Coordination Games, Multiple Equilibria and the Timing of Radio Commercials

Andrew Sweeting*

Northwestern University

January 29, 2005

Abstract

This paper presents a model of an imperfect information coordination game which has multiple equilibria if the incentive of players to choose the same action is strong enough. It shows how the existence of multiple equilibria in the model and in data from multiple independent repetitions of the game can identify the parameters. The model is estimated using new data on the timing of commercials by music radio stations in 147 local radio markets. Stations may have an incentive to choose the same times for commercials because many listeners try to avoid commercials by switching stations. There is evidence of multiple equilibria, with commercials clustered at different times in different markets, during drivetime hours. The estimated incentive to coordinate has a modest effect on Nash equilibrium timing strategies but commercials would overlap almost perfectly if each station internalized how its timing affects the audience of commercials on other stations. Most markets stay in the same equilibrium for the duration of my data and the incentive to coordinate is larger in smaller markets and in markets with more concentrated station ownership.

*Mailing address: Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston IL 60208. Email: asweeting@northwestern.edu. This paper is a revised version of chapter 2 of my MIT Ph.D. thesis. I thank Glenn Ellison, Paul Joskow and Whitney Newey for many helpful discussions and Pat Bajari, Victor Chernozhukov, Aviv Nevo, Nancy Rose and numerous participants at academic seminars and the 2003 National Association of Broadcasters/Broadcast Education Association Convention in Las Vegas for useful comments. I also thank Rich Meyer of Mediabase 24/7 for providing access to the airplay data and the National Association of Broadcasters for providing a research grant for the purchase of the BIAfn *MediaAccess Pro* database including the Arbitron share data. All views expressed in this paper, and any errors, are my own.

1 Introduction

A common aim in applied microeconomics is to identify whether an individual agent's choice of action is affected by interactions with other agents making similar choices. For example, Glaeser et al. (1996) examine whether peer effects affect crime, Ellison and Glaeser (1997) study whether the benefits of agglomeration affect the geographic distribution of industries and Bresnahan and Reiss (1991) analyze how the prospect of competition affects market entry. Models where these sorts of interaction are important frequently have multiple equilibria. For example, if there are large benefits to agglomeration then a model of firm location decisions may have several equilibria in which firms cluster in different places. Multiple equilibria are widely perceived to create problems in estimating these models. This paper shows that it is possible to estimate a model where there are multiple equilibria both in the model and in the data and, moreover, that the existence of multiple equilibria both in the model and in the data can identify the parameters of the model including the effect of the interactions. The intuition is straightforward. Multiple equilibria only arise when interactions are important and the existence of multiple equilibria in the data allows us to rule out parameters which cannot support all of the observed equilibria.

These results are developed in the context of a simple two action imperfect information game where each player may want to choose the same action as other players. The game has multiple equilibria if this incentive to coordinate is strong enough. The game is used to study whether commercial music radio stations in a local market have an incentive to choose the same times for commercial breaks (hereafter I call this "the incentive to coordinate"). The timing of commercial breaks plays a potentially important role in the economics of the radio industry because many listeners seek to avoid commercials by switching stations and advertisers' willingness to pay for commercial time should fall as more listeners avoid commercials. For example, the average in-car listener switches stations 29 times per hour primarily to avoid commercials and listens to less than half of the number of commercials she would hear if she never switched stations (Abernethy (1991), McDowell and Dick (2003)). The industry's annual advertising revenues are around $20 billion (Radio Advertising Bureau (2002), p. 4) and the avoidance of commercials by in-car listeners alone potentially costs the industry several billion dollars in revenue.[1] Not surprisingly advertisers have suggested that stations should play commercials at the same time to prevent listeners avoiding commercials.[2] A simple model illustrates this logic. Suppose that time is divided into discrete intervals. While actual time is continuous the scheduling of music radio programming is a largely discrete problem because it involves ordering songs, commercial breaks and other programming such as travel updates. There are N symmetric stations and in each discrete time interval each station decides to play commercials or music. Every listener has a first choice station (the "P1" in radio jargon) and a second choice station. Each station is the first choice of one unit of listeners who are equally divided between the other stations for their second choice. Independent of station tastes, a proportion $1-\theta$ of listeners never switch stations and always listen to their first choice station. Proportion θ listen to their first choice station unless it has commercials and their second choice station has music in which case they listen to their second choice. The audience of a commercial break when n_{-it} other stations have commercials at the same time is $1-\theta+\theta\frac{n_{-it}}{N-1}$. If $\theta > 0$ a station seeking to maximize the audience of its commercials should play them at the same time as a greater proportion of other stations.[3]

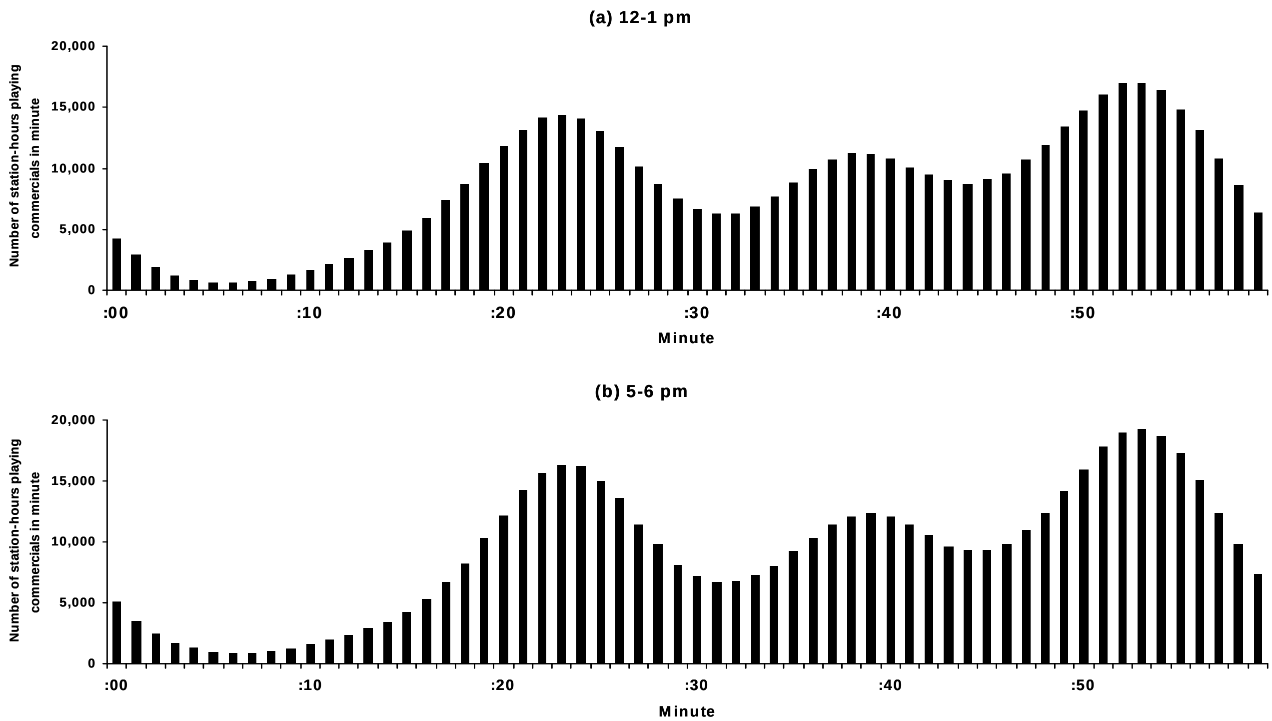
Figure 1 shows how many stations play commercials each minute for two different hours of the day based on a large sample of station-hours from 1,094 stations in 147 different local radio markets. Stations do tend to play commercials at the same time so that there are, for example, over 15 times as many commercials five minutes before the end of each hour as five minutes after the hour. While the incentive to coordinate provides one explanation for this pattern it could also be explained by the existence of "common factors" which make certain times, such as just before the end of the hour, attractive for all stations to play commercials independent of the times chosen by other stations. Two common factors consistent with Figure 1 are mentioned by people in the industry: stations "sweep"

[1] Arbitron Company and Edison Media Research (2003), p. 11 estimate that 34% of radio listening takes place in-car.

[2] For example, Brydon (1994), an advertising consultant, argues that "for advertisers, the key point is this: if, at the touch of a button, you can continue to listen to that [music] for which you tuned in, why should you listen to something which is imposing itself upon you, namely a commercial break." He suggests that either stations should play very short breaks which would not make switching worthwhile or stations should "transmit breaks at universally agreed uniform times. Why tune to other stations if it's certain that they will be broadcasting commercials as well?".

[3] Advertisers would like stations to try to maximize the audience of the commercials for a given price of commercial time but advertisers and stations are only able to measure the audience of commercials imperfectly. Dick and McDowell (2003) describe how commercial avoidance on different stations can be estimated from standard Arbitron ratings which are used by advertisers and advertising agencies. Models of television station timing choices, such as Epstein (1998) and Zhou (2000), also assume that stations maximize the audience of TV commercials which are also measured imperfectly (see, for example, Media Daily News (2004)).

Figure 1: Number of Stations Playing Commercials Each Minute



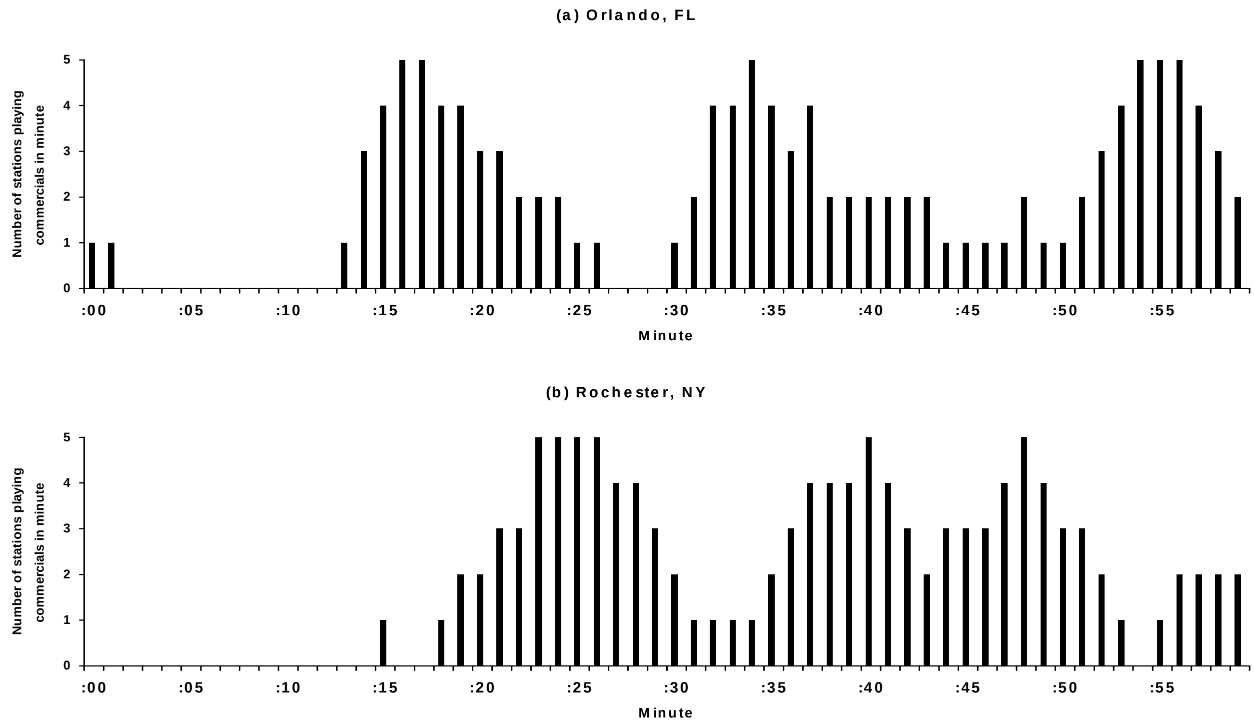
Note: based on airplay data described in Section 5.2 from 1,094 contemporary music stations in 147 local markets. 12-1 pm histogram based on 50,664 station-hours and 5-6 pm histogram based on 50,459 station-hours.

the quarter-hours with music because of how Arbitron estimates station ratings and they avoid playing commercials at the start of each hour because many listeners switch on then and listeners are believed to particularly dislike hearing commercials when they first tune-in.[4]

This leads to the question of how we can identify whether the incentive to coordinate affects timing choices. If commercials are clustered (i.e., played at the same time) in the same minutes in every market then it is clearly very difficult to tell whether the incentive to coordinate matters. On the other hand, if commercials are clustered in every market but at different times in different markets and we are willing to assume that common factors which favor particular times are the same across markets because, for example, Arbitron uses the same methodology in every market, then this may provide evidence of an incentive to coordinate. It is this clustering of commercials at different times in different markets which is interpreted as reflecting multiple equilibria in my model. Figure 2 shows

[4] Arbitron's methodology counts a listener as listening to a station for a full quarter-hour if she listens to it for five minutes during the quarter-hour. This means that a listener who can be kept over the quarter-hours points (:00, :15, :30 and :45) is likely to count for two quarter-hours (Warren (2001), p. 23-24). Keith (2000), p. 96 discusses the connection between the timing of commercials and when listeners tune-in.

Figure 2: Timing of Commercials in Orlando, FL and Rochester, NY on October 30, 2001 5-6 pm



an example of what clustering at different times in different markets looks like by showing histograms for 5-6 pm for Orlando, FL and Rochester, NY on October 30, 2001. Each histogram has 3 peaks but they occur in different minutes.

