include the true optimal market structure. Taking maximum and minimum values for the numbers of stations in each format-quality cell over this "surviving" set delivers bounds on the optimal market structure, in each market, as the final output of this exercise. In principle, one could possibly obtain tighter bounds on the optimal market structure by considering more complicated "deviations" from each vector (e.g., considering whether welfare could be improved by adding two stations to some format and removing three stations from another). The results reported below, however, demonstrate that the bounds are informative enough to provide a clear picture regarding the rates of excessive entry in almost all format-quality cells.

The algorithm employs a strategy for reducing the space of candidate vectors that need to be evaluated as optimal (clearly such a strategy is needed for otherwise the space of such vectors would be unbounded). We leave the full description of this algorithm, along with additional technical and computational details, to Appendix B.[24]

A couple of additional aspects of our approach are worth noting. First, if a given market-format-quality cell has no observed stations in the observed sample, we fix the number of stations in that market-format to zero when computing the optimal market structure. Thus, we do not capture under-provision situations where the market outcome leads to zero stations in the cell, whereas the social planner would have chosen a positive number of such stations. We can, however, capture under-provision situations where, say, one station is observed, and the social planner would prefer to have two.

Yet another issue that must be tackled is that quality was not determined in some format-quality cells, as discussed in subsection 3.1.2 above. In those

[23] This is an analogue of the computation of a set of "potential equilibria" given a partially-identified model in Eizenberg (2014).

[24] One could potentially construct a confidence set for the surviving vector set and its implied bounds on the excessive entry rates via an expensive bootstrap exercise, taking into account the estimation error in the model's parameters. Since these parameters are estimated with very high precision, such confidence sets are likely to be tight.

cases, all we know is that all stations in the market-format pair offer identical quality; but we do not know which quality level it is. The assignment of this quality affects the analysis of optimal market structures (e.g., it affects the estimated bounds on fixed costs). We address this issue using the following strategy: in markets where quality was undetermined (in Mainstream, News/Talk, or both) we proceed by estimating fixed costs bounds and computing bounds on the optimal market structure under each *possible* quality assignment. For example, in a market where quality was unassigned in the Mainstream format, we perform this analysis twice: once assuming that all Mainstream stations offer low quality, and once assuming they all offer high quality. We then compute upper (respectively, lower) bounds on the optimal number of stations in each component by taking the element-by-element maximum (minimum) over the two sets of market structures that cannot be ruled out as optimal. We address similarly markets where quality was undetermined in the News/Talk format (again considering two possible scenarios and computing two sets of vectors that cannot be ruled out as optimal), and markets where quality was undetermined in both Mainstream and News/Talk (leading to four possible scenarios and to the computation of four sets of market-structure vectors).

Results from this analysis are discussed next. Table 7 displays findings for the eight formats in which we did not allow quality differentiation, while Table 8 considers the two formats in which such differentiation was allowed.[25] Beginning with Table 7, the most-left column reports the mean (over the 163 markets) observed number of stations in each of the eight formats in which quality differentiation was not allowed. The second and third columns report the mean (again over the 163 markets) lower and upper bounds on the optimal number of stations in each of these formats. For example, the mean upper bound on the optimal number of stations in the Country music format is 1.086. Finally, the two most-right columns report upper and lower bounds on the excessive entry rate. To explain how the latter are computed, consider the Country format. The upper bound on the excessive entry rate is computed by (2.104-0.969)/2.104=54%, whereas the lower bound is given by (2.104-1.086)/2.104=48%.

Examining the eight formats, the excessive entry rate is generally bounded between 50% and 60%. The CHR and Oldies formats display lower rates of excessive entry, closer to 20%, but this is hardly surprising as these are small formats, where the observed number of stations is typically at most one, leaving very limited scope for excessive entry at the outset.

Table 8 provides similar information for the Mainstream and News/Talk formats. One important difference is that, in these formats, even the observed number of stations is sometimes partially identified (recalling that, for example, in some markets, all we know is that all Mainstream stations offer the same quality level, but that level is not identified). Mean upper and lower

[25]To be clear, while the description of these results is separated into two tables, both tables refer to the same analysis in which the bounds on the entire optimal market structure, referring to optimal numbers of stations in each format-quality combination, were computed.

Table 7: Bounds on excessive entry rates: formats without quality differentiation

| | | Optimal market structures | | Excessive entry rates | |
|---|---|---|---|---|---|
| | Mean observed stations | Mean optimal lower bound | Mean optimal upper bound | Lower bound | Upper bound |
| CHR | 1.061 | 0.847 | 0.853 | 20% | 20% |
| Country | 2.104 | 0.969 | 1.086 | 48% | 54% |
| Rock | 2.331 | 0.982 | 1.147 | 51% | 58% |
| Oldies | 1.025 | 0.847 | 0.859 | 16% | 17% |
| Religious | 1.663 | 0.730 | 0.810 | 51% | 56% |
| Urban | 1.497 | 0.638 | 0.767 | 49% | 57% |
| Spanish | 1.344 | 0.479 | 0.571 | 58% | 64% |
| Other | 2.123 | 0.939 | 1.049 | 51% | 56% |

Notes: the most-left column reports the mean (over the 163 markets) observed number of stations in each of the eight formats in which quality differentiation was not allowed. The second and third columns report the mean (again over the 163 markets lower) and upper bounds on the optimal number of stations in each of these formats. For example, the mean upper bound on the optimal number of stations in the Country format is 1.05. Finally, the two most-right columns report upper and lower bounds on the excessive entry rate. To explain how the latter are computed, consider the Country format. The upper bound on the excessive entry rate is computed by (2.10-0.93)/2.10=56%, whereas the lower bound is given by (2.10-1.05)/2.10=50%.

bounds on the *observed* numbers of stations are presented in the first two rows. The next couple of rows provide mean upper and lower bounds on the optimal numbers of stations in the relevant format/quality cell. For example, the mean (across the 163 markets) lower and upper bounds on the numbers of high quality Mainstream stations are 0.11 and 1.02, respectively.

The bounds on both the observed and the optimal numbers of stations in these formats are wider relative to those obtained for the formats in which quality differentiation was not allowed. This stems entirely from the partial identification of quality issue. Just the same, we conclude using similar calculations as described for the other eight formats (see the description of Table 7a above) that excessive entry obtains for high quality stations in both the Mainstream and the News/Talk formats. In the Mainstream format, the excessive entry rate is bounded between 25% and 95%, while in the News/Talk format, it is bounded between 19% and 98%. No clear conclusions regarding excessive entry of low-quality stations in these formats are available, however, as the bounds on the excessive entry rate contain the value zero.

Table 8: Bounds on excessive entry rates: formats with quality differentiation

| | Low quality | | High quality | |
|---|---|---|---|---|
| | Mainstream | News/Talk | Mainstream | News/Talk |
| Mean lower bound, observed | 1.18 | 1.83 | 1.40 | 1.06 |
| Mean upper bound, observed | 1.95 | 2.02 | 2.17 | 1.25 |
| Mean lower bound, optimal | 0.10 | 0.15 | 0.11 | 0.03 |
| Mean upper bound, optimal | 2.42 | 2.32 | 1.06 | 0.85 |
| Excessive entry rate, l. bound | -105% | -27% | 25% | 19% |
| Excessive entry rate, u. bound | 95% | 93% | 95% | 98% |

Notes: the first and second rows provide lower and upper bounds on the mean (over the 163 markets) observed number of stations in the relevant format and quality level. The reason why bounds are obtained on observed numbers of stations is the partial identification of quality levels in some cases. For example, the mean lower bound on the number of low-quality Mainstream stations is 1.18, and the mean upper bound is 1.95. The third and fourth rows provide mean lower and upper bounds on the optimal numbers of such stations. Finally, bounds on excessive entry rates are computed similarly as described in the notes to Table 7 (see also the text).

To summarize, Tables 7 and 8 imply substantial excessive entry across the board, the one exception being low-quality Mainstream and News/Talk stations where one cannot rule out that the numbers of stations in the free entry equilibrium coincide with the optimal numbers. It is worth noting that we also performed a simplified analysis in which, rather than using the bounds $\underline{f}_{gt}, \overline{f}_{gt}$ as explained above, we used the middle of the estimated interval $[\underline{f}_{gt}, \overline{f}_{gt}]$. This results in a point-identified optimal market structure (except for cases where quality is only partially identified, where the optimal market structure is still partially identified). This analysis yielded very similar results, implying excessive entry rates of 50%-60% in most formats (including the case of low-quality Mainstream and News/Talk stations, for which our baseline analysis could not unambiguously determine that excessive entry prevails). In discussing simpler models below (Section 5) we make use of this simplified "mid-interval" strategy for computing optimal market structures.

**How often is quality misallocated in equilibrium?** Having computed the optimal market structure, we next wish to pay particular attention to the optimality of quality allocations in the free-entry equilibrium. We address this issue by asking the following question: beginning with the market structure observed in equilibrium, how often can welfare be improved by converting an observed low-quality station into a high-quality one? And, vice versa, how often can welfare be improved by converting an observed high-quality station into a low-quality one?

Notice that in the current analysis we allow the social planner to shift a single station's quality level, holding the format-quality choices of all other stations fixed as in the observed equilibrium. In particular, this exercise leaves the total number of operating stations unaltered. This is conceptually different from the analysis above, in which we allowed the social planner to change the entire market structure vector such that both the total number of stations, and their allocation across formats and qualities could be altered. That analysis focused on excessive entry rates, whereas the "local" quality analysis addresses a different aspect of the efficiency (or lack thereof) of the equilibrium outcome: namely, that products may be misallocated across vertical and horizontal dimensions conditional on their total number. By relaxing station symmetry, our framework allows us to address such possibilities.