I use panel data on the timing of commercials in 147 different markets to estimate several specifications of games with multiple equilibria. In the simplest specification I find evidence of multiple equilibria, which allows the incentive to coordinate to be identified in my framework, during drivetime but not outside drivetime. This is consistent with a strong incentive to coordinate leading to multiple equilibria because during drivetime there are more in-car listeners who tend to switch stations more than those at home or at work (in the model θ should be higher during drivetime).[5] However, I find that the incentive to coordinate has a relatively modest effect on station timing strategies during drivetime. This is explained by the fact that it can be costly for a station, which has to fit commercials around other programming in a natural way, to play its commercials at the same time as

[5] MacFarland (1997), p. 89, reports that, based on a 1994 survey, 70% of in-car listeners switch at least once during a commercial break compared with 41% of at home and 29% of at work listeners. Arbitron estimates that 39.2% of listening is in-car during drivetime compared with 27.4% 10 am-3 pm and 25.0% 7 pm-midnight (Fall 2001 data from the Listening Trends section of Arbitron's website, www.arbitron.com).

other stations and each station fails to internalize how its timing affects the audience of commercials on other stations. The estimates imply that commercials would overlap almost perfectly if stations maximized their expected joint payoffs. I use the panel dimension of the data to show that markets tend to remain in the same equilibrium over time i.e., commercials are clustered in the same minutes from day-to-day. The inclusion of market characteristics shows that the incentive to coordinate is greater in smaller markets and in markets with more concentrated station ownership i.e., clustering is more pronounced in these markets.

The paper is related to four distinct literatures. The first literature examines whether social interactions or social learning matter in situations where we observe clustering of agents choices (e.g., Glaeser et al. (1996) and Duflo and Saez (2002)). I show how this clustering can be modelled as representing multiple equilibria and how this multiplicity can help to identify the parameters in a simple structural model where this kind of interaction may be present. Brock and Durlauf (2001), p. 3331, note the theoretical possibility of using multiple equilibria in the context of a discrete choice, social interactions model very similar to the one used here. I develop this idea in detail and apply it to real data. The second literature deals with the estimation of games with multiple equilibria.[6] Multiple equilibria, where a single set of parameters, data and unobservables, lead to more than one equilibrium outcome potentially create problems for estimation because they complicate the specification of the likelihood. This has been dealt with in different ways in different contexts. First, some researchers have structured their models to give unique outcomes (e.g., sequential entry in Berry (1992)) or focused on a prediction of the model which is unique (e.g., the number of firms in Bresnahan and Reiss (1991)). Second, some models are identified because, even though some parameters can support multiple equilibrium strategies, each set of equilibrium strategies is only supported by a unique set of parameters. The models are estimated under the assumption that there is a single equilibrium in the data (e.g., Salami (1986), Sargent and Wallace (1987), Moro (2003), Aguirregabiria and Mira (2004)). Third, Andrews and Berry (2003) and Ciliberto and Tamer (2004) have shown that if multiple equilibria create inequalities for the likelihood of observed outcomes then these can be used to identify some features of a model such as ranges of parameters. Fourth, Bjorn and Vuong (1985),

[6] Cooper (2004) summarizes some of this literature particularly in the context of macroeconomic and labor models.

Ackerberg and Gowrisankaran (2002) and Bajari et al. (2004) show that it is possible to estimate models with multiple equilibria by identifying or making assumptions about the mixture of different equilibria in the data. I also treat the existence of multiple equilibria as a mixture problem but the current paper has the difference from all of the above papers that the existence of multiple equilibria in the game and in the data is actually a source of identification. The third literature looks for evidence of multiple equilibria in the real world. Davis and Weinstein (2004) examine the recovery of population and industrial structure in Japanese cities after the Second World War and find that pre-war patterns were largely restored contrary to what might be expected given the existence of multiple equilibria in models of economic geography. Bajari et al. (2004) find some evidence of multiple equilibria, including mixed strategy equilibria, being played in a game where golf courses in the Carolinas choose to develop websites. I find some evidence of multiple equilibria when I examine station timing choices. The fourth literature examines the timing of commercials on radio and television. Epstein (1998) and Zhou (2000) provide theoretical models where television stations placing commercial breaks within well-defined programs choose to have them at the same times in Nash equilibrium. Epstein provides empirical evidence that the major US networks do tend to have their commercials at the same time especially at the beginning of programs. Sweeting (2004a) uses the same data but a very different empirical approach to test whether music radio stations have an incentive to coordinate. A model where stations have an incentive to coordinate gives predictions about how the equilibrium overlap of commercials should vary with observable market characteristics. A model where stations want to choose different times for commercial breaks (justified by some specifications of listener behavior) gives different predictions. Regressions of measures of how much commercials overlap on market characteristics provide consistent support for the coordination model with more overlap in markets with fewer stations, less listening to out-of-market stations, more concentrated ownership and in markets where stations' shares of listenership are more asymmetric. These relationships are stronger during drivetime than during other parts of the day which is consistent with listeners having a greater propensity to switch stations during drivetime.

Section 2 presents the model of an imperfect information coordination game. Section 3 provides the identification results. Section 4 describes how I estimate the model and test for multiple equilibria.

Section 5 analyzes the timing of commercial breaks. Section 6 concludes.

2 An Imperfect Information Coordination Game

This section develops a theoretical model which can have multiple equilibria and explains how it can be taken to data. I describe the model without explicitly relating it to the timing of radio commercials which I discuss in detail in Section 5.

2.1 Model

Consider a game where N players simultaneously choose one of two actions $t \in \{0, 1\}$. Each player i has the following reduced form payoff function from choosing action t

$$\pi_{it} = \beta_t + \alpha P_{-it} + \varepsilon_{it} \tag{1}$$

where P_{-it} is the proportion of players other than i choosing action t. The β_ts allow one action to have, on average, a higher payoff for reasons not connected with coordination. α reflects the strength of the incentive to coordinate. If $\alpha = 0$ each player's payoff is independent of other players' choices but if $\alpha > 0$ payoffs increase with the proportion of other players choosing the same action. I assume that $\alpha \geq 0$ but discuss in Section 3 what I might see in the data if $\alpha < 0$ so that players want to choose different actions. ε_{it} is an idiosyncratic term which is private information to player i and which allows players to have different preferences over actions. It is assumed to be independently and identically distributed across players and actions with a Type 1 extreme value ("logit") distribution.

i's strategy S_i consists of a rule for selecting its action as a function of its εs and, if $\alpha > 0$, the strategies of the other players (S_{-i}). i's optimal strategy will be to choose action 1 if and only if

$$\beta_1 + \alpha E(P_{-i1}|S_{-i}) + \varepsilon_{i1} \geq \beta_0 + \alpha E(P_{-i0}|S_{-i}) + \varepsilon_{i0} \tag{2}$$

This is a threshold crossing model so β_1 and β_0 cannot be separately identified and I normalize $\beta_0 = 0$. The distribution of ε implies that the probability, prior to the realization of the εs, that i chooses action 1 when it uses its optimal strategy is

$$p_i^* = \frac{e^{\beta_1 + \alpha E(P_{-i1}|S_{-i})}}{e^{\beta_1 + \alpha E(P_{-i1}|S_{-i})} + e^{\alpha E(P_{-i0}|S_{-i})}} \tag{3}$$

This probability is a convenient way to describe a player's strategy. It is straightforward to show that if $\alpha \geq 0$ all Bayesian Nash equilibria will involve players using symmetric strategies.[7] All players choose one action so that in equilibrium $p_i^* = E(P_{-i1}|S_{-i}) = 1 - E(P_{-i0}|S_{-i}) = p^*$ where

$$p^* = \frac{e^{\beta_1 + \alpha p^*}}{e^{\beta_1 + \alpha p^*} + e^{\alpha(1-p^*)}} \tag{4}$$

For any (β_1, α) the game has multiple equilibria if more than one value of p^* satisfies (4). As payoffs depend on the proportion of other players choosing action 1 equilibrium strategies are independent of N.

Figure 3(a)-(d) shows how player i's reaction function and the equilibria change with β_1 and α. In each diagram the probability that every player other than i chooses action 1 is shown on the horizontal axis and i's probability of choosing action 1 is shown on the vertical axis. There is an equilibrium at any point where the reaction function crosses the 45^0 line. In Figure 3(a) $\alpha = 0$ so i's optimal strategy is independent of the strategies of other players and the reaction function is flat. As $\beta_1 > 0$ p^* is greater than $\frac{1}{2}$. If $\alpha > 0$ i's reaction function slopes upwards and has an S-shape because of the logit distribution of ε. In Figure 3(b) there is still a single equilibrium but the benefit to coordination means that p^* is greater than in Figure 3(a). In Figure 3(c) α is higher and there are three equilibria. The middle equilibrium (slope of the reaction function is greater than 1) is unstable in the sense that if there was a small deviation from the equilibrium then the application of iterated best responses would not return strategies to the same equilibrium. In the rest of the paper, I will assume that only stable equilibria are played although, as I note below, unstable equilibria in the data could also help with identification. The S-shape of the reaction function, which has maximum slope at $p_i^* = \frac{1}{2}$, implies that there are a maximum of two stable equilibria and that one of them will involve players choosing

[7] Suppose that the equilibrium is not symmetric so that a player j chooses action 1 with higher probability than another player k. This implies that $E(P_{-k1}|S_{-k}) - E(P_{-k0}|S_{-k}) > E(P_{-j1}|S_{-j}) - E(P_{-j0}|S_{-j})$ which, from (3), implies that k would actually choose action 1 with higher probability than j, a contradiction. If $\alpha < 0$ there will be a symmetric equilibrium but there may also be asymmetric equilibria.

Figure 3: Station Reaction Functions and the Number of Equilibria for Different Values of β_1 and α

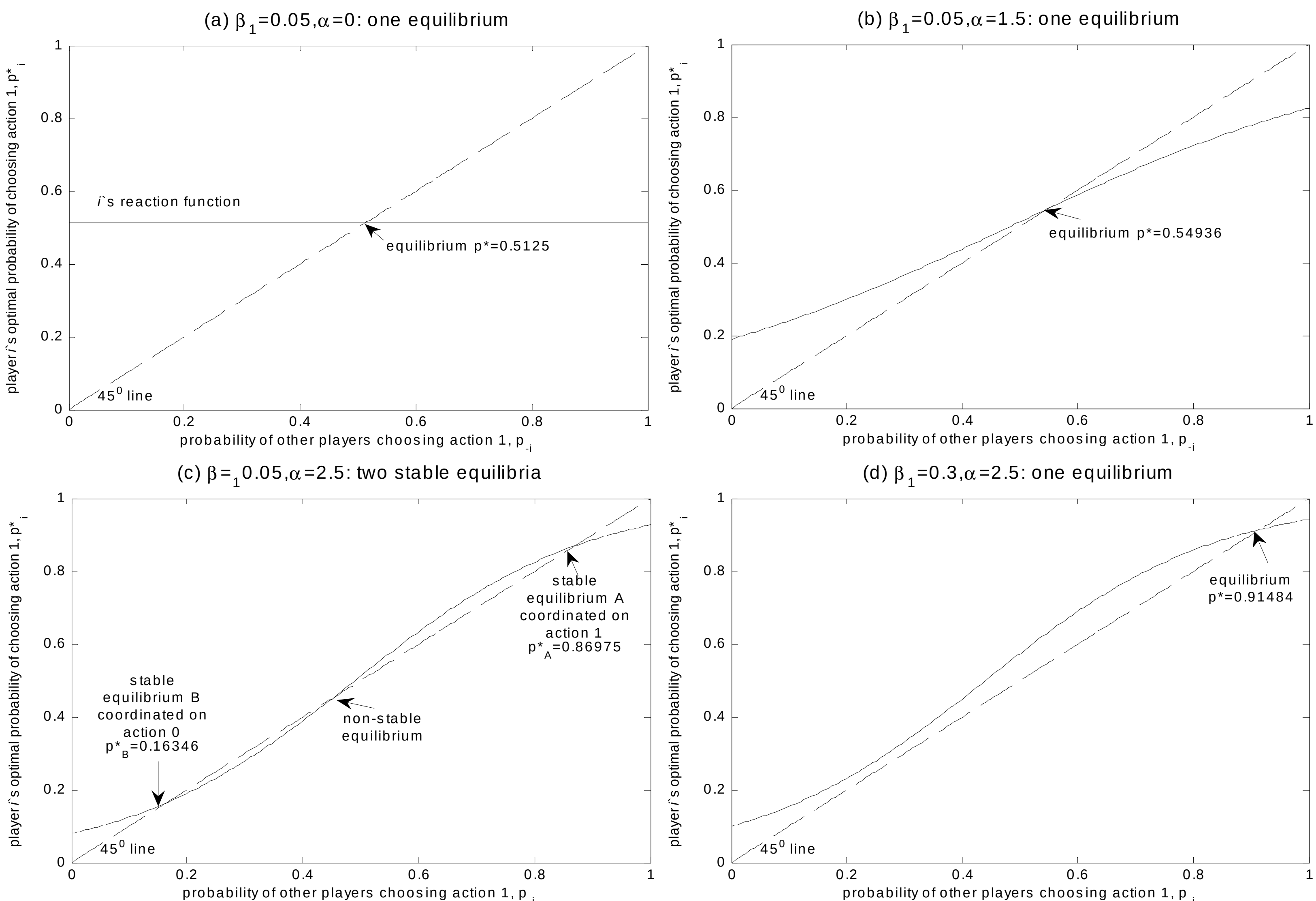
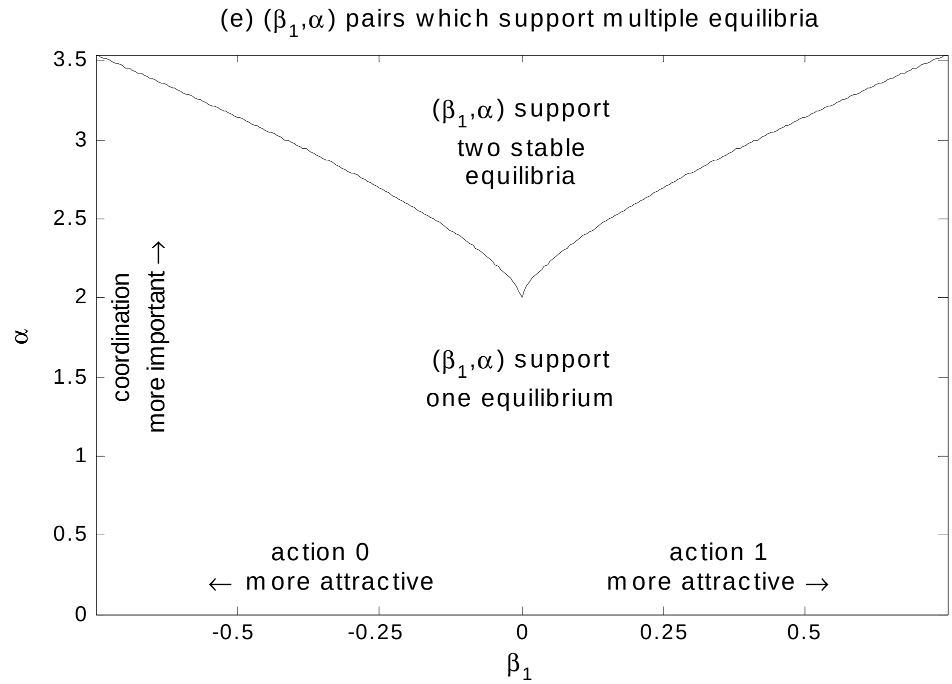


Figure 3: continued



action 1 with probability greater than $\frac{1}{2}$ (I label this equilibrium A, $p_A^* > \frac{1}{2}$) and that the other will involve players choosing action 1 with probability less than $\frac{1}{2}$ (equilibrium B, $p_B^* < \frac{1}{2}$). In Figure 3(c) players choose the same action with higher probability in equilibrium A than in equilibrium B because $\beta_1 > 0$. In Figure 3(d) β_1 increases, the reaction function shifts upwards and equilibrium B ceases to exist. The intuition for this result is that if action 1 is, on average, much more attractive than action 0 then an equilibrium involving coordination on action 0 cannot be sustained. Figure 3(e) summarizes these results by dividing the (β_1, α) parameter space into regions which support one equilibrium and two equilibria.

2.2 Empirical Model and "Equilibrium Selection"

Suppose that the data consists of observations from independent repetitions of the game, indexed by m, listing the number of players (N_m) and the number choosing action 1 (n_{1m}) with $N_m - n_{1m}$ choosing action 0. By assumption β_1 and α are the same across repetitions. If (β_1, α) support a single equilibrium $p^*(\beta_1, \alpha)$ then the probability that n_{1m} choose action 1 is the binomial probability

$$\Pr(n_{1m}|\beta_1, \alpha, N_m) = \frac{N_m!}{n_{1m}!(N_m - n_{1m})!} p^*(\beta_1, \alpha)^{n_{1m}} (1 - p^*(\beta_1, \alpha))^{N_m - n_{1m}} \tag{5}$$

If (β_1, α) support two equilibria, $p_A^*(\beta_1, \alpha)$ and $p_B^*(\beta_1, \alpha)$ which satisfy $\frac{\partial\left(\frac{e^{\beta_1+\alpha p^*}}{e^{\beta_1+\alpha p^*}+e^{\alpha(1-p^*)}}\right)}{\partial p^*} < 1$ (stable equilibria) and $p_A^*(\beta_1, \alpha) \geq p_B^*(\beta_1, \alpha)$, then the probability depends on the equilibrium played. If Z_m^A is an indicator which is 1 if equilibrium A is played in repetition m then the probability that n_{1m} choose action 1 is

$$\Pr(n_{1m}|\beta_1, \alpha, Z_m^A, N_m) = \frac{N_m!}{n_{1m}!(N_m - n_{1m})!} \begin{pmatrix} Z_m^A p_A^*(\beta_1, \alpha)^{n_{1m}} (1 - p_A^*(\beta_1, \alpha))^{N_m - n_{1m}} + \\ (1 - Z_m^A) p_B^*(\beta_1, \alpha)^{n_{1m}} (1 - p_B^*(\beta_1, \alpha))^{N_m - n_{1m}} \end{pmatrix} \tag{6}$$

This is the "complete data" probability of the observation. Of course, the Z^As are not observed so we have to specify how an equilibrium is selected. The simplest assumption is that in every repetition of the game equilibrium A is played with probability λ i.e.[8],

$$Z_m^A \sim \text{Bernoulli}(\lambda) \tag{7}$$

so that the "incomplete data" probability that n_{1m} choose action 1 is

$$\Pr(n_{1m}|\beta_1, \alpha, \lambda, N_m) = \frac{N_m!}{n_{1m}!(N_m - n_{1m})!} \begin{pmatrix} \lambda p_A^*(\beta_1, \alpha)^{n_{1m}} (1 - p_A^*(\beta_1, \alpha))^{N_m - n_{1m}} + \\ (1 - \lambda) p_B^*(\beta_1, \alpha)^{n_{1m}} (1 - p_B^*(\beta_1, \alpha))^{N_m - n_{1m}} \end{pmatrix} \tag{8}$$

This is the probability mass function (pmf) of a binomial mixture model with two components p_A^* and p_B^* which are the equilibrium strategies supported by β_1 and α. If β_1 and α support only a single equilibrium then (8) is equivalent to (5). Of course, the probability that equilibrium A is played may not be equal across repetitions.[9] For example, there may be factors such as laws, social conventions or previous play history which are not explicitly modelled that make equilibrium A more likely to

[8] Bjorn and Vuong (1985), Kooreman (1994), Ackerberg and Gowrisankaran (2002) and Bajari et al. (2004) also parameterize equilibrium selection using λ. The first two papers assume a value of λ for the purposes of estimation while the latter two papers estimate λ.

[9] Incorrectly assuming that the probability that equilibrium A is played is constant across repetitions can yield inconsistent estimates if this probability is different across repetitions with different numbers of players.

be played in certain repetitions. In Section 5 I estimate a basic model assuming that λ is constant across radio markets but then use the fact that I have repeated observations from the same market to estimate a model with market-specific λ_ms.

3 Identification

I now provide the identification results, explaining how multiple equilibria identify the parameters and how the data identify multiple equilibria.

3.1 Preliminaries

The data generating process is described in Section 2.2. The parameter space is $(\beta_1, \alpha, \lambda)$ with $-\infty \leq \beta_1 \leq \infty$, $\alpha \geq 0$ and $0 \leq \lambda \leq 1$. The sample space is (N_m, n_{1m}) from repetitions of the game indexed by $m = 1, ..., M$ where $N_m \geq 1$ and $n_{1m} \geq 0$. I define μ_N as the proportion of observations where $N_m = N$. I use the following definition of identification for a vector of parameters (β_1, α).

Definition (Identification of β_1 and α). *(β_1, α) are separately identified in this model if and only if for any pair (β_1', α'),*

$$\Pr(n_{1m}|\beta_1, \alpha, \lambda, N_m) = \Pr(n_{1m}|\beta_1', \alpha', \lambda', N_m) \ \forall N_m, n_{1m} \ \exists \lambda, \lambda' \tag{9}$$

implies that $(\beta_1, \alpha) = (\beta_1', \alpha')$.

3.2 Identification Results

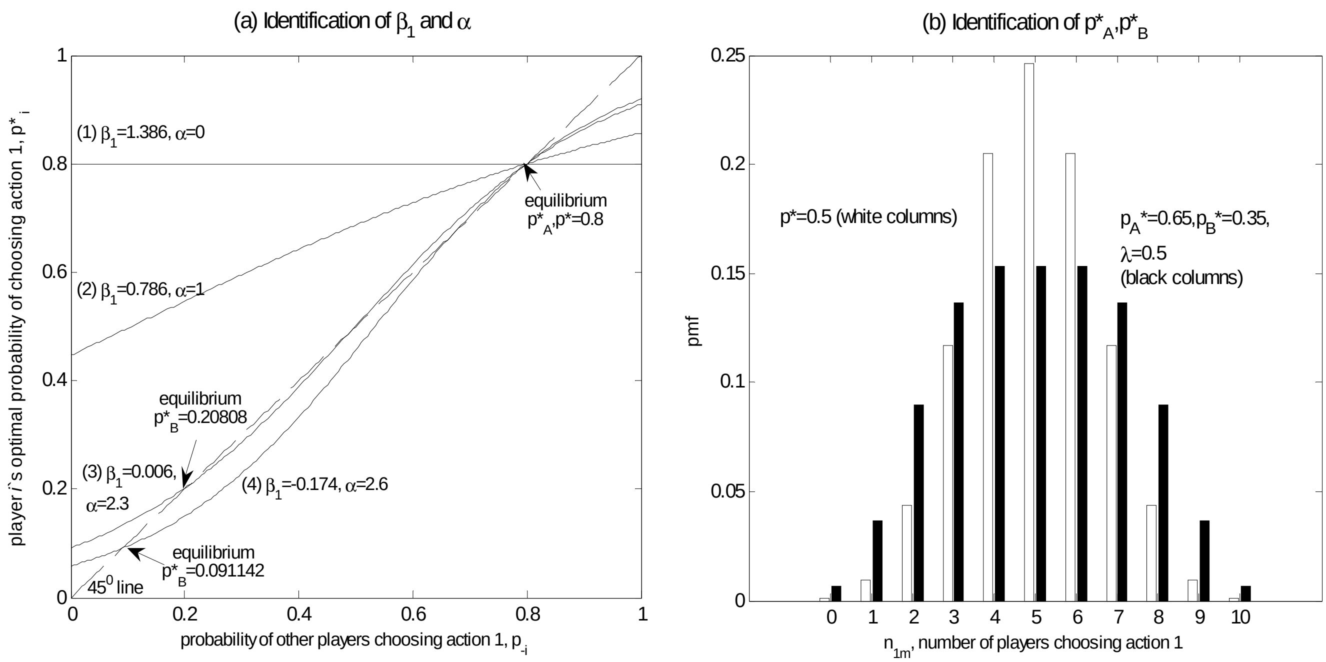
Parameter vectors (β_1, α) which support a single equilibrium are not separately identified.

Proposition 1. *All parameter vectors (β_1, α) where (β_1, α) support only one equilibrium are not separately identified.*

Proof. See Appendix A. ■

The proof involves showing that for any (β_1, α) which support a single equilibrium p^* we can find a different (β_1', α') which support the same p^* so that, from (5), the probability of every observation will

Figure 4: Identification



be the same. The intuition can be seen in Figure 4(a) where ($\beta_1 = 1.386$,$\alpha = 0$) and ($\beta_1 = 0.781$,$\alpha = 1$) support the same equilibrium $p^* = 0.8$. An alternative intuition is that when there is a single equilibrium we have only one equation (4) with two unknowns, β_1 and α. While β_1 and α are not separately identified the single equilibrium strategy p^* is identified.

If (β_1, α) support two stable equilibria then they are separately identified under two additional conditions.

Condition 1. *Some observations are generated from each equilibrium, i.e.,* $0 < \lambda < 1$.

Condition 2. *Some repetitions of the game have at least three players, i.e.,* $\sum_{j=3}^{\infty} \mu_j > 0$.

Proposition 2. *Parameter vectors* (β_1, α) *which support two distinct stable equilibria are separately identified if Conditions 1 and 2 hold.*

Proof. See Appendix A. ■

The intuition for Proposition 2 can also be seen in Figure 4(a). ($\beta_1 = 0.006, \alpha = 2.3$) and ($\beta_1 = -0.174, \alpha = 2.6$) both support $p_A^* = 0.8$ as an equilibrium but they support different p_B^*s. Therefore if we observe that $p_A^* = 0.8$ and $p_B^* = 0.20808$ then this is consistent with ($\beta_1 = 0.006, \alpha = 2.3$) and not consistent with ($\beta_1 = -0.174, \alpha = 2.6$), ($\beta_1 = 1.386, \alpha = 0$) or ($\beta_1 = 0.786, \alpha = 1$). The proof shows that any two distinct (β_1, α) pairs can support at most one equilibrium which is the same. Therefore if we can identify p_A^* and p_B^*, which requires Conditions 1 and 2, then we can identify β_1 and α. The proof also implies that if we can identify a non-stable equilibrium being played together with a stable equilibrium then this can also provide identification (i.e., my assumption that only stable equilibria are played is not necessary for identification). In identifying p_A^* and p_B^*, the components of the binomial mixture, we can also identify the incidental parameter λ.

Figure 4(b) illustrates how multiple equilibria are identified from the data. Suppose that $N_m = 10$. The white bars show the probability mass function (pmf) of n_{1m} for a single equilibrium $p^* = 0.5$ and the black bars show the pmf for $p_A^* = 0.65, p_B^* = 0.35$ and $\lambda = 0.5$. The expectation of n_{1m} is the same in both cases but with multiple equilibria the variance is greater with high or low values of n_{1m} having greater probability. As a binomial distribution has a fixed relationship between mean and variance, excess variance can only be created by a mixture of binomials. It is appropriate to consider what might be observed if $\alpha < 0$ so each player wants to choose a different action to the majority of other players. In this case there is one symmetric equilibrium but there may also be asymmetric equilibria where some players choose action 1 with relatively high probability and other players choose action 0 with relatively high probability. Asymmetric equilibria would generate pmfs with lower variance (very low probability that all players choose the same action) than the pmf from a single symmetric equilibrium.