In practice, let us consider market $t$ and format $g$, where $g$ is one of the two formats in which we allow quality differentiation. Let $f_{gt\ell}, f_{gth}$ correspond to the true fixed costs associated with operating a low quality station and a high quality station in this market-format cell, respectively. Suppose that this format has low-quality stations present in the observed equilibrium. We ask whether converting one such station into high-quality operation would increase social benefits (the sum of stations' revenue and advertisers' surplus) by more than the difference $\overline{f}_{gth} - \underline{f}_{gt\ell}$, where overlines and underlines correspond, as above, to upper and lower bounds, respectively. If this condition is met, underprovision of quality prevails in this case, in a local sense. Analogously, if high-quality stations are observed, we shall ask whether converting one of them to low quality would decrease social benefits by an amount smaller than the difference $\underline{f}_{gth} - \overline{f}_{gt\ell}$.[26]

A challenge arises in determining the welfare consequences of converting a low-quality station into a high-quality one: if no high quality stations are observed in the data, then no estimate of $\overline{f}_{gth}$ is available. This happens since upper bounds on fixed costs are generated from the observed revenue of stations in the relevant data cell. Importantly, however, since a low-quality station was observed, an estimate for $\psi_{gt}$, the market-format fixed effect, is available, and so it is possible to compute the mean-utility level of a hypothetical high-quality station as $\psi_{gt} + \gamma^h + \gamma^q$. With this mean-utility level at hand, it is possible to compute the lower bound $\underline{f}_{gth}$ from the hypothetical revenue of such a hypothetical entrant, conditional on the observed market structure.

Our approach in such cases is to set the value of the unknown $\overline{f}_{gth}$ equal to $\underline{f}_{gth} + \mu^h$, where $\mu^h$ is the maximum difference $\overline{f}_{gth} - \underline{f}_{gth}$ taken over all market-$t$, format-$g$ pairs in which high quality stations were observed (so that the computation of both $\overline{f}_{gth}$ and $\underline{f}_{gth}$ is possible). Similarly, when considering the conversion of a high quality station into a low quality one in a market-format cell where no low quality stations are observed, we use as our estimate of $\overline{f}_{gt\ell}$ the quantity $\underline{f}_{gt\ell} + \mu^\ell$ where $\mu^\ell$ is the maximum difference $\overline{f}_{gt\ell} - \underline{f}_{gt\ell}$ taken over all market-$t$, format-$g$ pairs with observed low-quality stations.

We perform this analysis in the two formats in which quality differentiation was allowed: Mainstream and News/Talk. Furthermore, markets in which

[26] Advertisers' surplus unambiguously increases when higher quality is offered, since that generates higher listenership and lower ad prices. The effect on stations' total revenue, in contrast, is ambiguous.

quality was undetermined in either the Mainstream or News/Talk formats where excluded from this exercise. The results of this exercise are quite telling: out of 90 markets with observed high-quality Mainstream stations, in 72 cases welfare can be unambiguously improved by converting one of those stations to low quality operation. In other words, overprovision of quality in the local sense occurs at a rate of 80%. An even higher rate, 94.9%, applies to the News/Talk format (74 out of 78 markets). On the other hand, there are no cases where a market has observed low-quality stations—in either format—and converting one of them to high quality would unambiguously improve welfare. Our analysis of local changes to quality offerings, therefore, reveals a highly-asymmetric pattern: over-provision of quality appears to be widespread, whereas under-provision is not encountered.

# 5 Comparison to simpler models

In this section we present simpler models that eliminate one or both of the dimensions of differentiation allowed for in our baseline model. Section 5.1 examines the implications of allowing horizontal (format) differentiation only, while Section 5.2 estimates a symmetric model *à la* BW99 in which no systematic station differentiation is allowed. The goal of these comparisons is twofold: to examine the robustness of our findings, as well as to shed light on the value of extending the simpler models, whose estimation is less demanding. Some further details on these models is provided in Appendix A.

## 5.1 A model with horizontal differentiation

This simplified version of the model assumes away quality differentiation and instead focuses on systematic format differentiation. We rely once again on the nested logit specification with the same eleven nests as before. Listener $i$'s utility from listening to station $j$, which belongs to format category $g$, in market $t$, is given by:

$$u_{i,j\in g,t} = \delta_{gt} + \nu_{igt}(\sigma) + (1-\sigma)\epsilon_{ijt}, \text{ with } \delta_{gt} = x_{gt}\beta + \xi_{gt} \tag{16}$$

where $x_{gt}$ is a vector of format and market characteristics, equivalent to the vector $d_{gt}$ from the baseline model. It includes the average income, the share of college educated, the shares of Black and Hispanic population, dummy variables for geographic regions and for format categories, and some intuitive interaction terms. The unobserved term $\xi_{gt}$ shifts the mean taste toward format $g$ in market $t$. The term $\nu_{igt}+\epsilon_{ijt}$ is, once again the idiosyncratic deviation from the mean utility.

Absent some details discussed in Appendix A, the nested logit specification then leads to a linear estimation equation for station $j$ operating in format $g$ in market $t$:

$$ln(s_{jt}) - ln(s_{0t}) = x_{gt}\beta + \sigma ln(s_{j/g,t}) + \xi_{gt} \tag{17}$$

where $s_{jt}$ is the market share of station $j$, $s_{j/g,t}$ is the share of this station as a fraction of the total listening share to format $g$ in market $t$, and $s_{0t}$ is the share of the outside option.

The endogeneity of $ln(s_{j/g,t})$ requires instrumental variables. To facilitate comparability across models, we use the same three instruments as in the baseline model: population, number of out-metro stations, and number of out-metro stations in the same format. All three are assumed to be exogenous, and they are likely to affect entry, and therefore the within-format market shares. Table 9 provides 2SLS estimation results. The patterns are qualitatively similar to those observed in the baseline model. Variables that have the strongest impact on a station's listenership are the dummy variables for in-metro status and for format categories, and the interactions of the format and the demographic effects. These effects are very precisely estimated. As expected, popular formats such as Mainstream or Rock have large estimated coefficients in this specification.

Table 9: The listening equation with horizontal differentiation only

| | | | | | |
|---|---|---|---|---|---|
| **Region** | northeast | 0.122*** | **Interactions** | hispXspan | 0.352*** |
| | | (0.042) | | | (0.036) |
| | midwest | 0.0974** | | blackXurban | 0.506*** |
| | | (0.041) | | | (0.050) |
| | south | -0.0506 | | southXreligious | 0.809*** |
| | | (0.041) | | | (0.095) |
| **Demographics** | black | -0.0681*** | | southXcountry | 0.316*** |
| | | (0.014) | | | (0.072) |
| | hisp | -0.0233** | **Substitution** | $\sigma$ | 0.519*** |
| | | (0.0097) | | | (0.063) |
| | income | -0.00258 | **In-metro** | | 0.639*** |
| | | (0.017) | | | (0.082) |
| | college | -0.0630** | **Constant** | | -5.325*** |
| | | (0.027) | | | (0.15) |
| **Format** | included | (not reported) | | | |
| Observations | 1919 | | | | |
| R-squared | 0.72 | | | | |

Notes: standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Also expected is the strong and significant effect of the interactions between the fraction of Black population and the Urban format, the interaction between the fraction of Hispanic population and the Spanish format, and the interactions of the South region dummy with the Religious and Country formats. The fraction of the market's population with college education is negatively related to listenership. Finally, and importantly, the substitution parameter $\sigma$ is estimated at 0.519, a similar value to that obtained in the baseline analysis.

We proceed by estimating fixed costs and computing optimal market structures in the same fashion as in the baseline analysis, where quality differentiation was allowed. The analysis is simpler in the sense that the equilibrium market structure is fully observed, whereas in the baseline analysis, the number of stations in certain cells was not always uniquely determined since the quality level of stations was unobserved (and in some cases only partially-identified by the model). We also simplify the computation of optimal market structures by using the middle of the estimated interval for fixed costs, rather than working with the bounds directly. As discussed above, this simplification was not found to have an important effect in the baseline model, supporting its use here.

The comparison of observed vs. optimal market structures given the simplified listening model is provided in Table 10. The table averages over the numbers of stations in each format in all 163 markets. Excessive entry, on average across markets, is apparent in all ten formats. Much like in the baseline analysis of Section 4, in most formats, an average reduction of about 50%-60% in the number of in-metro stations is optimal, and the least amount of excessive entry occurs in the CHR and Oldies formats, where the average optimal reductions are 20% and 14%, respectively. In total, the average market has 19.58 in-metro stations, whereas the optimal number of such stations is 50% lower: 9.79. The results of the simplified analysis in which only horizontal differentiation is allowed are, therefore, highly consistent with the findings from our baseline model in which unobserved vertical differentiation was also allowed for. This is apparent both in the estimated parameters of Table 9 (notably, in the similar value of the business-stealing parameter $\sigma$), and in the extent of excessive entry presented in Table 10. Our findings are, therefore, robust to the exclusion of a vertical differentiation dimension in the model.

Table 10: Optimal and observed market structures (horizontal differentiation only)

| Format | Observed | Optimal | % Excessive entry |
|---|---|---|---|
| Mainstream | 3.35 | 1.38 | 0.59 |
| CHR | 1.06 | 0.85 | 0.20 |
| Country | 2.10 | 1.05 | 0.50 |
| Rock | 2.33 | 1.09 | 0.53 |
| Oldies | 1.02 | 0.88 | 0.14 |
| Religious | 1.66 | 0.81 | 0.51 |
| Urban | 1.50 | 0.73 | 0.51 |
| Spanish | 1.34 | 0.60 | 0.56 |
| News/Talk | 3.08 | 1.35 | 0.56 |
| Other | 2.12 | 1.07 | 0.50 |
| **Total In-metro** | **19.58** | **9.79** | 0.50 |

Notes: The first column computes the mean (over the 163 markets) number of observed stations in each format, whereas the second column reports the mean optimal number of stations in each format. The third column computes the excessive entry rates similarly as in the baseline analysis. For instance, in the Mainstream format, the rate is (3.35-1.38)/3.35=0.59.

One may wonder whether this similarity stems from the fact that the baseline model allowed for vertical differentiation in two formats only. That does not appear to be the case: in yet another version of our model, not reported in this paper, we allowed for *observed* vertical differentiation measured using an index of observed station characteristics such as Antenna height, wattage and FM status. Stations were discretely classified into two quality levels such that high (low) quality stations were those with a quality index above (below) the median quality index. This model allowed for horizontal differentiation as well as for observed vertical differentiation *in all formats*, and yielded similar findings to the models presented above.[27]

### 5.2 A symmetric model following BW99

Both our baseline model and the simplified model presented in Section 5.1, in which only horizontal differentiation was allowed, indicated excessive entry rates of 50%-60% in most formats. In contrast, BW99's symmetric model yielded a higher excessive entry rate of about 74%. A natural question to ask is whether the introduction of systematic differentiation softens the excessive entry finding. Intuitively, this may be the case if the differentiated models allow stations to be more distant substitutes thus alleviating business stealing.