3.3 Further Comments on Identification

The identification results describe how the existence of multiple equilibria in data from multiple repetitions of a two action game can identify the parameters. It is appropriate to make some additional comments about these results and how they generalize to more complicated settings.

1. **Logit Estimation**. Suppose one tried to estimate the parameters using a simple logit specification $\Pr(I_{im} = 1) = \frac{e^{(\beta_1-\alpha)+2\alpha\widetilde{p}_{-im}}}{1+e^{(\beta_1-\alpha)+2\alpha\widetilde{p}_{-im}}}$ where $I_{im} = 1$ if player i in repetition m chooses action 1. This is a correct specification for estimating β_1 and α if and only if $\widetilde{p}_{-im}$ is equal to the equilibrium strategies of the other players in repetition m (e.g., p_A^* in equilibrium A). In this case, if a single equilibrium

is played in every repetition then $\widetilde{p}_{-im}$ is just a constant and β_1 and α are not separately identified. On the other hand, if there are multiple equilibria then β_1 and α are separately identified because $\widetilde{p}_{-im}$ varies across repetitions. Note that β_1 and α cannot be consistently estimated by simply using the *observed* proportion of other players choosing action 1 as $\widetilde{p}_{-im}$ because this mismeasures their strategies.[10] Therefore consistent estimation of β_1 and α requires estimation of the full model which combines the data and the model to find what the equilibrium strategies are.

2. **More than 2 actions**. The results extend to the game with more than two actions. To be specific, suppose that players choose one of T actions, so that there are $T-1$ β_t parameters and one α parameter and player i's payoff from choosing action t is given by (1) where ε_{it} is Type I extreme value. If the true parameters support only one set of equilibrium choice probabilities then $(\beta_1, .., \beta_{T-1}, \alpha)$ cannot be separately identified but the parameters are separately identified if two or more sets of equilibrium choice probabilities can be identified: the number of required equilibria does not increase with the number of actions or parameters.[11] The parameters are therefore identified (or overidentified) if two or more sets of equilibrium choice probabilities can be identified in a T action game which may have more than 2 stable equilibria. This problem is equivalent to identifying the components of a multinomial mixture model with a possible E_T components where E_T is the maximum number of stable equilibria in the T choice game. A necessary condition is that there are repetitions of the game with $2E_T - 1$ players (Kim (1984) and Elmore and Wang (2003)).

3. **Interpretation of a single equilibrium**. Estimation can give parameters which are in the region of the parameter space which only supports a single equilibrium.[12] An important point to note is that while α must be sufficiently large to generate multiple equilibria, a single equilibrium is consistent with any value of α. To be precise a single equilibrium p^* is consistent with any β_1 and α which satisfy $\beta_1 + \alpha(2p^* - 1) - \log\left(\frac{p^*}{1-p^*}\right) = 0$ so a single equilibrium places no bound on α.

[10] The mismeasurement becomes less of a problem when the number of players is large. Brock and Durlauf (2001) propose the use of logit models to estimate social interactions in reasonably large populations.

[11] This reflects, at least in part, the Type I extreme value distribution of the εs. This distribution implies that the relative choice probabilities of actions t and T (with β_T normalized to zero) depend only on the parameters β_t and α. Consequently two sets of equilibrium choice probabilities identify α and all of the β_ts. The Type I extreme value assumption is not necessary in the two action model where other continuous distributions, such as the normal distribution, give the same results. The Type I extreme value assumption is, however, computationally convenient in the two action model.

[12] Feng and McCulloch (1996) show that maximum likelihood estimates converge to the non-identified subset of the parameter space containing the true parameters if the data is generated from a single distribution (here, a single equilibrium) rather than a true mixture (here, multiple equilibria).

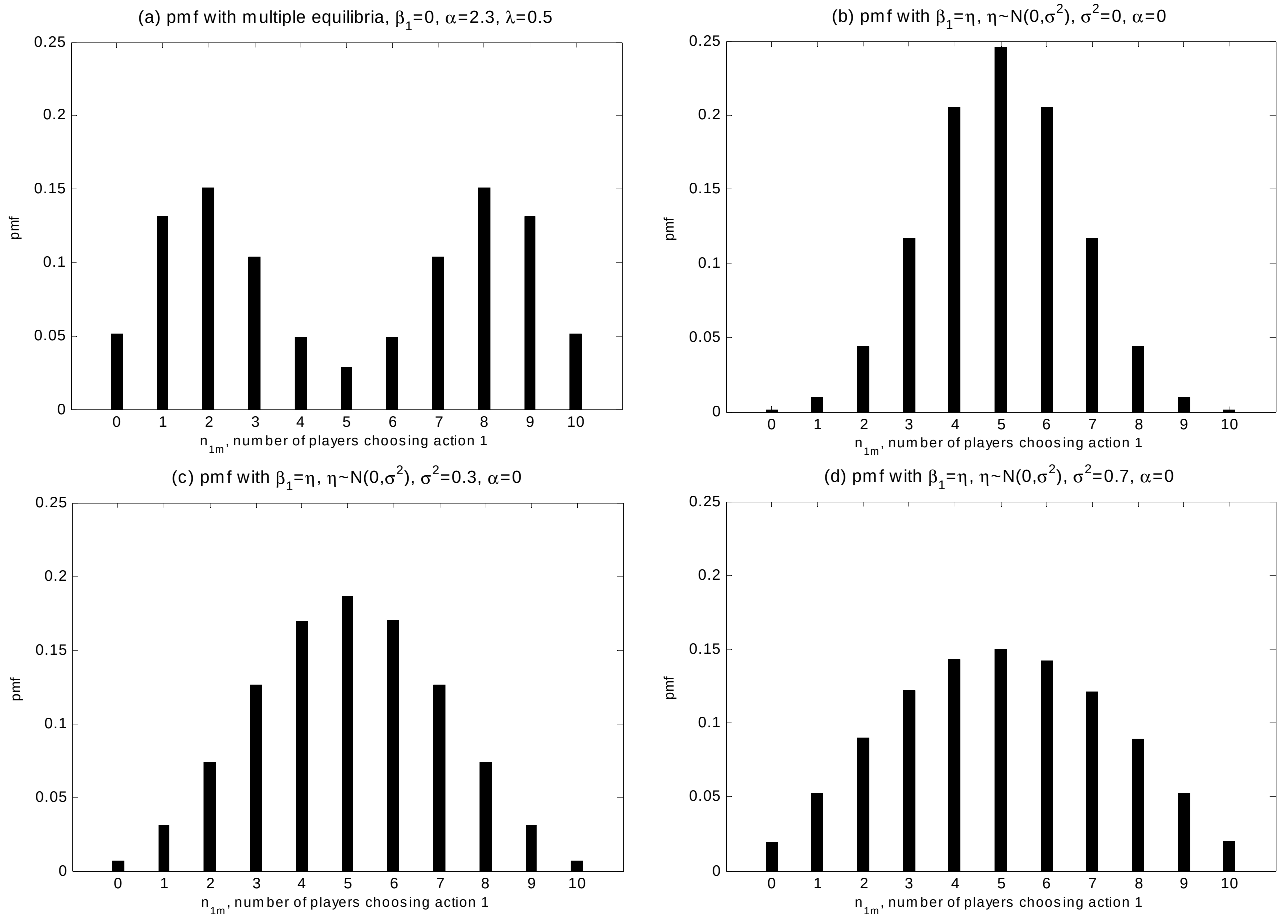
4. **Heterogeneity in β_1.** An assumption of the model is that β_1 is constant across repetitions of the game. Heterogeneity in β_1 can also explain why there are more repetitions in which action 1 is chosen by very many or very few players than predicted by a model with a single equilibrium where β_1 is fixed. Whether one is willing to assume that β_1 is constant across repetitions or varies only with covariates (in Section 5.5 I estimate a model where β_1 can vary with a measure of observed traffic patterns and α varies with covariates and unobserved heterogeneity) will depend on the situation. For example, a key common factor which affects the timing of radio commercials is Arbitron's methodology for estimating station ratings and the same methodology is used in every market. Similarly one would probably be willing to assume that the decision of people in England to drive on the left and people in the USA to drive on the right is not explained by country-specific preferences for one side of the road. In any event, multiple equilibria combined with a strong incentive to coordinate can generate patterns in the data which cannot be explained by some plausible forms of heterogeneity. This can be seen in a specific example. Figure 5(a) shows the pmf generated by the model with $\beta_1 = 0, \alpha = 2.3$ and $\lambda = 0.5$ when each repetition of the game has 10 players. Figure 5(b)-(d) show pmfs of the model with $\beta_{1m} = \eta_m, \alpha = 0$ and $\eta_m \sim N(0, \sigma^2)$ for various values of σ^2.[13] Increasing σ^2 increases the variance in the pmf but it cannot generate the twin-peaked pmf which characterizes the model without heterogeneity but with multiple equilibria and a strong incentive to coordinate. Of course, the pmf in Figure 5(a) would be matched by a model where $\alpha = 0$ and $\beta_1 = 1.3621$ with probability 0.5 and $\beta_{1m} = -1.3621$ with probability 0.5 but in many situations one might regard this kind of two-point heterogeneity as implausible. In the radio timing data the degree of clustering is modest and I maintain the assumption that β_1 is constant or depends only on a covariate.

4 Estimation and Testing for Multiple Equilibria

This section describes how I estimate the model using maximum likelihood (ML) and how I test for whether there is statistically significant evidence of multiple equilibria in the data.

[13] These pmfs are estimated using 10,000 simulation draws of η_m.

Figure 5: Multiple Equilibria vs. Heterogeneity in β_1



4.1 Estimation with a Single Equilibrium

It is straightforward to estimate the model if it is assumed that there is a single equilibrium. Ignoring the binomial coefficient which does not depend on the parameters the log-likelihood is

$$\ln L = \sum_{m=1}^{M} n_{1m} \ln p^*(\beta_1, \alpha) + (N_m - n_{1m}) \ln(1 - p^*(\beta_1, \alpha)) \tag{10}$$

and $p^*(\beta_1, \alpha)$ can be estimated by $\left(\frac{\sum_{m=1}^{M} n_{1m}}{\sum_{m=1}^{M} N_m}\right)$. As explained in Section 3 β_1 and α are not separately identified but if α is arbitrarily set equal to zero then

$$\widehat{\beta_1} = \ln\left(\frac{\sum_{m=1}^{M} n_{1m}}{\sum_{m=1}^{M} N_m - n_{1m}}\right) \tag{11}$$

4.2 Estimation with Multiple Equilibria

Ignoring the binomial coefficient the incomplete data log-likelihood is

$$\ln L = \sum_{m=1}^{M} \ln \begin{pmatrix} \lambda p_A^*(\beta_1, \alpha)^{n_{1m}} (1 - p_A^*(\beta_1, \alpha))^{N_m - n_{1m}} \\ +(1-\lambda) p_B^*(\beta_1, \alpha)^{n_{1m}} (1 - p_B^*(\beta_1, \alpha))^{N_m - n_{1m}} \end{pmatrix} \tag{12}$$

There are two ways to maximize (12). The first method is to maximize (12) directly with respect to the parameters $(\beta_1, \alpha, \lambda)$.[14] For each value of (β_1, α), p_A^* and p_B^* can be found using an iterative fixed point procedure. Even if there are multiple equilibria in the data the log-likelihood has a set of local maxima in the region of parameter space which supports a single equilibrium, so to find a maximum where (β_1, α) support multiple equilibria it is necessary to start the estimation procedure using good starting values. Therefore a coarse grid search over the region of the parameter space which supports multiple equilibria is used first. Feng and McCulloch (1996) show that ML estimates converge to the non-identifiable subset of the parameter space containing the true parameters if the data is generated by a single component (here, a single equilibrium p^*) rather than a true mixture (multiple equilibria with $0 < \lambda < 1$). A second method involves estimating p_A^*, p_B^* and λ as the parameters of a two component binomial mixture model and then solving two equations of the form $p^* = \frac{e^{\beta_1 + \alpha p^*}}{e^{\beta_1 + \alpha p^*} + e^{\alpha(1-p^*)}}$ to estimate β_1 and α. p_A^*, p_B^* and λ can be estimated using the EM Algorithm (Dempster et al. (1977)) which is widely used to estimate mixture models and which uses only analytic formulae. The details of the EM Algorithm used are given in Appendix B. This method is potentially quicker because it avoids using the fixed point procedure for each value of the parameters. However it is possible that the estimated p_A^* and p_B^* cannot be supported as stable equilibria for any values of β_1 and α. If so, it is necessary to use the first method.[15]

4.3 Testing for Multiple Equilibria

Identification of β_1 and α requires there to be multiple equilibria in the data so it is necessary to test for whether there is statistically significant evidence of multiple equilibria. This is equivalent

[14] The likelihood is discontinuous for values of β_1 and α on the boundary between the regions where one and two equilibria are supported. For this reason I use a Nelder-Mead simplex minimization routine (fminsearch in MATLAB) which does not require the calculation of derivatives.

[15] For example, if p_A^* and p_B^* are distinct stable equilibria then it must the case that $p_A^* > \frac{1}{2} > p_B^*$.

to testing whether a binomial mixture model has one or two components where the economic model places the additional constraint that if there are two components then they must be supported as stable equilibria.[16] I use the likelihood ratio test statistic (LRTS) to compare the likelihood of the model which allows two equilibria with the more restricted model which allows only a single equilibrium. The LRTS does not have its standard χ^2 distribution in testing for the homogeneity of mixtures because two regularity conditions are violated: under the null hypothesis of a single component some parameters are not identified and λ may be on the boundary of its [0,1] parameter space. Chen and Chen (2001) show that the LRTS for a binomial mixture model has an asymptotic distribution which is equivalent to the distribution of the supremum of a centered Gaussian process with a specific covariance which is a continuous function of the single binomial probability under the null hypothesis of a single component (here p^*). It is hard to estimate this distribution but Chen and Chen argue that because the distribution exists and p^* can be consistently estimated under the null hypothesis, it is appropriate to use a parametric bootstrap to estimate the critical values of the LRTS.[17] I use the following bootstrap testing procedure:

1. use the actual data to estimate $\widehat{p^*}$ and the value of the log-likelihood under the null hypothesis of a single equilibrium;

2. use the actual data to estimate $\widehat{\beta_1}, \widehat{\alpha}, \widehat{\lambda}$ and the log-likelihood under the alternative hypothesis that there may be two stable equilibria. The LRTS for the actual data is calculated;

3. use $\widehat{p^*}$ as the binomial choice probability to create a new set of data, under the assumption that there is only one equilibrium, with the same number of repetitions as the actual data and the same number of players in each repetition. Repeat steps 1 and 2 using this data.

4. repeat step 3 B times (I use $B = 249$). The jth-order statistic of the test statistics calculated in step 3 estimates the $\frac{j}{(B+1)}$th quantile of the distribution of the LRTS under the null hypothesis (McLachlan and Peel (2000), p. 193).

[16] Additional issues arise in testing for multiple equilibria in models with more than two possible equilibria. For example, certain testing criteria might reject a model with two equilibria in favor of a model with three equilibria but also reject a model with three equilibria in favor of a model with one equilibrium. The resolution of issues of this kind is beyond the scope of this paper.

[17] Lemdani and Pons (1997), Theorem 3, give the form of the asymptotic distribution where N_m varies across observations.

It is relatively computationally expensive to estimate $\widehat{\beta_1}, \widehat{\alpha}$ and $\widehat{\lambda}$ if the binomial components p_A and p_B which fit the data best cannot be supported as stable equilibria. This happens frequently when the data is generated from a single equilibrium as, of course, is the case for the bootstrapped data. I therefore use a less expensive method to calculate the LRTS for the bootstrap replications. This fits a two component binomial mixture (p_A, p_B, λ) under the alternative hypothesis without imposing the constraint that p_A and p_B must be supported as stable equilibria. This gives a log-likelihood which is weakly greater than if I imposed the constraint. The constraint is imposed in calculating the LRTS for the actual data and the same model is estimated for the actual and bootstrap data under the null hypothesis of a single equilibrium (any probability strictly greater than zero and strictly less than one can be supported as a stable equilibrium in a one equilibrium model). As a result my assessments of the significance of multiple equilibria are conservative i.e., I am less likely to reject the null of a single equilibrium. Appendix B presents some simulation results regarding the performance of the bootstrap in testing for the homogeneity of binomial mixtures and the extent to which my tests are conservative.

5 The Timing of Commercial Breaks by Music Radio Stations

I estimate the model using data on the timing of commercials by contemporary music radio stations. Section 5.1 explains the relevance of the model to stations' timing decisions, Section 5.2 describes the data and Section 5.3 presents some summary statistics and the results of estimating some simple logit models. Section 5.4 presents the empirical results from estimating the basic model described above and Section 5.5 presents the results from two extensions.

5.1 Listener Behavior and Station Timing Decisions

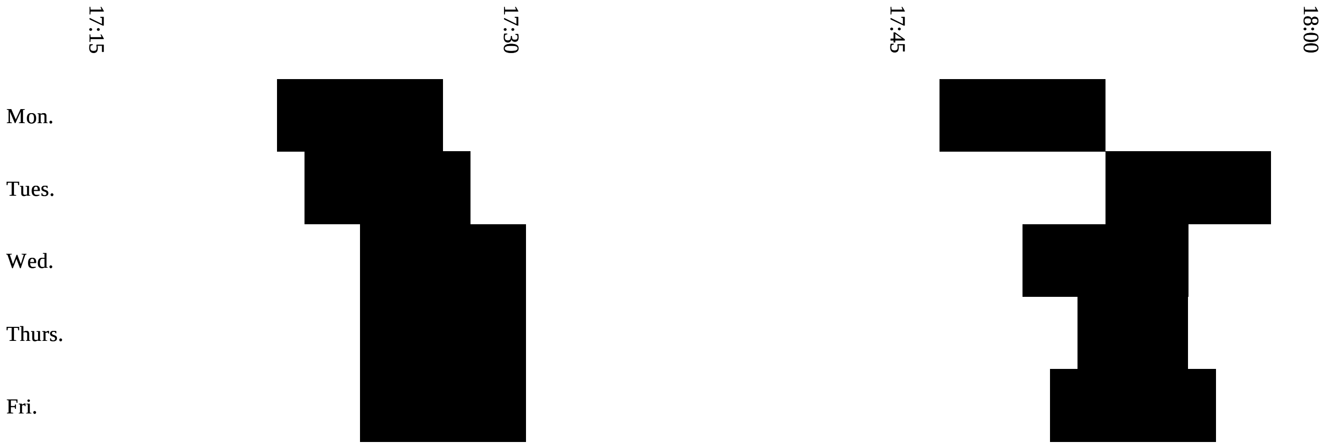
Section 2 presented an incomplete information discrete choice game where a player's payoff from choosing a particular action t was given by $\beta_t + \alpha P_{-it} + \varepsilon_{it}$. I now explain the relationship between this game and station timing decisions where an action is having a commercial break in a particular time interval.

While actual time is continuous the scheduling of commercials on music stations has a strong element of discreteness because it involves planning the order of songs and commercial breaks, so that, for example, a programmer decides whether to have a commercial break with one or two songs remaining in the hour (Warren (2001), p. 27 and Lynch and Gillespie (1998), p. 111 provide sample schedule "clocks"). The three terms in the payoff function can be rationalized in the following way. β_t allows some times to be more desirable for commercials than others. Some times may be more desirable because of how Arbitron estimates station ratings and because many listeners tune-in at the start of each hour. I assume that β_t is the same across markets which is plausible but I do consider below whether differences in driving patterns across markets could also explain the results. αP_{-it}, with $\alpha \geq 0$, allows a station's payoff to increase with the proportion of other stations in i's market choosing t for a commercial break (P_{-it}). The simple model of listener behavior presented in the Introduction justifies why a station's audience for its commercials might increase with the proportion of other stations having commercials at the same time. α should increase with listeners' propensity to switch stations. ε_{it} allows each station to have idiosyncratic preferences over the timing of commercials. It is assumed to be private information to player i and to be independent and identically distributed across players and actions. The εs represent two features of station timing decisions. First, a programmer may have idiosyncratic preferences over scheduled timing arrangements because, for example, he wants to develop a reputation for having "travel on the 3s". Second, other programming, such as songs, travel news or competitions, can vary and be unpredictable in length and a station would not want to annoy its listeners by cutting short this programming in order to play commercials at precise times.[18] For this reason the exact time at which a station plays commercials tends to vary from day-to-day. Figure 6 illustrates this by showing the timing of commercials on a Boston Rock station during the hour 5-6 pm during one week in 2001.

A station's decision for an hour might be, for example, to have 14 songs with three commercial breaks placed after the 5^{th}, 8^{th} and 12^{th} songs. Estimating a model with this kind of choice is well beyond the current literature and the scope of this paper. I therefore estimate the model by analyzing a simple choice. I define two timing choices during the last part of the hour which is the part with the most commercials (see Figure 1). I then look at those stations which choose one of these choices (very

[18] Warren (2001) p. 24 describes how sweeping the quarter-hours "can be done some of the time. But it can't be done consistently by very many stations. Few songs are 2:30 minutes long any more". A station manager confirmed this as an accurate description of the situation.

Figure 6: The Timing of Commercials on WROR-FM October 29-November 2 2001, 5-6pm



Note: based on airplay data described in Section 5.2. Shaded areas are commercial breaks.

few choose both) and examine which one they choose to estimate a two action model. I present some simple logit specifications which show that focusing on this selected sample of stations and ignoring, for example, how many commercial breaks a station has during an hour, does not produce spurious evidence that there is clustering of commercials at different times in different markets.[19]

5.2 Data

The timing data is taken from airplay logs provided by Mediabase 24/7 which electronically monitors stations to collect data on music airplay. Table 1 shows an extract of an airplay log for a Classic Hits station. I describe the set of stations in the Mediabase sample before explaining how I use the logs to analyze station timing decisions.

5.2.1 Coverage of the Mediabase Sample

I have logs for 1,094 contemporary music stations. BIAfn's *MediaAccess Pro* database identifies each station's home metro-market (defined by Arbitron) and its music category in each ratings quarter. The sample stations are home to 147 different metro-markets although 14 of these markets have only one sample station. The stations are in seven contemporary music categories: Adult Contemporary,

[19] Formally, the incentive to coordinate could be consistently estimated in this model by just focusing on players' choices over two particular actions if the expected proportion of players choosing either of these actions is the same across repetitions of the game. In fact there are some radio markets where relatively few stations have commercials at the end of the hour and ignoring this tends to lead me to underestimate the degree of clustering in the data.

Table 1: Extract from a Daily Log of a Classic Hits (Rock) station

Time	Artist	Title	Release Year
5:00PM	CLAPTON, ERIC	Cocaine	1980
5:04PM	BEATLES	While My Guitar Gently Weeps	1968
5:08PM	GRAND FUNK	Some Kind of Wonderful	1974
5:12PM	TAYLOR, JAMES	Carolina in My Mind	1976
5:16PM	RARE EARTH	Get Ready	1970
5:18PM	EAGLES	Best of My Love	1974
Stop Set	*BREAK*	*Commercials and/or Recorded Promotions*	-
5:30PM	BACHMAN-TURNER	Let It Ride	1974
5:34PM	FLEETWOOD MAC	You Make Loving Fun	1977
5:38PM	KINKS	You Really Got Me	1965
5:40PM	EDWARDS, JONATHAN	Sunshine	1971
5:42PM	ROLLING STONES	Start Me Up	1981
5:46PM	ORLEANS	Dance with Me	1975
Stop Set	*BREAK*	*Commercials and/or Recorded Promotions*	-
5:56PM	JOEL, BILLY	Movin' Out (Anthony's Song)	1977

Album Oriented Rock/Classic Rock, Contemporary Hit Radio/Top 40, Country, Oldies, Rock and Urban.[20] A music category aggregates similar music formats. For example, BIAfn classifies the Classic Hits format station in Table 1 in the Rock category. The sample does not include every station in these categories in the 147 markets but, as shown in Table 2, it does include stations which account for the majority of listenership especially in the largest markets and in categories other than Oldies. Stations in different categories are treated symmetrically in the model. This is appropriate as the available evidence suggests that listeners switch as much between stations in different music categories as between stations in the same category.[21] The prevalence of cross-category switching may reflect listeners' taste for music variety as well as the fact that stations classified in the same category may play quite different kinds of music (see Sweeting (2004b) for an analysis).

I use data from the first five weekdays of each month in 2001. The panel is unbalanced because Mediabase's sample of stations and markets expands over time and many individual station-days are

[20] The music categories that are not in the sample are Classical, Easy Listening, Jazz and Nostalgia/Big Band which appeal to older listeners than contemporary music categories. I drop station-quarters where a station in the airplay sample is categorized outside of the seven categories. This only affects two stations.

[21] For example, Arbitron Company (2003) estimates the number of listeners each pair of stations in the Boston market had in common in Fall 2002 (one year after the end of the airplay data). There were 6 Rock stations and 8 non-Rock contemporary music stations which had been in the airplay sample in Fall 2001. On average, 15.8% of the listeners to a Rock station listened to each of the other Rock stations and 17.1% of the listeners to a Rock station listened to each of the non-Rock stations. Rock listeners were more likely to listen to other contempary music stations than non-contemporary music stations apart from news/talk station WBZ-AM which is the largest station in Boston.

Table 2: Coverage of Airplay Sample Stations by Music Category

Contemporary Music Category	Number of Airplay Metro-Markets	Number of Home Market Rated Stations	Number of Home Market Rated Airplay Stations	Average % of Fall 2001 Home Listening Covered by Airplay Sample
Arbitron Metro-Markets Ranked 1-70 (1 is New York and 70 is Ft. Myers, FL)				
All categories	69	1003	720	86.1
Adult Contemporary	66	221	162	89.2
AOR/Classic Rock	65	111	98	95.9
CHR/Top 40	64	131	112	95.6
Country	64	141	94	92.1
Oldies	44	64	44	92.1
Rock	61	147	122	94.0
Urban	44	133	88	86.0
Arbitron Metro-Markets Ranked 71 and above (71 is Knoxville, TN)				
All categories	78	759	374	68.8
Adult Contemporary	56	135	78	78.7
AOR/Classic Rock	34	66	45	82.5
CHR/Top 40	59	96	75	91.4
Country	60	137	76	85.7
Oldies	1	3	1	40.7
Rock	42	81	60	86.7
Urban	27	58	39	86.2

Note: to understand how to read the table look at the Country entry for Arbitron metro-markets ranked 1-70. I have airplay data on at least one home-market Country station in 64 of these markets. In these 64 markets there are 141 rated Country stations with 94 of them in the airplay data. The airplay stations account for 92.1% of Country listening in the 64 markets (i.e., they have more listeners than the average station). "All categories" combines the seven categories. A rated station has a non-zero listening share listed by Arbitron.

missing so that there are 51,601 station-days of data.