Such conclusions are, however, difficult to draw. One reason is that the BW99 results were obtained using a different dataset. While the data sources are similar, that paper was estimated using 1993 data, whereas the current paper uses 2001 data. Motivated by this observation, we estimate a symmetric model following the spirit of BW99 using the current dataset. We briefly describe this symmetric model and estimation results, referring the reader to the BW99 paper for additional details.

Following BW99, we apply the nested logit model once again, with two nests: one containing all commercial stations, and one containing the outside option only. We assume symmetry among all in-metro stations in market $t$, implying that they are characterized by the same mean-utility level, and, hence, by the same market share. Denote this common mean-utility level by $\delta_{1,t}$. Similarly, we assume that all out-metro stations are also identical, characterized by an identical mean-utility level of $\delta_{2,t}$. Each market is represented by two observations, pertaining to the typical in-metro station, and to the typical out-metro station.[28] We parameterize the mean-utility to depend on market-level characteristics $x_t$. We estimate the nested model by OLS, as well as by 2SLS using population and the number of out-metro stations as excluded instruments (the same instruments that were used in BW99). Further details are in Appendix A.

Table 11 provides results from this estimated model. As the table shows, both OLS and 2SLS yield very similar values for $\sigma$ of about 0.9. This value is somewhat higher than that obtained in BW99, which was about 0.8. Much

[27] Details are available from the authors upon request.

[28] In about 20 markets we have no out-metro stations. For those markets we have one observation only - a "representative" in-metro station. In total, our 163 markets generate 306 observations in this model.

like in BW99, therefore, this symmetric model is consistent with substantial business stealing. The signs of the other coefficients are also largely consistent with those reported in BW99.

Table 11: Symmetric model: listening equation estimates

| Utility coefficients | OLS | 2SLS |
|---|---|---|
| northeast | 0.0944*** | 0.0948*** |
| | (0.0228) | (0.0239) |
| midwest | 0.0446** | 0.0450** |
| | (0.0208) | (0.0220) |
| south | 0.0492*** | 0.0493*** |
| | (0.0186) | (0.0187) |
| income | 0.0102 | 0.0101 |
| | (0.0103) | (0.0106) |
| college | -0.0485*** | -0.0484*** |
| | (0.0150) | (0.0151) |
| home | 0.135*** | 0.138*** |
| | (0.0241) | (0.0447) |
| $\sigma$ | 0.910*** | 0.908*** |
| | (0.0141) | (0.0294) |
| Constant | -2.341*** | -2.348*** |
| | (0.0723) | (0.131) |
| Observations | 306 | 306 |
| R-squared | 0.984 | 0.984 |

Notes: 2SLS results utilize the market population and its number of out-metro stations as instruments, see text.

We continue to follow BW99 by assuming that fixed costs are log-normal, with the per-station fixed cost in market $t$ taking the following form: $log(F_t) = x_t\mu + \lambda\nu_t$, where $\nu_t|x \sim N(0,1)$ and is independent across markets, $x_t$ is a vector of market characteristics, and $\lambda$ and $\mu$ are parameters to be estimated. This specification allows the estimation of these parameters using a Maximum Likelihood approach. This stems from the fact that the number of in-metro stations observed in the market is uniquely determined in the equilibrium of the symmetric model. This stands in contrast to the nonsymmetric models presented above, in which uniqueness of the market structure is not guaranteed, and, as a consequence, only partial identification of fixed costs is available.

Referring the reader to BW99 for additional details on the construction of the likelihood function, we report estimation results in Table 12. The main qualitative difference from the BW99 results is that we get insignificant regional effects, but other than that, the pattern of the coefficients is the same (in particular, we obtain positive effects for college, income and population).

Finally, we simulate the optimal market structure (i.e., the optimal number

Table 12: Symmetric model: estimated fixed cost parameters

| Fixed cost coefficients | COEFF | SE |
|---|---|---|
| Constant | -1.6210 | 0.2673 |
| northeast | -0.1346 | 0.1290 |
| midwest | -0.0538 | 0.1232 |
| south | 0.1097 | 0.1123 |
| income | 0.3936 | 0.0676 |
| college | 0.0238 | 0.1079 |
| pop | 0.2959 | 0.0137 |
| $\lambda$ | 0.4910 | 0.0304 |
| Observations | 163 | |

Notes: standard errors are not corrected for the error in the preliminary estimation of $\hat{\beta}, \hat{\eta}$.

of in-metro stations) in each of our 163 markets. Again, we refer the reader to BW99 for additional details. The results are striking: over the 163 markets, an average excessive entry rate of 74.13% is obtained—a near-identical finding to that reported in BW99. Our results therefore suggest that the symmetric model consistently yields higher excessive entry rates than those implied by the non-symmetric models.

This result, however, appears difficult to generalize to other settings, since it is difficult to characterize the exact mechanism that yields this discrepancy between the symmetric and non-symmetric models. These models differ from each other along several dimensions: first, the symmetric model is point-identified and is estimated given a parametric assumption on the distribution of fixed costs, whereas the non-symmetric models imposed no such assumptions. Second, the non-symmetric model utilized a third instrumental variable—the number of out-metro stations in the same format—which cannot be utilized in the symmetric model. Third, and perhaps most importantly, the symmetric model yields a high value for $\sigma$, 0.9, relative to much lower values of about 0.51-0.57 in our non-symmetric models. This higher value of $\sigma$ naturally results in higher excessive entry rates.

The major takeaway from the comparison to the symmetric model is, therefore, that the estimation of non-symmetric models, while technically more demanding, can be important, as they yield quantitatively and qualitatively different results relative to a simple symmetric model.

## 6 Robustness

In this section we address the robustness of our analysis to a sample selection issue, to certain symmetry assumptions, and to the issue of joint station ownership. In addition, we address a limitation of the analysis: the fact that the social planner maximizes the welfare of market participants, failing to consider the unmeasured benefits to listeners. We provide some suggestive evidence that the finding of excessive entry is likely to hold, even when we take listeners' welfare into account.

### 6.1 Robustness to sample selection and to symmetry assumptions

**Sample selection.** As discussed in Section 2, sample selection is a relevant concern with respect to the Urban, Spanish and Religious formats. One may suspect that we only observe such stations in markets where the unobserved taste for such broadcasting is sufficiently strong, leading to an upward bias in the estimates of the coefficients on the relevant format dummy variables.[29] Addressing such selection problems in the context of a product-choice model with complete information is quite complicated, since the selection mechanism depends on the error terms of all products (rather than on the error of the specific product), and is not uniquely determined in equilibrium. Following traditional selection-correction mechanisms, such as Heckman (1976, 1979), is infeasible.

Eizenberg (2014) offers a partial-identification strategy to formally address the product selection issue in a study of the PC industry. In contrast, we do not formally address the selection issue within the estimation procedure. Instead, we offer several analyses of the robustness of our results to the selection issue. The details of these analyses are provided in Appendix C.

First, we demonstrate that the presence of stations in the Urban and Spanish formats is very strongly driven by observables for which we control—namely, the demographic makeup of the market's population. Second, we re-estimate our horizontal-differentiation model using a sub-sample of markets that are predicted to have stations in the relevant format (e.g., Urban) with a very high probability. This subsample should be viewed as selection-free to a large extent. Our results indicate, reassuringly, that the estimates do not change in a way that is consistent with sample selection bias (though the evidence is consistent with some degree of selection bias in the case of the Spanish format).

**Symmetry assumptions.** Our baseline model assumes that stations within market-format-quality cells are symmetric, in the sense of having the same mean-utility level and predicted market shares. Observed deviations of the empirical market shares from this assumption are formally attributed to measurement error in the listening data. Our simplified horizontal-differentiation model, presented in Section 5.1, imposes symmetry within market-format cells and computes market shares in a way that is consistent with this hypothesis. The symmetry assumptions are not dictated by constraints associated with estimating the listening equation. Rather, they are motivated by the need

[29] While our baseline listening model (section 3.1) controls for unobserved market-format tastes via fixed effects, this does not solve the problem of selection on unobservables to the extent that no station is observed within the relevant market-format cell.

to have a well-defined mean-utility level for each market-format (or, market-format-quality) in the estimation of foxed costs, and in the welfare analysis.

To demonstrate this point, Appendix C reports estimates from a non-symmetric version of our horizontal-differentiation model, i.e., we estimate a model which is identical to that presented in Section 5.1, except that symmetry within the market-format is no longer imposed. This model allows observed station characteristics (Antenna height, wattage, FM) to differentiate such stations. As shown there, the estimation results are qualitatively similar to those of the symmetric model, suggesting that the analysis of the listening equation is not sensitive to the symmetry assumptions.

### 6.2 Listeners' welfare

**Listener welfare.** The analysis thus far has ignored the positive externalities conferred upon listeners from broadcasting. Simply put, the social planner's elimination of stations reduces total listening, and hence listener's utility. Measuring the lost listener surplus is difficult on account of the radio signal being non-rival and non-excludable. At the same time, observing some benchmark figures can provide some idea about the lost listener surplus. We therefore perform several exercises which, for simplicity, use our horizontal-differentiation model (Section 5.1).

The first exercise we perform works as follows: the station elimination prescribed by the social planner discussed above leads to a reduction in the total listening share to in-metro stations. Multiplying the lost shares by the relevant market populations, and summing over all 163 markets, we obtain that a total of 6.02 million listeners are "lost" to the radio industry.[30] To offset the welfare gains to advertisers and stations, totalling \$1.8 billion, the average "lost" listener would need to be willing to pay at least \$299 for a year's worth of radio listenership. While learning about listeners' willingness to pay for a free product is very difficult, we may derive a useful benchmark from subscription fees to Satellite radio.

A monthly subscription to XM Sirius's most basic satellite radio services cost \$14.49 in August 2013, translating into an annual subscription cost of \$173.8.[31] We view this amount as an upper bound on the true willingness to pay for the terrestrial radio broadcasting that we analyze in this paper, given that satellite radio is a premium product. If that is the case, we may conclude that the willingness to pay is much lower than \$299. While this is clearly a very rough calculation, it does suggest that the lost listener welfare resulting from the station elimination prescribed by our social planner may not

[30] A subtle issue is that eliminating in-metro stations in one market also eliminates them as out-metro stations in another market, and our count of lost listeners does not take that into account. Adjusting the analysis to account for this issue would be quite demanding as it requires the mapping of each in-metro station in each market to all other markets in which it is an out-metro station. Since our goal is to provide a back-of-the-envelope calculation only, and since listenership to out-metro stations is systematically lower than listenership to in-metro stations to begin with, we ignore this cross-market externality in our analysis.

[31] http://www.sirius.com/ourmostpopularpackages-sirius.

be large enough to offset the combined gains for market participants (stations and advertisers).

As an alternative approach, we consider an exercise that involves listener's expected utility, as defined by the nested logit formulae. The benefit of this approach is that expected utility is calculated from an "ex-ante" perspective (i.e. prior to the realization of the listener-specific utility shocks), and takes into account some gains from variety which would be clearly reduced if the social planner were to reduce the number of active stations in the market. Summing over all 163 markets, we find that the total loss of expected utility of all listeners in all markets is given by 6.56 million "utility units."[32] If this lost listeners' utility is to exactly offset a $1.8 billion gains to the market participants (again, stations and advertisers), each utility unit must, therefore, be worth $274.[33]

With this conversion rate at hand, we can try to evaluate its implications for listeners' willingness to pay. For concreteness, let us focus on the New York City radio market. Beginning with the observed set of stations, let us contemplate an elimination of all four stations in the Rock format. We can calculate the impact of this elimination on the expected utility of the representative member of NYC's population (twelve years of age and above), and transform it to monetary terms using the conversion rate computed above. Each such person would incur a loss of 0.013 utility units, or, $3.62. A similar calculation would show that the monetary surplus lost as a consequence of eliminating all the market's stations would be $44.12.

What we learn from this exercise is that, if listeners' willingness-to-pay for radio is high enough to offset the welfare gains from station elimination, it would have to be true that each person living in NYC, prior to drawing her listener-specific utility shocks, would have to be willing to pay about $44 per year to be able to access the radio market. Given that the share of radio listenership in NYC is lower than 15%, this may seem like a large number. While it is hard to draw strong conclusions from this exercise, it again provides some suggestive evidence that listener's willingness to pay may not be high enough to offset the welfare gains from the station elimination prescribed by our social planner.

Another way to gain perspective of the issue is to use the conversion rate computed above to compute the expected surplus loss to listeners if one were to remove a single Rock station from the NYC market. This loss amounts to $5.3 million, which falls short of the combined gains to advertisers and stations, amounting to $6.4 million. In other words, even if listeners' willingness to pay is high enough to offset the gains to market participants from the massive stations elimination prescribed by our social planner, it would still be true that removing one Rock station would be optimal, i.e., the lost listener surplus from this modest station elimination would not be large enough to offset the gains.

[32] Where expected utility is given by $\log\left(\sum_g exp(I_g)\right)$ with $I_g$ being nest $g$'s "inclusive value."

[33] The same caveat about ignored cross-market externalities involving out-metro stations, discussed with respect to the previous analysis above, continues to apply to this analysis as well.

To sum, we view the above analysis as providing some perspective on the issue of the negative externalities on listeners' welfare that would result from station elimination. Overall, we view this evidence as consistent with there being at least some degree of excessive entry, even if one would take into account the externalities on listeners.

### 6.3 Joint ownership

An explicit treatment of multiproduct firms would introduce a number of interesting additional issues that we do not address in the paper. Importantly, though, as we explain below, our current approach is robust to the existence of jointly owned stations under the simplifying assumption that those firms differ only in their ownership structure but not in their quality or fixed costs (conditional on format-quality cell). Thus, our model is consistent with the *strategic* concerns of multi-product firms, although not with cost or demand-side "synergies" of multi-product ownership. We leave these synergies to future research.

It is easy to see that our bounds are still valid under multi-product ownership. Formally, consider a model of endogenous joint ownership with an infinite (or sufficiently large) number of *ex ante* identical potential entrepreneurs. Each potential firm can enter into any combination of the different format-quality cells in a given metro area and can operate any number of stations in a given cell. Assume that the profits of a station depend only on the number of stations in each cell, and not directly on the joint ownership structure of the station (this last assumption ruling out synergies of joint ownership.) As in our main model, the entry equilibrium is pure-strategy complete information Nash. Typically, in any given market there will be many equilibria, often involving different degrees of joint ownership.

In any equilibrium of the model with multi-station firms, it is easy to see that the bounding expression in (12) still holds. That is, the fact that all stations in format-quality cell $g$ are profitable in equilibrium still requires

$$f_{gt} \leq s_{gt} \times pop_t \times p_t \equiv \overline{f}_{gt}$$

and so the previous upper bound on fixed cost is still valid. Similarly, the lower bound in (13) still holds. That is, we must have:

$$f_{gt} \geq \mathcal{S}_{gt}(N_t + e^g, d_t, \theta_0) \times pop_t \times \mathcal{P}_t(N_t + e^g, d_t, \theta_0) \equiv \underline{f}_{gt}$$

or else some singly-owned station could choose to enter the format-quality cell, increase the number of stations to $N_t + e^g$ and earn positive profits. Thus, equilibrium still implies that the lower bound condition holds. Note also that our social-planner counterfactual does take account of business stealing and so it is still correct if we have correct bounds on fixed costs.

Note that joint ownership can imply *further* bounds that we do not use in the analysis. For example, a joint owner should consider the implications of entry in a given cell for the profits of its stations in other cells. These bounds might be useful in exploring the possibility that joint ownership lowers costs, perhaps including lower costs of high quality. However, addressing such synergies would require a substantial departure from our econometric techniques: specifically, one would need to estimate a parametric fixed cost function and allow joint ownership to shift that function. This stands in contrast to our approach that abstracts from such parametric assumptions. These more complicated methods are outside the spirit of the current paper and so we leave the question of joint-ownership synergies to future research. For the present work, we rely on the fact that our method is consistent with joint ownership (absent synergies).

## 7 Concluding remarks

The goal of this paper is to introduce horizontal and vertical differentiation into the analysis of excessive entry in local radio markets. Introducing such systematic heterogeneity creates three main challenges: first, we deal with the non-uniqueness of equilibrium and the resulting partial identification of fixed costs. Second, we deal with estimating a model that allows for quality as an unobserved station characteristic. Finally, given our set-estimated model, we address the conceptual challenge of computing an identified set of market structures that are not ruled out as socially optimal.

Notwithstanding these challenges, the results indicate that allowing for discrete station differentiation is important. It appears to soften the excessive entry finding to some extent, placing it at 50%-60% in most formats, compared to 74% in a specification that follows BW99 in that it does not allow for systematic station differentiation. The optimal elimination rates are quite robust to whether we allow for both horizontal and vertical differentiation, or horizontal differentiation only. Considering *local* changes to the observed equilibria, there is a very high incidence of quality overprovision: welfare can be improved by converting a high quality station into a low quality one.

In sum, extending traditional entry models to allow for horizontal and vertical differentiation can create a rich framework in which various questions concerning the divergence of free-entry equilibria and optimal market structures can be addressed. Such an agenda is particularly attractive given that theoretical analyses of such questions often provide ambiguous predictions that depend on specific parameter values. This motivates empirical work that attempts to estimate such parameters (e.g., the magnitude of fixed costs).

## References


[1] Andrews, D.W.K., and P. Jia Barwick (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure", forthcoming, *Econometrica*

[2] Berry, S.T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25(2): 242-262

[3] Berry, S., Levinsohn, J., and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4): 841-90

[4] Berry, S.T., and E. Tamer (2006): "Identification in Models of Oligopoly Entry," in *Advances in Economics and Econometrics: Theory and Applications*, Ninth World Congress, vol. 2, R. Blundell, W.K. Newey and T. Persson, eds., Cambridge University Press

[5] Berry, S.T., and J. Waldfogel (1999): "Free Entry and Social Inefficiency in Radio Broadcasting," *RAND Journal of Economics*, 30(3): 397-420

[6] Berry, S.T., and J. Waldfogel (2001): "Do Mergers Increase Product Variety? Evidence From Radio Broadcasting," *The Quarterly Journal of Economics* 116, 1009-1025

[7] Bresnahan, T.F., and P.C. Reiss. (1991): "Entry and Competition in Concentrated Markets," *The Journal of Political Economy*, 99(5): 977-1009

[8] Cardell, N.S. (1997): "Variance Components Structures for the Extreme Value and Logistic Distributions with Application to Models of Heterogeneity," *Econometric Theory*, (13): 185-213

[9] Chernozhukov, V., Hong, H. and Tamer, E. (2007), Estimation and Confidence Regions for Parameter Sets in Econometric Models. *Econometrica*, 75: 1243-1284

[10] Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77(6): 1791-1828

[11] Crawford, G.S., and A. Yurukoglu: "The Welfare Effects of Bundling in Multichannel Television ," Accepted at *American Economic Review*

[12] Duncan, J.H.: "American Radio," Spring 2001 edition (compiled and edited by Tony Sanders), Cincinnati, Ohio: Duncan's American Radio, L.L.C.

[13] Duncan, J.H.: "Duncan's Radio Market Guide," 2001 and 2002 editions, Cincinnati, Ohio: Duncan's American Radio, L.L.C.