5.2.2 Definition of Timing Choices

The logs identify the start time of each song and whether there was a commercial break between songs. I use the logs in two different ways to define the timing of commercials. The first way, which provides most of the results presented here, estimates the length of each song and then, assuming that commercial breaks fill the gaps between songs where "Commercial Breaks and/or Recorded Promotions" are indicated, identifies the median minute of the commercial break. I then examine stations which have commercial breaks with median minutes in the intervals :48-:52 or :53-:57 (but not both).[22] The sec-

[22] I estimate the length of each individual song using observations where the song is followed by another song without a commercial break. Its length is then defined as the median number of minutes between the start time of the song and the start of the next song. If a song is played less than 10 times without being followed by a commercial break then I assume that the song is 4 minutes long. I then form a minute-by-minute log assuming that each song is played its median

ond way uses the order of songs and commercial breaks. I examine stations which have a commercial break with one or two songs remaining in the hour (but not both).

There is measurement error in identifying which minutes have commercials because the logs do not identify non-music programming which may be placed between a song and a commercial break. If there is clustering on different times in different markets then as long as this measurement error is not correlated across stations within a market then the measurement error will tend to make the clustering less pronounced (less evidence of multiple equilibria) and the incentive to coordinate will be underestimated. The potential for measurement error is greater if a station plays only a few songs so I only use station-hours with at least 8 songs.

I focus attention on the hours 3-7 pm which Arbitron classifies as the afternoon drive. I do not use the morning drive because many stations have a lot of talk programming in the morning with more than half of all station-hours having less than 8 songs. On the other hand, less than 5% of the sample has less than eight songs during the afternoon drive. I also estimate the basic model for four randomly-chosen non-drivetime hours (3-4 am, 12-1 pm, 9-10 pm and 10-11 pm) when I expect the incentive to coordinate, α, to be weaker than during drivetime.

5.3 Summary Statistics and Logit Specifications

Table 3 presents summary statistics on station timing choices. Almost all station-hours have at least one commercial break except 3-4 am when 20% of station-hours are commercial-free. Part (a) of the table presents statistics based on allocating commercial breaks to 5 minute time intervals. Around 60% of station-hours have a commercial break in either the :48-:52 interval or the :53-:57 interval apart from 3-4 am when, once again, fewer stations have commercials. Roughly equal numbers of stations choose each interval in every hour. Part (b) of the table presents statistics based on the order of songs and commercial breaks. Fewer stations make each choice than in part (a) as most songs are shorter than 5 minutes. In both parts of the table the number of stations making both choices is much smaller than would be expected if the decisions about each choice were independent. As I estimate a discrete

length unless this would completely eliminate a commercial break in which case I assume that the song is shortened to allow at least one minute of commercials. I assume that commercial breaks fill the remaining gaps between songs where commercial breaks are indicated and identify the median minute of each break. The break is allocated to one of the time intervals :48-:52 and :53-:57 based on the median point of the break. If the median point is :52 minutes and 30 seconds then I allocate the break to the earlier (:48-:52) interval.

Table 3: Summary Statistics on Station Timing Choices at the End of the Hour

(a) Timing Defined by Median Minute of Commercial Break and Five Minute Time Intervals

	Number of station-hours with				
Hour	at least 8 songs	any break in hour	break in :48-:52	break in :53-:57	breaks in both :48-:52 & :53-:57
3-4 am	50,694	39,968	9,974	7,965	64
12-1 pm	50,664	49,300	15,431	13,915	87
3-4 pm	50,963	50,375	16,737	13,559	70
4-5 pm	50,617	49,879	15,916	14,592	88
5-6 pm	50,459	48,978	16,043	15,194	73
6-7 pm	50,694	50,003	16,794	13,106	74
9-10 pm	49,927	48,187	13,558	14,640	64
10-11 pm	48,619	46,116	13,106	13,459	58

(b) Timing Defined by Order of Songs and Commercial Breaks

	Number of station-hours with				
Hour	at least 8 songs	any break in hour	break with 2 songs left in hour	break with 1 song left in hour	breaks with 1 and 2 songs left in hour
3-4 am	50,694	39,968	7,955	6,874	88
12-1 pm	50,664	49,300	13,558	12,440	350
3-4 pm	50,963	50,375	14,974	13,093	481
4-5 pm	50,617	49,879	14,735	13,767	570
5-6 pm	50,459	48,978	14,705	13,959	521
6-7 pm	50,694	50,003	15,165	12,616	481
9-10 pm	49,927	48,187	12,131	14,020	783
10-11 pm	48,619	46,116	11,467	12,981	618

choice model where each player makes a single choice the stations making both choices are ignored in the estimation of the basic model.

Table 4 shows the results of estimating some simple logit models. The specification is $\Pr(I_{1ihmd} = 1|X, \gamma) = \frac{e^{X\gamma}}{1+e^{X\gamma}}$ where I_{1ihmd} is an indicator equal to 1 if station i in hour h in market m on day d chooses action 1 defined as having a commercial in the :53-:57 time interval (action 0 :48-:52) or a commercial break with one song remaining in the hour (action 0 two songs). The X variables include the proportion of other stations in i's metro-market m choosing action 1 on day d in hour h ($PROPORTION$). The coefficient on $PROPORTION$ will be positive if commercials are clustered at different times in different markets, although, as explained in Section 3.3, the coefficients cannot be used to form a consistent estimate of α.

In column (1) the X variables are $PROPORTION$, dummies for i's music category, dummies for the number of blocks of commercials i has during the hour (which can vary between 1 and 7 with a

Table 4: Logit Specifications

(a) Timing of Commercials Defined by Median Minute of Commercial Break

	(1)	(2)	(3)	(4)
PROPORTION				
Drivetime	0.3134 (0.0891)	0.2635 (0.0871)	0.2346 (0.0578)	0.1551 (0.0573)
Non-drivetime	0.1158 (0.0810)	0.0444 (0.0793)	-0.0065 (0.0452)	0.0006 (0.0432)
Dummies	Hour Day of Week Music Category Number of Breaks	Hour Day of Week Music Category Number of Breaks Hour * *COMMUTE* Hour * *MARKETRANK* Hour * Region Dummies	Hour	Hour Day of Week Music Category Number of Breaks Hour * *COMMUTE* Hour * *MARKETRANK* Hour * Region Dummies
Number of station-hours	376,607	376,607	212,420	212,420

(b) Timing of Commercials Defined by the Order of Songs

	(1)	(2)	(3)	(4)
PROPORTION				
Drivetime	0.3481 (0.0811)	0.3163 (0.0783)	0.1925 (0.0391)	0.1694 (0.0396)
Non-drivetime	0.1142 (0.0789)	0.0474 (0.0772)	-0.0748 (0.0375)	-0.0551 (0.0360)
Dummies	Hour Day of Week Music Category Number of breaks	Hour Day of Week Music Category Number of Breaks Hour * *COMMUTE* Hour * *MARKETRANK* Hour * Region Dummies	Hour	Hour Day of Week Music Category Number of Breaks Hour * *COMMUTE* Hour * *MARKETRANK* Hour * Region Dummies
Number of station-hours	376,607	376,607	184,980	184,980

Note: Standard errors in parentheses allow for correlation across observations from the same station across hours and across days. If only one station is observed in a market-day then observations on that station cannot be used as *PROPORTION* is not defined.

mean of 2.08 and standard deviation of 0.67), hour dummies and day of week dummies. In columns (1) and (2) I use all stations whether or not they have a commercial break in one of the two time slots both in estimation and in defining $PROPORTION$. The $PROPORTION$ coefficient is allowed to vary between drivetime (3-7 pm) and non-drivetime hours (the four other hours listed above). The standard errors are calculated to allow for correlation across choices by the same station across hours and across days. The $PROPORTION$ coefficients are positive and statistically significant at the 0.1% level during drivetime indicating that commercials are clustered at different times in different markets during drivetime. This is consistent with multiple equilibria. The coefficients are much smaller and statistically insignificant at the 10% level outside drivetime. The results are very similar for the two different definitions of when commercials are played.

The most obvious concern with interpreting clustering as reflecting multiple equilibria is that market-specific factors might also lead stations in different markets to have commercials at different times even if there is no incentive to coordinate. A candidate is local commuting patterns as clustering is more pronounced during drivetime. Arbitron provides an estimate of average one-way commuting times (for people not working at home) in each market on its website and I use this to create a variable $COMMUTE$.[23] The mean and median $COMMUTE$ is 24 minutes and the standard deviation across the sample markets is 3.3 minutes. Commuting patterns might also vary with the size of the market so I create a $MARKETRANK$ variable which is simply Arbitron's rank of the market, based on population, and which varies from 1 (New York City) to 222 (Muskegon, MI). They may also vary with region, so I create region dummies for 4 geographic regions (North East, South, Mid-West and West).[24] In column (2) I include $COMMUTE$, $MARKETRANK$ and the region dummies allowing their effects to vary by hour. The drivetime $PROPORTION$ coefficients fall only slightly and, in fact, they fall by less than the non-drivetime coefficients. This shows that clustering is not explained by those aspects of commuting patterns which are correlated with commuting time, market size or region. A station manager also said that while commuting patterns may affect how

[23] Arbitron's estimates are based on data from the long-form version of the 2000 Census. Arbitron's website is www.arbitron.com.

[24] The North East region includes markets in ME, NH, VT, MA, RI, CT, NY, PA and NJ. The Mid-West region includes OH, IN, MI, IL, WI, IA, MO, MN, KS, NE, ND and SD. The South region includes FL, GA, SC, NC, VA, WV, MD, DE, KY, TN, AL, MS, LA, AK, OK, TX and Washington DC. The West region includes the remaining states including Hawaii.

many commercials are played during an hour but they would be unlikely to affect a fine choice such as whether to have a break at 4:50 rather than 4:55 for which timing relative to other stations would be more important.

Columns (3) presents a specification which is close to the basic model which I estimate. Only stations choosing either action 0 or action 1 (but not both) are used in estimation and in defining $PROPORTION$. The column (3) specification includes only hour dummies. The $PROPORTION$ coefficients are positive and highly significant during drivetime and insignificant or negative outside drivetime. In column (4) I add the same controls as in column (2). The drivetime coefficients fall but remain statistically significant at the 1% level. The drivetime coefficients in columns (3) and (4) are smaller than those in columns (1) and (2) indicating that focusing on timing choices between actions 0 and 1 ignores some clustering as there are some markets in which relatively few stations choose either of these actions.[25] Therefore estimating the model using only stations which choose these actions will lead me to underestimate the incentive to coordinate.

5.4 Results from the Basic Model

Table 5(a) presents the results from estimating the model of Section 2 separately for each drivetime hour. Action 0 is defined as having a commercial in the :48-:52 time interval and action 1 is defined as having a commercial in the :53-:57 time interval. Stations not choosing either action are excluded from the estimation. Every market-day observation is treated as a separate and independent observation of the game and β_1,α and λ are assumed to be identical across market-days. Standard errors are calculated using a resampling bootstrap procedure. In each replication of the procedure markets are drawn, with replacement, from the data. Markets, rather than market-days, are sampled to allow for correlations across observations from the same market. The procedure is repeated 25 times.

The first set of results in each column show the estimates when it is assumed that the same equilibrium is played in every market-day observation. β_1 and α are not separately identified and the reported estimate of β_1 assumes that α is equal to zero. The coefficient is negative in hours

[25] For example, the drivetime $PROPORTION$ coefficient in column (2), median minute specification, implies that a 10% increase from the mean in the proportion of other stations in the market choosing action 1 is associated with a 0.515% increase in the probability of a station choosing action 1. The column (4) coefficient implies that a 10% increase in the proportion of other stations choosing action 1 is associated with a 0.379% increase in the probability of a station choosing action 1.

Table 5: Basic Model Results with Timing Defined by the Median Minute of a Commercial Break

(a) Afternoon Drivetime Hours

	3-4 pm	4-5 pm	5-6 pm	6-7 pm
One Equilibrium Model				
β_1 (assuming $\alpha = 0$)	-0.2116 (0.0450)	-0.0874 (0.0406)	-0.0546 (0.0424)	-0.2492 (0.0437)
Log-likelihood	-20734.8	-20995.6	-21539.0	-20393.3
Implied equilibrium p^*	0.447	0.478	0.486	0.438
Two Equilibria Model				
β_1	-0.0008 (0.0012)	-0.0007 (0.0012)	0.0008 (0.0009)	-0.0009 (0.0012)
α	2.0140 (0.0072)	2.0127 (0.0082)	2.0147 (0.0069)	2.0154 (0.0069)
λ	0.2582 (0.1165)	0.4735 (0.1512)	0.2127 (0.1502)	0.2419 (0.1016)
Log-likelihood	-20731.6	-20989.6	-21535.2	-20390.4
Implied equilibria p_A^*, p_B^*	0.534,0.417	0.541,0.421	0.585,0.459	0.517.0413
Joint-payoff maximizing p^{JP}	0.021	0.021	0.979	0.020
Test for Multiple Equilibria				
LRTS	8.3	12.1	7.7	5.8
90^{th},95^{th},99^{th} percentiles of LRTS distribution	3.1,4.3,6.1	3.6,5.7,7.6	2.7,4.4,6.8	2.7,4.3,8.1
Number of market-days	7,598	7,656	7,702	7,657
Number of station-days	30,156	30,332	31,091	29,752

(b) Non-Drivetime Hours

	3-4 am	12-1 pm	9-10 pm	10-11 pm
One Equilibrium Model				
β_1 (assuming $\alpha = 0$)	-0.2266 (0.0440)	-0.1040 (0.0412)	0.0771 (0.0377)	0.0267 (0.0372)
Log-likelihood	-12232.1	-20181.1	-19435.8	-18330.7
Implied equilibrium p^*	0.444	0.474	0.5193	0.507
Two Equilibria Model				
β_1	-0.2266 (0.0440)	-0.1040 (0.0412)	0.0001 (0.0003)	0.0267 (0.0372)
α	-	-	2.0041 (0.0039)	-
λ	-	-	-	-
Log-likelihood	-12232.1	-20181.1	-19435.6	-18330.7
Implied equilibria p_A^*, p_B^*	0.444	0.474	0.545,0.499	0.507
Test for Multiple Equilibria				
LRTS	0	0	0.3	0
90^{th},95^{th},99^{th} percentiles of LRTS distribution	3.2,4.8,7.1	2.3,3.3,7.2	3.0,4.5,7.6	3.1,4.5,9.1
Number of market-days	6,520	7,549	7,482	7,360
Number of station-days	17,811	29,172	28,070	26,449

Notes: Standard errors in parentheses. Log-likelihoods do not include binomial coefficients which are constants and independent of the parameters.