[14] Eizenberg, A. (2014): "Upstream Innovation and Product Variety in the U.S. Home PC Market," *The Review of Economic Studies*, 81, 1003-1045

[15] Heckman, J.J., and S. Navarro (2007): "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics*, 136(2): 341-396

[16] Ho, K.E. (2009): "Insurer-Provider Networks in the Medical Care Market," *American Economic Review*, 99(1): 393-430

[17] Ishii, J. (2008): "Interconnection Pricing, Compatibility, and Investment in Network Industries: ATM Networks in the Banking Industry," *mimeo*

[18] Jeziorski, P. (2012): "Effects of Mergers in Two-sided Markets: Examination of the U.S. Radio Industry," forthcoming, *American Economic Journal: Microeconomics*

[19] Jeziorski, P. (2013a): "Estimation of cost efficiencies from mergers: Application to U.S. radio," forthcoming, *RAND Journal of Economics*

[20] Jeziorski, P. (2013b): "Empirical Model of Dynamic Merger Enforcement - Choosing Ownership Caps in U.S. Radio," *mimeo*

[21] Mankiw, G.N., and M.D. Whinston (1986): "Free Entry and Social Inefficiency," *RAND Journal of Economics*, 17(1): 48-58

[22] Maskin, E., Riley, J. (1984), "Monopoly with incomplete information," *Rand Journal of Economics*, 15, 171-96

[23] Mazzeo, M.J. (2002), "Product Choice and Oligopoly Market Structure," *Rand Journal of Economics*, 33(2), 221-42

[24] Mussa, M., Rosen, S. (1978), "Monopoly and product quality," *Journal of Economic Theory*, 18, 301-17

[25] 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

[26] Newey, W..K., and D. McFadden (1986): "Large Sample Estimation and Hypothesis Testing," Handbook of Econometrics, in: R. F. Engle & D. McFadden (ed.), Handbook of Econometrics, edition 1, volume 4, chapter 36, pages 2111-2245

[27] Rochet, J.C., and L. Stole (2002): "Nonlinear Pricing with Random Participation." *Review of Economic Studies*, 69(1): 277-311

[28] Rogers, R.P., and J.R. Woodbury (1996): "Market Structure, Program Diversity, and Radio Audience Size," *Contemporary Economic Policy*, 14(1): 81-91

[29] Smith, H., and C. O'Gorman (2008): "Efficiency Gain from Ownership Deregulation: Estimates from the Radio Industry," Discussion paper no. 385, Department of Economics, Oxford University

[30] Spence, M. (1976): "Product Selection, Fixed Costs, and Monopolistic Competition," *The Review of Economic Studies*, 43(2): 217-235

[31] Steiner, P.O. (1952): "Program Patterns and Preferences, and the Workability of Competition in Radio Broadcasting," *Quarterly Journal of Economics*, 66: 194-223

[32] Sweeting, A. (2010): "The Effects of Mergers on Product Positioning: Evidence from the Music Radio Industry," *RAND Journal of Economics*, 41(2)

[33] Sweeting, A. (2013): "Dynamic Product Positioning in Differentiated Product Industries: The Effect of Fees for Musical Performance Rights on the Commercial Radio Industry," *Econometrica*, 81(5): 1763-1803

[34] Pakes, A., J. Porter, K. Ho, and J. Ishii (2015): "Moment Inequalities and their Application," *Econometrica* 83(1):315-334.

# Appendices

## A Details on Simpler Models

This appendix provides some further details on the simpler models of section 5.

In the model of section 5.1 with only horizontal differentiation, the "mean utility" component, $\delta_{gt}$, is restricted to be identical within a given market-format data cell. This restriction is not dictated by estimation considerations, and we report a robustness check in which this symmetry assumption is relaxed. Rather, much like in our baseline model (where symmetric mean utilities were assumed within market-format-quality cells), this assumption simplifies counterfactual analyses in which stations enter a particular format, since it allows one to compute a well-defined mean utility for any such station. In practice, the within-format symmetry is modified to account for an important feature of the data: stations' "in-metro" vs. "out-metro" status. To account for this, we assume symmetry among all in-metro stations in the format, as well as among all the out-metro stations in the format. We therefore have two distinct mean utility levels within the market-format: $\delta_{gt,in-metro}$ and $\delta_{gt,out-metro}$. We also include an in-metro (or, "home") dummy variable in the $x$ covariate vector. For simplicity, the notation above (as well as in other parts of the paper) does not reflect the distinction between in-metro and out-metro stations. While mean-utilities are symmetric, individual stations within the format are still allowed to bring unique benefits via the $\epsilon_{ijt}$ term.

The within-nest symmetry in mean-utility levels implies that stations of a given format are predicted to garner identical market shares. Unlike the baseline model, here we wish to rely on a simpler estimation procedure, and therefore impose that stations within the market-format cell have identical shares (rather than attributing non-identical shares to measurement error). This means that $s_{jt}$ should be calculated as $S_{gt}/N_{gt}$, where $S_{gt}$ and $N_{gt}$ are the observed total market share and total number of stations in the market-format cell, respectively. Similarly, the fraction $s_{j/g,t}$ should equal $(1/N_{gt})$. In practice, however, the differentiation of in-metro vs. out-metro stations leads to slightly different calculations: for example, we obtain the share of a typical in-metro station in format $g$ by dividing the total observed share of *in-metro* stations in format $g$ by the total number of such in-metro stations.

The symmetry assumption further implies that we only retain (at most) two observations for each market-format data cell: an observation pertaining to the typical in-metro station in this format-market, and an observation pertaining to the typical out-metro station. Since it is possible that there are no in-metro stations, or out-metro stations, in the format-market cell, the number of observations pertaining to this format-market may in fact be zero, one or two.

In the symmetric model of section 5.2, the representative observations for in-metro and out-metro stations in market $t$ takes the following form:

$$ln(\frac{S_{1t}}{N_{1t}}) - ln(S_{0t}) = x_{1t}\beta + \sigma ln(\frac{S_{1t}/N_{1t}}{S_{1t}+S_{2t}}) + \xi_{1t}$$
$$ln(\frac{S_{2t}}{N_{2t}}) - ln(S_{0t}) = x_{2t}\beta + \sigma ln(\frac{S_{2t}/N_{2t}}{S_{1t}+S_{2t}}) + \xi_{2t}$$

where $S_{1t}$ and $S_{2t}$ are total listening shares to in-metro and out-metro stations, respectively, and $N_{1t}$ and $N_{2t}$ are the numbers of in-metro and out-metro stations, respectively. The share of the outside option is $S_{0t}$, and $x_{1t}, x_{2t}$ are market-level characteristics vectors, the difference between them being that $x_{1t}$ also contains an in-metro dummy variable. The representative observations for in-metro and out-metro stations in market $t$ takes the following form:

$$ln(\frac{S_{1t}}{N_{1t}}) - ln(S_{0t}) = x_{1t}\beta + \sigma ln(\frac{S_{1t}/N_{1t}}{S_{1t}+S_{2t}}) + \xi_{1t}$$
$$ln(\frac{S_{2t}}{N_{2t}}) - ln(S_{0t}) = x_{2t}\beta + \sigma ln(\frac{S_{2t}/N_{2t}}{S_{1t}+S_{2t}}) + \xi_{2t}$$

where $S_{1t}$ and $S_{2t}$ are total listening shares to in-metro and out-metro stations, respectively, and $N_{1t}$ and $N_{2t}$ are the numbers of in-metro and out-metro stations, respectively. The share of the outside option is $S_{0t}$, and $x_{1t}, x_{2t}$ are market-level characteristics vectors, the difference between them being the value of an in-metro dummy variable. The specification in BW99 is very similar, but not entirely identical to that employed here.

# B Computing optimal market structures

Introducing the algorithm requires some notation. Define $W^{+g}(N)$ as the change in welfare resulting from adding an in-metro station to format-quality $g$, $1 \leq j \leq G$ (where $G$ stands for the total number of format-quality cells, in our case 12: we have eight formats in which quality differentiation is not allowed, plus four format-quality combinations in the Mainstream and News/Talk formats), compared with a benchmark market structure $N$:

$$W^{+g}(N) = pop \times \int_{S_1(N)}^{S_1(N+e_g)} p(x)dx - f_g$$

where $e_g$ is a G-vector with all entries set to zero except the $g^{th}$ entry which is set to 1. Let $f$ be a known $G$-vector of fixed costs for the market's format-quality cells. In practice, we obtain an estimated interval for the fixed costs of operating in format-quality $g$, $[\underline{f}_g, \overline{f}_g]$. The change in welfare from adding an in-metro station to format-quality $g$ can, therefore, be bounded from above and from below by defining:

$$\overline{W}^{+g}(N) = pop \times \int_{S_1(N)}^{S_1(N+e_g)} p(x)dx - \underline{f}_g, \; \underline{W}^{+g}(N) = pop \times \int_{S_1(N)}^{S_1(N+e_g)} p(x)dx - \overline{f}_g$$

One can similarly obtain upper and lower bounds on the welfare gain from *removing* an in-metro station from format-quality $g$:

$$\overline{W}^{-g}(N) = -pop \times \int_{S_1(N-e_g)}^{S_1(N)} p(x)dx + \overline{f}_g, \; \underline{W}^{-g}(N) = -pop \times \int_{S_1(N-e_g)}^{S_1(N)} p(x)dx + \underline{f}_g$$

Finally, denote the set of market structure vectors that cannot be rule out as socially-optimal by $\mathcal{N}^o$. The following algorithm calculates this set in a given market:

1. Let $\underline{N}_\ell$ be a lower bound on the optimal number of in-metro stations in format-quality $1 \leq \ell \leq G$ (initially set $\underline{N}_\ell = 0 \; \forall \ell$). Fix format-quality $g$, $1 \leq g \leq G$. We obtain an integer, $\overline{N}_g$, interpreted as an upper bound on the optimal number of in-metro stations in format-quality $g$, as follows: we set $\overline{N}_g = 0$ if the following condition is met:

$$\Delta\overline{W}^{+g}(\underline{N}_1, \underline{N}_2, ..., \underline{N}_{g-1}, 0, \underline{N}_{g+1}, ..., \underline{N}_G) < 0 \tag{18}$$

   Otherwise, set $\overline{N}_g = \tilde{N}_g$ where $\tilde{N}_g$ is the smallest positive integer that satisfies:[34]

$$\overline{W}^{+g}(\underline{N}_1, \underline{N}_2, ..., \underline{N}_{g-1}, \tilde{N}_g - 1, \underline{N}_{g+1}, ..., \underline{N}_G) \geq 0 \tag{19}$$

$$\overline{W}^{+g}(\underline{N}_1, \underline{N}_2, ..., \underline{N}_{g-1}, \tilde{N}_g, \underline{N}_{g+1}, ..., \underline{N}_G) < 0 \tag{20}$$

   We Repeat the above for $g = 1, ..., G$ to obtain a vector of upper bounds, $\overline{N} = (\overline{N}_1, ..., \overline{N}_G)$

2. Again fix a format $g$. We compute a new lower bound $\underline{N}_g$ as follows: we set $\underline{N}_g = 0$ if the following holds:

$$\underline{W}^{+g}(\overline{N}_1, \overline{N}_2, ..., \overline{N}_{g-1}, 0, \overline{N}_{g+1}, ..., \overline{N}_G) < 0 \tag{21}$$

   Otherwise, we set $\underline{N}_g = \tilde{N}_g$ where $\tilde{N}_g$ is the smallest positive integer that satisfies:

$$\underline{W}^{+g}(\overline{N}_1, \overline{N}_2, ..., \overline{N}_{g-1}, \hat{N}_g - 1, \overline{N}_{g+1}, ..., \overline{N}_G) \geq 0 \tag{22}$$

$$\underline{W}^{+g}(\overline{N}_1, \overline{N}_2, ..., \overline{N}_{g-1}, \hat{N}_g, \overline{N}_{g+1}, ..., \overline{N}_G) < 0 \tag{23}$$

   We repeat the above for $j = 1, ..., G$ to obtain a vector of lower bounds, $\underline{N} = (\underline{N}_1, ..., \underline{N}_G)$

3. Repeat step 1 and step 2 for a fixed number of iterations (in our empirical application, 40 iterations). Stop the iterations (and skip step 4 below) if, at some point, $\underline{N} = \overline{N}$. In this case, convergence to a unique optimal vector was achieved and the set $\mathcal{N}^o$ is a singleton.

4. Enumerate all vectors $N$ such that $\underline{N} \leq N \leq \overline{N}$ (this is an inequality in vector sense). For each such vector $N$, conclude that $N \in \mathcal{N}^o$ if *both* the following conditions hold:

   (a) $\overline{W}^{+g}(N) \leq 0$ for each $g = 1, ..., G$

   (b) $\overline{W}^{-g}(N) \leq 0$ for each $g = 1, ..., G$

**Discussion.** The intuition underlying the above algorithm is the following: in steps 1 and 2 (and their iteration according to step 3), one obtains upper and lower bounds on the optimal market structure.[35] This provides a cost-effective way of

[34] It is easy to prove that $\tilde{N}_g$ exists, and is unique.

[35] A proof that these are valid bounds is available from the authors upon request. The intuition is that, due to the nested-logit structure, the marginal welfare gain from adding a station to any

limiting the space of vectors that need to be considered as candidates for socially-optimal market structures. In principle, these iterations may result in convergence to a unique optimal market structure. Alternatively, we are left with a set of vectors that satisfy the inequality $\underline{N} \leq N \leq \overline{N}$. Step 4 then enumerates each of these vectors. At each such vector, we examine whether adding or removing an in-metro station from any format-quality cell is *unambiguously* welfare enhancing. Only if no such welfare improvements are possible, is the vector admitted into the set $\mathcal{N}^o$.

It is interesting to examine the size of the set $\mathcal{N}^o$, i.e., the number of vectors that have not been ruled out as optimal, over the 163 markets. In certain markets, however, this set was computed several times due to the partial identification of quality levels. So, in total, this set was computed 229 times. Over these 229 computations, the largest $\mathcal{N}^o$ set contained 58 vectors, while the average and median sizes for this set are 5.3 and 4, respectively. Convergence to a unique optimal structure (i.e., the situation where the set $\mathcal{N}^o$ is a singleton) obtained in 30 of the 229 markets examined. All in all, therefore, the algorithm obtains rather small sets that provide informative bounds on optimal market structures across the different markets.

# C Robustness: within-format symmetry and selection

In this section we investigate the robustness of the listening equation estimation in the case where only horizontal differentiation is allowed (Section 5.1), focusing on two issues: the symmetry assumption, and the potential selection bias.

**The symmetry assumption**. Our horizontal-differentiation model assumes that stations within a market-format cell are symmetric, i.e., they have the same mean utility level.[36] We investigate robustness by estimating a listening equation that allows for horizontal differentiation only (as in the model presented in Section 5.1) but does not impose this restriction. Dropping this assumption allows one to control for station-level variables such as an FM dummy variable, transmission power (in 100 MHz units) and Antenna height (in thousand feet above average terrain).

A comparison of the symmetric vs. non-symmetric specifications is provided in Table 13. Relaxing the symmetry assumption causes the key parameter $\sigma$ to increase slightly, from about 0.52 to about 0.61. The regional and demographic effects appear robust, although the "home" and some of the format effects do change. In total, it appears that the symmetry assumption is reasonable, and, in particular, does not drive the excessive entry results: imposing this assumption actually reduces the value of $\sigma$, thus weakening the excessive entry finding.

**Selection**. As discussed above, three formats raise potential selection issues: "Religious," "Urban," and "Spanish". To be clear, the concern is that we may only observe an "Urban" station, say, in markets where such a format is likely to be popular. If this likelihood is affected by the unobserved $\xi_{gt}$, the format-market taste error, estimation of mean-utility parameters could be biased, and, in particular, the estimate of the coefficient on the "Urban" dummy variable may be expected to be biased upward.

given format-quality cell is diminishing in the number of stations already added. This marginal gain is also decreasing in the numbers of stations in the other formats.

[36] As explained above, in practice we differentiated between in-metro and out-metro stations, and assumed that all in-metro (out-metro) stations within such data cells are symmetric.

Table 13: Comparing symmetric vs. nonsymmetric specifications

| | | Symmetric | Non-Symmetric | | | Symmetric | Non-Symmetric |
|---|---|---|---|---|---|---|---|
| **Regional Dummies** | northeast | 0.122*** | 0.139*** | **Interactions** | hispXspan | 0.352*** | 0.286*** |
| | | (0.042) | (0.027) | | | (0.036) | (0.018) |
| | midwest | 0.0974** | 0.100*** | | blackXurban | 0.506*** | 0.460*** |
| | | (0.041) | (0.026) | | | (0.050) | (0.029) |
| | south | -0.0506 | -0.0906*** | | southXreligious | 0.809*** | 1.034*** |
| | | (0.041) | (0.026) | | | (0.095) | (0.061) |
| **Demograpics** | black | -0.0681*** | -0.0467*** | | southXcountry | 0.316*** | 0.274*** |
| | | (0.014) | (0.0090) | | | (0.072) | (0.048) |
| | hisp | -0.0233** | -0.0153** | **Correlation** | $\sigma$ | 0.519*** | 0.614*** |
| | | (0.0097) | (0.0063) | | | (0.063) | (0.034) |
| | income | -0.00258 | 0.0233** | **Station covariates** | FM | | 0.305*** |
| | | (0.017) | (0.011) | | | | (0.030) |
| | college | -0.0630** | -0.0797*** | | power | | 0.367*** |
| | | (0.027) | (0.017) | | | | (0.033) |
| **Home Dummy** | home | 0.639*** | 0.325*** | | HAAT | | 0.0537*** |
| | | (0.082) | (0.031) | | | | (0.016) |
| **Format Dummies** | mainstream | 0.595*** | 0.296*** | | FMXnt | | -0.237*** |
| | | (0.058) | (0.039) | | | | (0.064) |
| | chr | 0.431*** | 0.0306 | | Constant | -5.325*** | -5.247*** |
| | | (0.056) | (0.038) | | | (0.15) | (0.094) |
| | country | 0.389*** | 0.0939** | | | | |
| | | (0.053) | (0.038) | | | | |
| | rock | 0.561*** | 0.238*** | | | | |
| | | (0.049) | (0.033) | | | | |
| | oldies | 0.0447 | -0.331*** | | | | |
| | | (0.061) | (0.042) | | | | |
| | religious | -1.264*** | -1.336*** | | | | |
| | | (0.072) | (0.052) | | | | |
| | urban | -0.406*** | -0.556*** | | | | |
| | | (0.098) | (0.062) | | | | |
| | spanish | -1.165*** | -1.044*** | | | | |
| | | (0.096) | (0.058) | | | | |
| | nt | 0.214*** | 0.320*** | | | | |
| | | (0.053) | (0.032) | | | | |
| Observations | | 1919 | 4362 | | | | |
| R-squared | | 0.72 | 0.79 | | | | |

Notes: standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Power is in 100MHz units, HAAT is Antenna Height in 000s feet over average terrain.

To address this concern, we must look into the underlying selection mechanism that determines whether a metro would have an "Urban" (or "Spanish", or "Religious") station. Both observed (by the econometrician) factors, such as the size of the metro's Black population, and unobserved factors, such as the popularity of certain musical styles in the metro, can potentially be important determinants of whether an "Urban" station would be observed. While selection on observables would not cause a bias, selection on unobserved variables is a potential problem.

As a starting point, we estimate a probit model which relates the probability of observing the format to market characteristics. The results are presented in Table 14. The size of the metro's Black population significantly increases the probability of observing an Urban-formatted station, and the size of the Hispanic population has a similar effect on the probability of observing a Spanish station. Total population seems to have no explanatory power (for the Urban format case) or even a negative effect (for the Spanish format). Location of the metro in the South region has a positive and significant effect on the probability of observing both Urban and Spanish stations. The probability of observing a Religious-formatted station is increasing in the size of the Black population, but is not significantly affected by other characteristics. In particular, location in the "South" region has a positive, but insignificant, effect on the probability of observing a Religious station.[37]

Table 14: Probit Results for Formats

| | Urban | Spanish | Religious |
|---|---|---|---|
| Population (000s) | 0.0005 | -0.00251*** | -0.0003 |
| | (0.00088) | (0.00069) | (0.00047) |
| income | -0.121 | 0.628* | 0.275 |
| | (0.27) | (0.35) | (0.23) |
| Black Population (000s) | 0.0338*** | 0.00319 | 0.0116** |
| | (0.0095) | (0.0023) | (0.0046) |
| Hispanic Population (000s) | -0.0022 | 0.0741*** | -0.0004 |
| | (0.0019) | (0.014) | (0.0014) |
| college | 0.0588 | -0.501 | -0.367 |
| | (0.36) | (0.42) | (0.28) |
| northeast | 0.772 | 0.934 | -0.212 |
| | (0.50) | (0.72) | (0.43) |
| midwest | 0.243 | 1.322* | -0.394 |
| | (0.48) | (0.73) | (0.40) |
| south | 1.348*** | 1.866** | 0.369 |
| | (0.52) | (0.79) | (0.41) |
| Constant | -0.855 | -4.385*** | -0.0618 |
| | (0.99) | (1.60) | (0.89) |
| Observations | 163 | 163 | 163 |

Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[37] Since a Religious station is observed in almost 80% of the markets, it is difficult to estimate the probit parameters with precision.