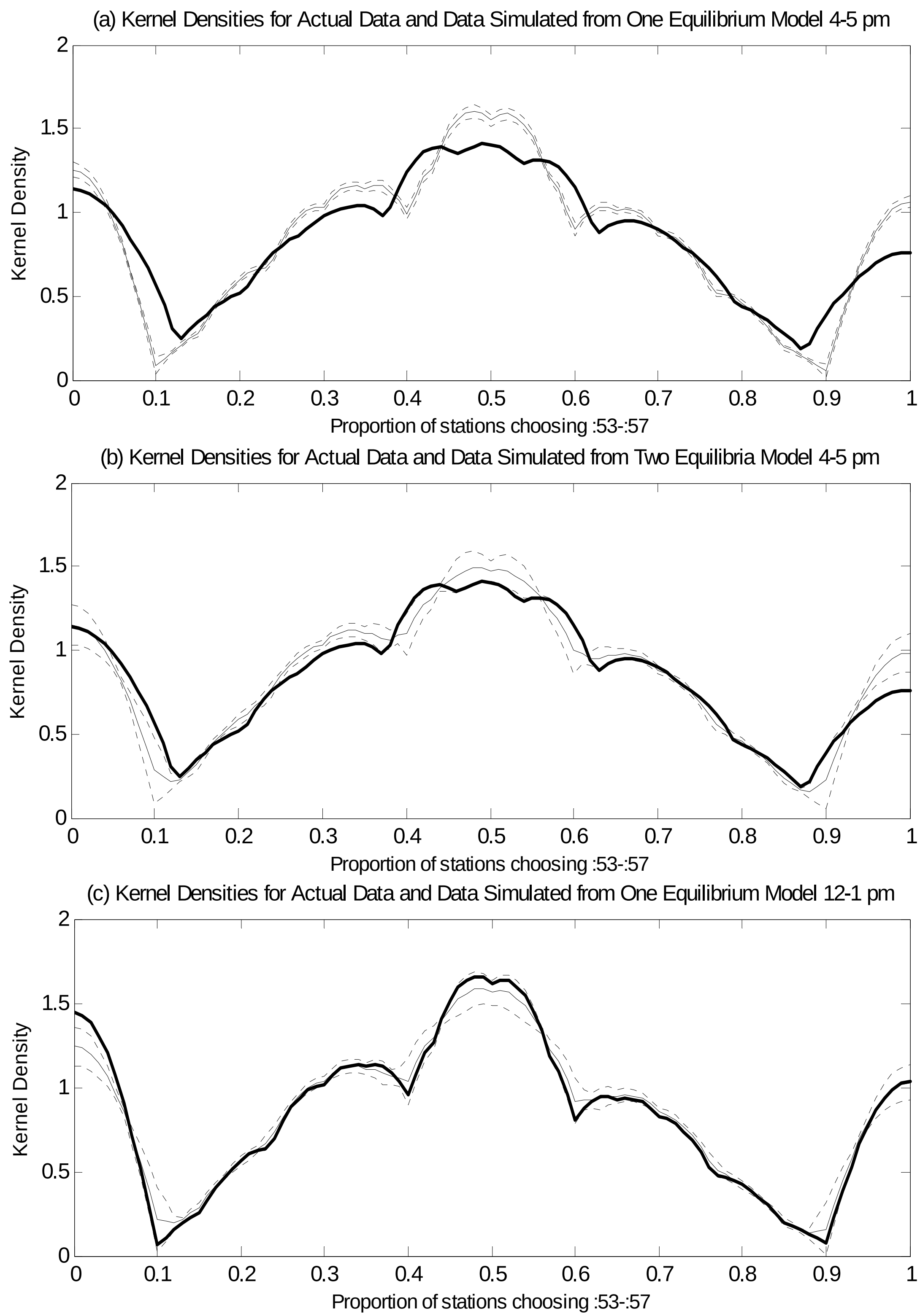
when more stations choose :48-:52 than :53-:57. The second set of results allow for two equilibria to be played. β_1 and α are separately identified if there are multiple equilibria in the data. The two equilibria model fits better than the one equilibrium model in each drivetime hour. This is reflected in the LRTS which is reported together with the estimated 90^{th}, 95^{th} and 99^{th} percentiles of the LRTS's distribution under the null hypothesis of a single equilibrium. Based on these percentiles, the LRTS is statistically significant at the 1% level in three of the four hours (3-6pm) and significant at the 5% level in the remaining hour (6-7 pm). The β_1 coefficient measures whether :53-:57 is more attractive for commercials independent of any incentive to coordinate. In none of the hours is the estimate of β_1 significantly different from zero which is consistent with both of these time intervals being equally distant from the quarter-hours which are known to be unattractive times for commercials. The magnitude of α is best understood by examining the implied equilibrium probabilities (p_A^*, p_B^*) of choosing :53-:57 for a commercial break. For example, the 4-5 pm coefficients imply that in equilibrium A stations choose :53-:57 with probability 0.541 (:48-:52 with probability 0.459) and in equilibrium B stations choose :53-:57 with probability 0.421 (:48-:52 with probability 0.579). The equilibrium degree of clustering is clearly modest in all of the drivetime hours.

Table 5(b) shows the results for the four non-drivetime hours. In three of the four hours the estimates of the two equilibria model imply that there is a single equilibrium in the data (LRTS equals zero). For 9-10 pm the two equilibria model fits slightly better than the one equilibrium model but the LRTS is only 0.3 and is not close to being statistically significant. Without evidence of multiple equilibria we cannot infer whether there is an incentive to coordinate during these hours. However, the presence of multiple equilibria during drivetime but not outside drivetime is consistent with the model only having multiple equilibria if the incentive to coordinate is strong enough because the incentive should be greater during drivetime when there are more in-car listeners.

In the spirit of Figure 4(b), Figure 7(a) and (b) illustrate why the data for 4-5pm support the existence of multiple equilibria but only modest clustering. The proportion of stations choosing :53-:57 is on the horizontal axis. The bold solid lines show kernel density estimates for the actual data.[26] The shape of the density is irregular because of the different numbers of stations observed in different

[26] The kernel density is estimated using an Epanechnikov kernel at 101 evenly spaced points between 0 and 1.

Figure 7: Proportion of Stations Choosing Action 1 (:53-:57) in Each Market-Day Observation in the Actual Data and in Data Simulated from the Estimated Models



market-day observations. In (a) the light solid line shows the mean kernel density from 50 sets of simulated data generated from the estimated one equilibrium model and the dashed lines indicate one standard deviation from this mean density. The density for the actual data lies below the density for the simulated data for most proportions between 0.25 and 0.7 and slightly above the simulated density for most of the remaining proportions, i.e., there are more observations where very high or very low proportions of stations choose :53-:57 than the single equilibrium model predicts. In (b) the light solid line shows the mean density for data simulated from the estimated two equilibria model. The density for the actual data is within one standard deviation of the mean density for the simulated data for almost all proportions. (c) repeats (a) using data for 12-1 pm. The density for the actual data lies slightly above the density for the simulated data for proportions between 0.4 and 0.6. This is consistent with the one equilibrium model fitting the data as well as the model which allows two equilibria.

The modest nature of equilibrium clustering during drivetime may seem surprising given the evidence on how many listeners switch stations to avoid commercials. Does this imply that the avoidance of commercials has little effect on the value of advertising time or industry revenues? It does not because of the presence of externalities in the timing game. Each station makes it timing decision without internalizing how its decision affects the payoff (audience) of other stations. This externality, combined with the fact that each station may find it privately costly to choose the time which it expects most of the other stations in its market to choose because commercials have to be fitted in around other programming (the εs in the model), leads to significantly less clustering in Nash equilibrium than there would be if the stations in a radio market maximized their joint payoffs. Calculating the strategies which maximize expected joint payoffs is straightforward under the assumption that the εs remain private information so a station's strategy can only depend on its own εs.[27] If other stations choose action 1 with probability p_{-i} then the expected benefit to other stations from i choosing action 1 is $(N_m - 1)p_{-i}\frac{\alpha}{N_m-1} = \alpha p_{-i}$. Similarly, if i chooses action 0 then the expected benefit to other

[27] Brock and Durlauf (2001), p. 3318, provide a similar calculation for a social planner's problem.

stations is $\alpha(1 - p_{-i})$. i maximizes expected joint payoffs by choosing action 1 if and only if

$$\beta_1 + 2\alpha p_{-i} + \varepsilon_{i1} \geq 2\alpha(1 - p_{-i}) + \varepsilon_{i0} \tag{13}$$

and the strategies which maximize expected joint payoffs will satisfy (p^{JP})

$$p^{JP} = \frac{e^{\beta_1 + 2\alpha p^{JP}}}{e^{\beta_1 + 2\alpha p^{JP}} + e^{2\alpha(1-p^{JP})}} \tag{14}$$

The difference between (14) and (4) is that α is replaced by 2α: joint payoff maximization effectively doubles the incentive to coordinate. More than one value of p^{JP} may satisfy (14) but if $\beta_1 \geq 0$ then the expected joint payoff maximizing strategy will have $p^{JP} \geq 0.5$. The joint payoff maximizing strategies implied by the estimates for the four drivetime hours are shown in Table 5(a). They involve almost perfect coordination with each station choosing the most attractive interval with probability around 0.980.

While this calculation relies on the functional forms assumed in the model it is interesting for three reasons. First, it shows that the modest observed clustering is not necessarily inconsistent with the claim that the value of radio commercials would be maximized by all stations playing commercials at the same time so that listeners could not avoid them. Second, it shows that the private costs associated with the precise timing of commercials may have a large effect on equilibrium timing patterns. It is interesting to note that television stations, which use more pre-recorded programming in which commercials can be placed quite precisely, appear to have commercials which overlap more than radio stations.[28] Third, it illustrates that Nash equilibrium outcomes in coordination games can be quite different from joint payoff maximizing outcomes even if, unlike in, for example, a price-setting game, there is no incentive for an individual player to undercut other players.

The results when the basic model is estimated using the order of songs are shown in Appendix C. The results are similar to those presented above with no significant evidence of multiple equilibria outside drivetime and significant evidence during drivetime except that the LRTS for 4-5 pm is marginally insignificant at the 10% level while the results for 5-6 pm indicate greater clustering of commercials

[28] Epstein (1998) provides empirical evidence on the timing of commercials by the major television networks during primetime.

with a more significant LRTS.

5.5 Enriching the Model

I extend the model with multiple equilibria in two stages. The first extension allows the proportion of observations from each equilibrium to vary across markets using the fact that the data contains multiple observations from the same market. I find that most markets stay in the same equilibrium (commercials clustered at the same time) over time. The second extension introduces covariates into the model and allows for unobserved market-specific heterogeneity in the incentive to coordinate. The incentive to coordinate is stronger in smaller markets and in markets with more concentrated station ownership.

5.5.1 Market-Specific λs

The basic model assumes that the probability that any observation is generated by equilibrium A strategies is the same. This implies that if there are multiple equilibria across markets then there are also multiple equilibria within markets whereas it seems plausible that if stations are trying to coordinate on timing then they would stay in the same equilibrium from day-to-day. This assumption can be relaxed by using the repeated observations from each market to estimate a market-specific $\lambda_m, 0 \leq \lambda_m \leq 1$.[29] I continue to assume that β_1 and α are constant across markets so that only the proportion of observations in each equilibrium varies. The log-likelihood, ignoring the binomial coefficients, is

$$\ln L = \sum_{m=1}^{M} \sum_{d=1}^{D_m} \ln \left(\begin{array}{c} \lambda_m p_A^*(\beta_1, \alpha)^{n_{1md}} (1 - p_A^*(\beta_1, \alpha))^{N_{md} - n_{1md}} \\ +(1-\lambda_m) p_B^*(\beta_1, \alpha)^{n_{1md}} (1 - p_B^*(\beta_1, \alpha))^{N_{md} - n_{1md}} \end{array} \right) \tag{15}$$

Even with over 140 markets the parameters can be estimated easily using maximum likelihood. The estimated values of β_1 and α and the implied equilibrium strategies are shown in Table 6 for the four drivetime hours when action 1 is defined as having a commercial in the :53-:57 time interval. Standard errors are, as before, calculated using a resampling bootstrap procedure which is repeated

[29] Consistency of the estimates requires that the number of observations per market tends to infinity. In the data there are up to 60 observations per market.