A non-parametric investigation of the relationship between Black (Hispanic) population and the likelihood of observing an Urban (Spanish) station is available in Figure 2 (and 3). Figure 2 plots a dummy variable that takes the value 1 if an Urban station (either in-metro or out-metro) is broadcasting to the metro, and zero otherwise, against the metro's Black population. Figure 3 does the same for the Spanish format, and Hispanic population.[38]

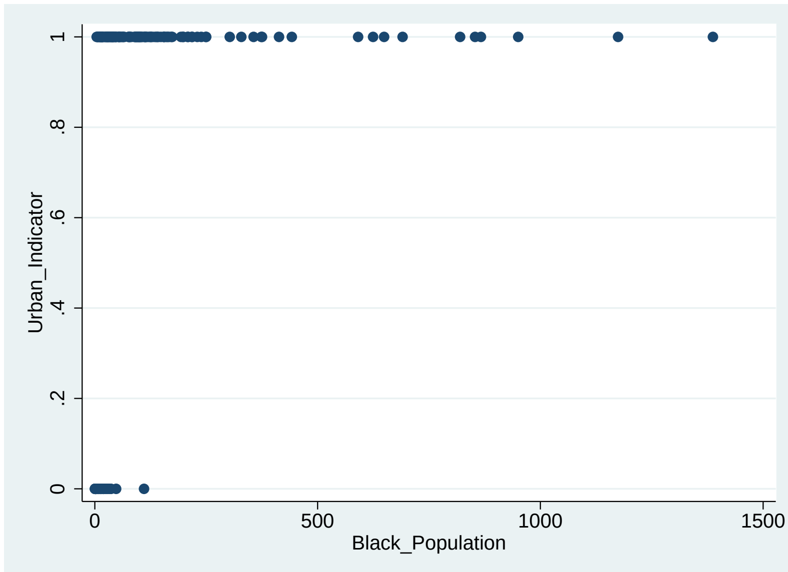


Figure 2: Presence of Urban Station Plotted against Black Metro Population, in 1000s (NYC Excluded)

[38]In both cases, NYC is excluded from the graph, as it has large Black and Hispanic populations (4.2 million in the case of Black population, 3.7 Million in the Hispanic case), and so excluding it allows for a clearer plot. In Figure 3, LA is excluded on similar grounds.

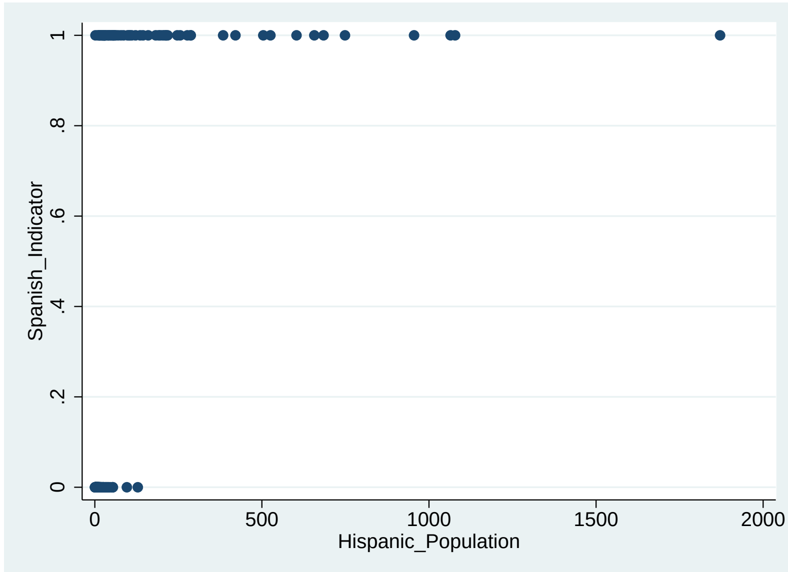


Figure 3: Presence of Spanish Station Plotted against Hispanic Metro Population, in 1000s (NYC, LA Excluded)

Figure 2 shows that, when the Black population is small, the metro may or may not have an Urban station. On the other hand, once the size of that population crosses a certain threshold, the metro *always* has an Urban station. This threshold can be characterized by the largest Black population in a metro that does not have an Urban station, and this happens in the city of Las Vegas, that has a Black population of about 110,000 persons. However, Las Vegas may be somewhat of an outlier, and the "true" threshold may be lower.[39]

A possible conclusion is that, if selection on unobservables occurs in the context of the Urban format, it is probably concentrated in those markets that have a small Black population. Figure 3 reveals similar patterns for the relationship between Hispanic population and the existence of a Spanish station, and the "threshold" population, of 129,000, is observed in Seattle-Tacoma.

The probit models of Table 14, as well as Figure 2 and 3, imply that the probability of observing an Urban (Spanish) station appears to be strongly driven by observables, i.e., the size of the main target population in the metro. For the Religious format, it is harder to locate a demographic "smoking gun" that would explain the probability of observing stations in this format, although Black population again emerges as having explanatory power.

While the evidence above is encouraging, it does not rule out selection on unobservables, and the resulting potential for selection bias. To further address this possibility, we perform a robustness check, motivated by an "Identification at infinity" approach.[40] The idea of this approach is to restrict attention to those markets where the probability of observing, say, an Urban station, as predicted by the probit

[39]Las Vegas is the most-right dot for which the "Urban Indicator" takes the value zero in Figure 2. Second-in-line is Omaha-Council Bluffs, with a Black population of about 48,000 people, and no Urban-formatted station.

[40]See for example Heckman and Navarro (2007).

specification in Table 14, is higher than, say, 99%. In this group of markets, there should be virtually no selection problem, since only a huge negative taste shock $\xi$ could prevent an Urban station from broadcasting to this metro. As a consequence, estimating the listening model using only observations from this restricted subsample of markets should yield estimates that are robust to selection bias.

We, therefore, compare the estimates of the listening function obtained using the full sample (again referring to the horizontal-differentiation model of Section 5.1 as the base model) with those obtained from subsamples that include only those markets in which the probability of observing an Urban, Spanish, or Religious station is higher than 99%. The results are presented in Table 15. Note the estimates which appear in bold text: these are the estimates that are most likely to be biased by selection, i.e., those pertaining to the dummy variables for "Urban," "Spanish," and "Religious," as well as relevant interactions of these dummies with demographic variables. If selection bias is important, these estimated coefficients might be lower when using the restricted subsamples compared with the results obtained using the full sample.

Rather than offering a formal test, we simply examine the relevant coefficients in Table 15, and ask whether these results are consistent with selection bias. The emerging picture appears to be mixed; the estimated coefficient on the "Urban" dummy variable is (-0.40) for the full sample, and actually increases to a statistically insignificant estimate of (-.17) in the "selection free" subsample. The coefficient on the interaction of "Urban" with the percentage of the market's Black population is decreasing very slightly, from 0.50 in the full sample, to 0.46 in the restricted subsample. There seems to be little evidence, therefore, that estimates concerning the Urban dummy variable are upward-biased due to selection.

In the case of the "Spanish" dummy, both the coefficient on this dummy variable itself, and on its interaction with the percentage of the market's Hispanic population, appear to drop when we shift from the full sample to the restricted subsample (from (-1.16) to (-1.37) for the dummy variable itself, and from .35 to .27 for the interaction term). This is consistent with a certain degree of sample selection bias. Finally, the results for the "Religious" format do not appear consistent with selection bias.

Summing up, for both the Urban and Spanish formats, there is both parametric and non-parametric evidence that the selection is strongly driven by observed metro characteristics, for which we control. Our robustness check does not indicate selection bias concerning the Urban or Religious formats. However, some findings are consistent with selection bias in the case of the Spanish format (remembering that this format is only present in 40% of the markets).

# D A closed-form solution for the fixed effects

The multinomial likelihood function of stations' market shares (as opposed to the likelihood of within-format shares, presented in the text), can be written as:

$$logL(s,x;\beta,\sigma,\xi) = \sum_t \sum_g \sum_{j \in g} n_{jt} \times \left[ \frac{\delta_{jt}}{1-\sigma} - \sigma log D_{g(j)t} - log(1 + \sum_m D_{mt}^{1-\sigma}) \right]$$

Also note that:

Table 15: Restricting Attention to "Selection-free" Sub-samples

| | Full sample | "Urban Subsample" | "Spanish Subsample" | "Religious Subsample" |
|---|---|---|---|---|
| $\sigma$ | 0.519*** | 0.656*** | 0.390** | 0.765*** |
| | (0.063) | (0.11) | (0.17) | (0.19) |
| black | -0.0681*** | -0.0311 | -0.0767 | -0.0424 |
| | (0.014) | (0.024) | (0.051) | (0.044) |
| hisp | -0.0233** | -0.0287 | -0.0194 | -0.0121 |
| | (0.0098) | (0.019) | (0.017) | (0.025) |
| income | -0.00258 | 0.00917 | 0.0680* | 0.0588 |
| | (0.017) | (0.026) | (0.039) | (0.064) |
| college | -0.0630** | -0.125*** | -0.209*** | -0.127 |
| | (0.027) | (0.042) | (0.073) | (0.12) |
| home | 0.639*** | 0.432*** | 0.699*** | 0.328 |
| | (0.083) | (0.16) | (0.21) | (0.30) |
| mainstream | 0.595*** | 0.803*** | 0.150 | 0.887*** |
| | (0.058) | (0.099) | (0.095) | (0.17) |
| chr | 0.431*** | 0.383*** | -0.183 | 0.250 |
| | (0.056) | (0.084) | (0.23) | (0.17) |
| country | 0.389*** | 0.163* | -0.323** | -0.408** |
| | (0.054) | (0.094) | (0.15) | (0.20) |
| rock | 0.561*** | 0.635*** | -0.0104 | 0.616*** |
| | (0.049) | (0.077) | (0.097) | (0.13) |
| oldies | 0.0447 | -0.0100 | -0.380* | 0.00510 |
| | (0.062) | (0.099) | (0.22) | (0.20) |
| religious | **-1.264*** ** | -1.238*** | -1.627*** | **-1.244*** ** |
| | (0.073) | (0.11) | (0.23) | (0.21) |
| urban | **-0.406*** ** | **-0.173** | -0.458** | -0.250 |
| | (0.099) | (0.14) | (0.23) | (0.30) |
| spanish | **-1.165*** ** | -1.311*** | **-1.369*** ** | -1.226*** |
| | (0.097) | (0.14) | (0.17) | (0.22) |
| nt | 0.214*** | 0.494*** | -0.0702 | 0.669*** |
| | (0.054) | (0.091) | (0.096) | (0.16) |
| hispXspan | **0.352*** ** | 0.550*** | **0.271*** ** | 0.564*** |
| | (0.036) | (0.085) | (0.055) | (0.12) |
| blackXurban | **0.506*** ** | **0.461*** ** | 0.345*** | 0.554*** |
| | (0.051) | (0.067) | (0.13) | (0.14) |
| southXreligious | **0.809*** ** | 0.982*** | -0.0883 | **0.954*** ** |
| | (0.095) | (0.14) | (0.18) | (0.24) |
| southXcountry | 0.316*** | 0.614*** | 0.306** | 1.108*** |
| | (0.072) | (0.11) | (0.14) | (0.22) |
| Constant | -5.325*** | -5.031*** | -5.195*** | -5.161*** |
| | (0.15) | (0.28) | (0.46) | (0.55) |
| Observations | 1919 | 846 | 444 | 341 |
| R-squared | 0.72 | 0.72 | 0.73 | 0.66 |

Notes: Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Regional effects not reported for lack of space.

$$D_{gt} = e^{\psi_{gt}/1-\sigma}\sum_{j\in g} e^{\kappa_1\cdot q_{jt}+\kappa_2\cdot h_{jt}} \Rightarrow \frac{\partial D_{gt}}{\partial\psi_{gt}} = D_{gt}/(1-\sigma),\ \ logD_{gt} = \psi_{gt}/(1-\sigma)+log\big[\sum_{j\in g} e^{\kappa_1\cdot q_{jt}+\kappa_2\cdot h_{jt}}\big]$$

Taking the FOC of the likelihood function with respect to $\psi_{kt}$ yields:

$$\begin{aligned} 0 = \frac{\partial \log L}{\partial \psi_{kt}} &= \sum_{j\in k} n_{jt}\left[1 - \frac{D_{kt}^{1-\sigma}}{1+\sum_m D_{mt}^{1-\sigma}}\right] - \sum_{g\neq k}\sum_{j\in g} n_{jt}\left[\frac{D_{kt}^{1-\sigma}}{1+\sum_m D_{mt}^{1-\sigma}}\right] \\ &= \sum_{j\in k} n_{jt}[1-\mathbf{s}_{kt}] - \sum_{g}\sum_{j\in g} n_{jt}\mathbf{s}_{kt} = \sum_{j\in k} n_{jt} - \mathbf{s}_{kt}\times n_t \Rightarrow s_{kt} = \mathbf{s}_{kt} \end{aligned}$$

where the second equality utilizes the nested-logit formula for the share of nest (format) $k$, and $s_{kt}, \mathbf{s}_{kt}$ are the observed and predicted shares of listening to format $k$, respectively. The above derivations show that the optimal (i.e., likelihood-maximizing) solution for the fixed effects sets predicted and observed format shares to be equal. Taking logs on both sides of $s_{kt} = \mathbf{s}_{kt}$ yields:

$$\begin{aligned} log(s_{kt}) &= (1-\sigma)log\big[D_{kt}\big] + log(\mathbf{s}_{0t}) \\ &\Rightarrow log(s_{kt}) = \psi_{kt} + (1-\sigma)log\left[\sum_{j\in k} e^{\kappa_1\cdot q_{jt}+\kappa_2\cdot h_{jt}}\right] + log(\mathbf{s}_{0t}) \\ &\Rightarrow \psi_{kt} = log(s_{kt}) - log(s_{0t}) - (1-\sigma)log\left[\sum_{j\in k} e^{\kappa_1\cdot q_{jt}+\kappa_2\cdot h_{jt}}\right] \end{aligned}$$

The last step replaced the predicted share choosing the outside option $\mathbf{s}_{0t}$ by its empirical counterpart $s_{0t}$. This is a valid replacement since all nests' predicted shares are matched to their empirical counterparts, including the nest which only element is the outside option.

# E Standard errors for the two-step estimation procedure

Standard errors for $\kappa$ are estimated by standard MLE formulae given 4,362 observations (stations). To derive the standard errors for the remaining parameters, estimated in the second stage, we let:

$$\begin{aligned} &\theta = (\sigma, \lambda),\ \ \kappa = (\kappa_1, \kappa_2) \\ &G_n(\theta,\kappa) = \frac{1}{1433}\sum_{i=1}^{1433} \xi_i(\theta,\kappa)\cdot Z_i \end{aligned}$$

where $\xi_i$ is the unobserved market-format taste shifter satisfying $\xi_i = \psi_i - d_i\lambda$, and

$Z_i$ is a 22-vector (row) of instruments for that market-format (population, number of out-metro stations, number of out-metro stations in same format and all the 19 $d$ covariates). The number of observations is equal to the number of market-format pairs, 1,433.

Importantly, however, we wish to allow this estimator to utilize as many observations as the ML estimator for $\kappa$. Therefore, we replicate each of these 1,433 observations in the following fashion: the observation pertaining to market-$t$, format-$g$ is replicated $r_{gt}$ times, where $r_{gt}$ is the number of stations in market-$t$, format-$g$. Each such observation is also divided by $r_{gt}$.

The two-step estimator $\hat{\theta}_n$ minimizes:

$$Q_n(\theta) = \big[G_n(\theta, \hat{\kappa}_n)\big]'(A_n' A_n)\big[G_n(\theta, \hat{\kappa}_n)\big]/2$$

where $\hat{\kappa}_n$ is the ML estimator for $\kappa$. $A_n' A_n$ is the weight matrix (in our case it equals $((1/n)Z'Z)^{-1}$). We assume the following:

$$\sqrt{n}\begin{pmatrix} G_n(\theta_0, \kappa_0) \\ \hat{\kappa}_n - \kappa_0 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \sim N\left(0, \begin{pmatrix} V_{10} & V_{20} \\ V_{20}' & V_{30} \end{pmatrix}\right)$$

This result can be verified by applying the multivariate CLT to the GMM moments $G_n$ and by applying a linear expansion that expresses $\sqrt{n}(\hat{\kappa_n} - \kappa_0)$ as the sum of an asymptotically-negligible term (i.e., $o_p(1)$), and a term which multiplies the information matrix by a function of the scores. It is then possible to rewrite the above assumption as:

$$\sqrt{n}\begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} \xi_i(\theta_0, \kappa_0) \cdot Z_i \\ \frac{1}{n}\sum_{i=1}^{n} -(I_0)^{-1} s_i(w_i, \kappa_0) \end{pmatrix}_{5\times 1} \xrightarrow{d} N\left(0, \begin{pmatrix} V_{10} & V_{20} \\ V_{20}' & V_{30} \end{pmatrix}\right)$$

The elements $V_{10}, V_{20}, V_{30}$ are therefore given as follows. To compute $V_{10}$, denote:

$$g_i = \begin{bmatrix} \xi_i \cdot Z_{i1} \\ \xi_i \cdot Z_{i2} \\ \vdots \\ \xi_i \cdot Z_{i22} \end{bmatrix}$$

We assume that (i) $g_i$ are IID (ii) $E(g_i) = 0$ (iii) $Var(g_i) = \Sigma$, where $\Sigma$ is a $22 \times 22$ matrix. Then by the CLT: $\sqrt{n} G_n \xrightarrow{d} N(0, \Sigma)$, implying: $V_{10} = \Sigma$. In other words, $V_{10}$ is the variance of the moments of our GMM step.

To compute $V_{20}$, we use the multivariate CLT to obtain:

$$\begin{aligned} V_{20} &= E\Big[\big(g_i(w_i, \theta_0, \kappa_0) - E(g_i(w_i, \theta_0, \kappa_0))\big)\big( - (I_0)^{-1} s_i(w_i, \kappa_0) - E(-(I_0)^{-1} s_i(w_i, \kappa_0))\big)'\Big] \\ &= E\Big[\big(g_i(w_i, \theta_0, \kappa_0)\big)\big( - (I_0)^{-1} s_i(w_i, \kappa_0)\big)'\Big] \end{aligned}$$

Finally, to compute $V_{30}$, we obtain, by the limit variance of the ML estimator, and the Information Matrix Equality:

$$V_{30} = \left[ E s_i(\kappa_0) s_i(\kappa_0)' \right]^{-1}_{1\times 1}$$

where $s_i$ is the Score function.

Given the above, we obtain the following result:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, B_0^{-1}\Omega_0 B_0^{-1})$$

The asymptotic variance to be estimated is, therefore, $B_0^{-1}\Omega_0 B_0^{-1}$. We estimate those matrices by:

$$\hat{B}_n = \left(\frac{\partial}{\partial\theta'} G_n(\hat{\theta}_n, \hat{\tau}_n)\right)' A_n' A_n \frac{\partial}{\partial\theta'} G_n(\hat{\theta}_n, \hat{\kappa}_n)$$

$$\hat{\Omega}_n = \left(\frac{\partial}{\partial\theta'} G_n(\hat{\theta}_n, \hat{\kappa}_n)\right)' A_n' A_n \big(\hat{V}_{1n} + \hat{\Delta}_n \hat{V}_{2n}' + \hat{V}_{2n}\hat{\Delta}_n' + \hat{\Delta}_n \hat{V}_{3n} \hat{\Delta}_n'\big) \times A_n' A_n \frac{\partial}{\partial\theta'} G_n(\hat{\theta}_n, \hat{\kappa}_n)$$

$$\hat{\Delta}_n = \frac{\partial}{\partial\kappa'} G_n(\hat{\theta}_n, \hat{\kappa}_n)$$

Where the expressions $\frac{\partial}{\partial\theta'} G_n(\hat{\theta}_n, \hat{\kappa}_n)$, $\frac{\partial}{\partial\kappa'} G_n(\hat{\theta}_n, \hat{\kappa}_n)$ are computed using numerical derivatives. Estimators for $\hat{V}_{1n}, \hat{V}_{2n}, \hat{V}_{3n}$ are obtained by computing the sample analogs of the expressions for $V_{10}, V_{20}, V_{30}$.