Table 6: Results With Market-Specific λ_m Equilibrium Selection Probabilities for Four Drivetime Hours

	3-4 pm	4-5 pm	5-6 pm	6-7 pm
β_1	-0.0085 (0.0014)	-0.0068 (0.0027)	-0.0076 (0.0020)	-0.0085 (0.0014)
α	2.0690 (0.0077)	2.0617 (0.0094)	2.0645 (0.0083)	2.0690 (0.0076)
Log-likelihood	-19917.3	-20124.4	-20733.6	-19612.6
Implied equilibria p_A^*, p_B^*	0.573,0.321	0.602,0.331	0.591,0.327	0.571,0.321
Joint-payoff maximizing p^{JP}	0.018	0.018	0.018	0.018
Number of market-days	7,598	7,656	7,702	7,657
Number of station-days	30,156	30,332	31,091	29,752

Notes: Standard errors in parentheses. Log-likelihoods do not include binomial coefficients which are constants and independent of the parameters.

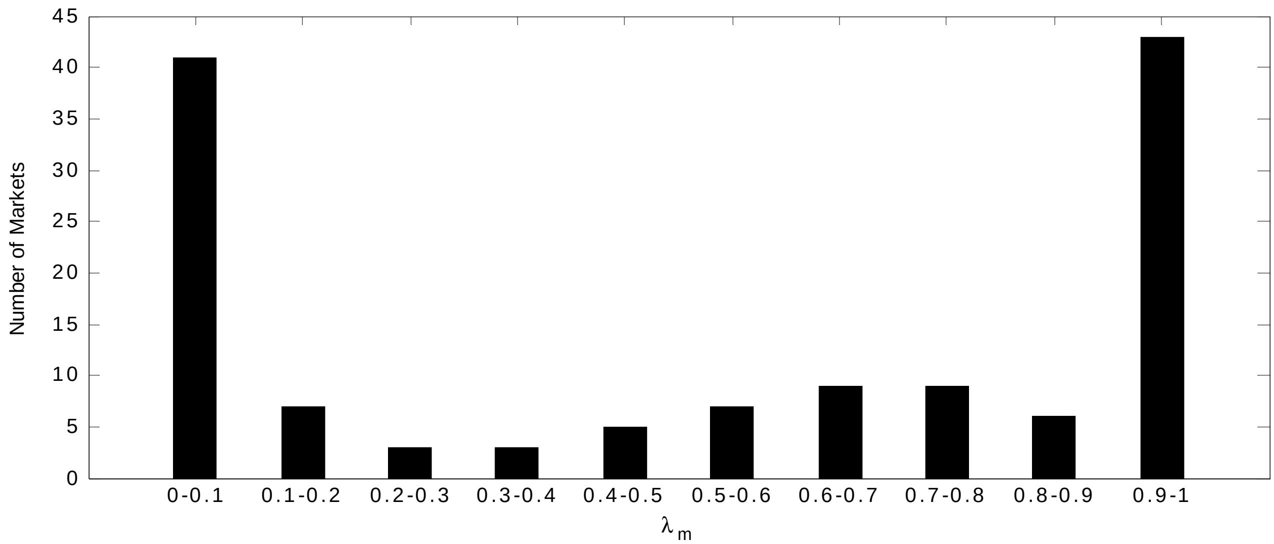
25 times. Comparison of the log-likelihoods in Tables 5 and 6 shows a sizeable improvement in the model's fit. The estimated values of α increase and the equilibrium strategies imply slightly more clustering of commercials than was implied by the basic model. For example, the 4-5 pm equilibrium A involves each station choosing :53-:57 with probability 0.602, compared with 0.541 in the basic model, and equilibrium B involves each station choosing :48-:52 with probability 0.669, compared with 0.579 in the basic model. The estimates of β are negative implying that :48-:52 is more attractive for commercials than :53-:57. As in the basic model joint payoff maximization would involve almost perfect coordination on timing.

Figure 8 shows the distribution of the estimated values of λ_m for 4-5 pm (the histograms for the other drivetime hours are similar) for markets with at least two stations in the sample. Approximately two-thirds of the $\widehat{\lambda}_m$s are less than 0.1 or greater than 0.9 which shows that most markets tend to have commercials clustered at the same time from day-to-day even if these times differ across markets.[30] Markets with $\widehat{\lambda}_m$ around 0.5 may either have commercials which are clustered but at different times on different days or they may have commercials which are always fairly evenly distributed across the two time intervals.

Figure 9 maps which markets have high or low $\widehat{\lambda}_m$s. Hollow shapes are markets with $\widehat{\lambda}_m > 0.5$ (more stations choose :53-:57). Stars represent markets with very high or very low values of $\widehat{\lambda}_m$. Three features of the map are of interest. First, the equilibria are not concentrated in different regions

[30] These results are consistent with estimates from logit specifications similar to those in Table 4 with station-hour fixed effects. In several specifications the coefficient on $PROPORTION$ is positive but statistically insignificant, and much smaller than the estimates in Table 4, during drivetime.

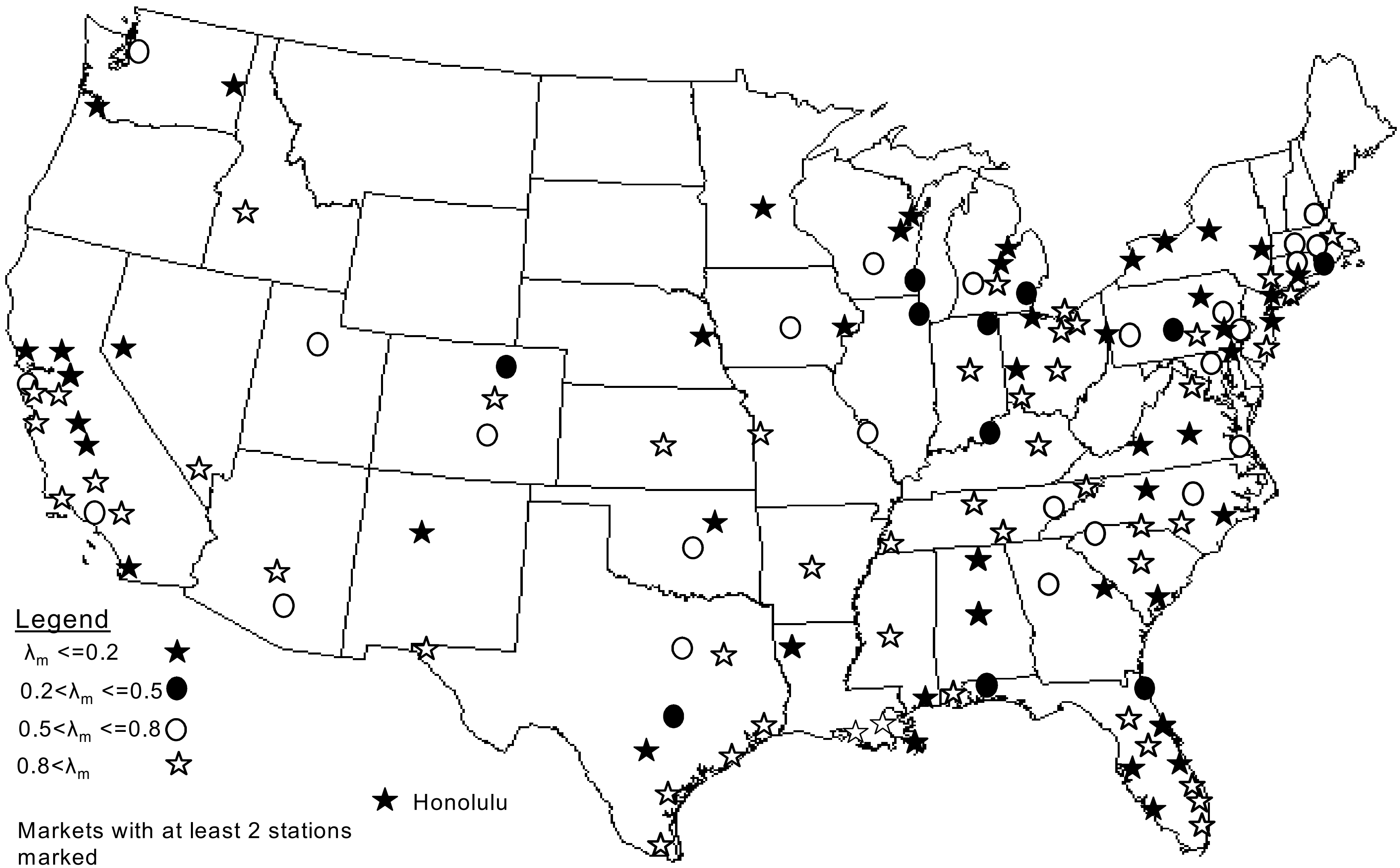
Figure 8: Distribution of $\widehat{\lambda_m}$ for 4-5 pm for Markets with at least Two Stations in the Sample



of the country. For example, states such as California, Florida, Ohio and North Carolina have some markets with $\widehat{\lambda}_m > 0.8$ and other markets with $\widehat{\lambda}_m \leq 0.2$. This suggests that factors such as different regional tastes for programming cannot explain why commercials are clustered at different times in different markets. Second, there appears to be some clustering at a very local level. For example, Akron-Canton-Cleveland, Appleton-Green Bay, Boston-Portsmouth-Worcester, Los Angeles-Riverside and San Francisco-San Jose are examples of pairs or groups of very close markets which are largely in the same equilibrium. In these markets many listeners in the smaller market may listen to stations in the larger market and, as a result, stations in the smaller market may want to choose the same times for commercials as stations in the larger market. For example, 55% of listening in Worcester, MA in Fall 2001 was to stations which were home to the Boston market which is approximately 40 miles away.[31] Third, the largest markets are disproportionately represented by circles. For example, 45% of the largest 20 markets in the data based on their Arbitron market size rank (New York City to Tampa, FL) are represented by circles compared with 24% for the remaining markets. One explanation is that clustering is greater in smaller markets because coordination is, in some sense, easier when there

[31] It is much rarer for listeners in a large market to listen to stations in a smaller market. For example only 1.8% of listening in the Boston market in Fall 2001 was to Worcester stations. I estimated a model with the λ_ms restricted to be the same in markets where more than 30% of rated listening in the smaller market was to stations in the larger market. These restrictions were rejected at the 1% significance level.

Figure 9: Market-Specific λ_m for 4-5 pm for Timing Defined by the Median Minute of a Commercial Break



are fewer stations. I investigate this possibility by allowing the incentive to coordinate to depend on covariates such as market size or the number of stations.

5.5.2 Market Characteristics

I now enrich the market-specific λ model to allow market characteristics to affect the attractiveness of each interval and the incentive to coordinate. For example, the relative attractiveness of having early commercials may depend on drivetime commuting patterns. The incentive to coordinate may also vary with observable and unobservable market characteristics.

Specifically I assume that player i in market m's payoff from choosing action t on day d is given by

$$\pi_{imdt} = X_m^{\beta}\beta_t + (X_m^{\alpha}\alpha + \xi_m^{\alpha})P_{-imdt} + \varepsilon_{idt} \tag{16}$$

where P_{-imdt} is the proportion of other stations in the market choosing action t on day d. X_m^{α} and X_m^{β} are observed market-specific covariates and ξ_m^{α} reflects unobserved market-specific factors which affect the incentive to coordinate. I normalize β_0 to be equal to zero. ξ_m^{α} is assumed to be independent of X_m^{α} and X_m^{β}, common across stations within a market, fixed within a market over time and to be independently and identically distributed across markets with $\xi_m^{\alpha} \sim N(0, \sigma^2)$. σ^2 is identified because the variance of the εs is fixed. There are multiple equilibria in market m if $X_m^{\alpha}\alpha + \xi_m^{\alpha}$ is large enough relative to $X_m^{\beta}\beta$. The log-likelihood, ignoring the binomial coefficients, is

$$\ln L = \sum_{m=1}^{M} \ln \int_{-\infty}^{\infty} \prod_{d=1}^{D_m} \left(\begin{array}{c} \lambda_m p_A^*(\beta_1, \alpha, X, \xi_m^{\alpha})^{n_{1md}}(1 - p_A^*(\beta_1, \alpha, X, \xi_m^{\alpha}))^{N_{md}-n_{1md}} \\ +(1-\lambda_m) p_B^*(\beta_1, \alpha, X, \xi_m^{\alpha})^{n_{1md}}(1 - p_B^*(\beta_1, \alpha, X, \xi_m^{\alpha}))^{N_{md}-n_{1md}} \end{array} \right) f(\xi_m^{\alpha}|\sigma) d\xi_m^{\alpha} \tag{17}$$

The integral cannot be computed analytically so I estimate the model using Simulated Maximum Likelihood (SML). As σ is a parameter to be estimated I draw u_{ms} from a standard uniform ($U[0,1]$) distribution and calculate $\xi_{ms}^{\alpha} = \sigma\Phi^{-1}(u_{ms})$ where $\Phi^{-1}()$ is the inverse of the cumulative distribution function of a standard normal random variable. The simulated log-likelihood is

$$\ln L^{SIM} = \sum_{m=1}^{M} \ln \frac{1}{S} \sum_{s=1}^{S} \prod_{d=1}^{D_m} \left(\begin{array}{c} \lambda_m p_A^*(\beta_1, \alpha, X, \sigma\Phi^{-1}(u_{ms}))^{n_{md1}} (1 - p_A^*(\beta_1, \alpha, X, \sigma\Phi^{-1}(u_{ms})))^{N_{md} - n_{md1}} \\ +(1-\lambda_m) p_B^*(\beta_1, \alpha, X, \sigma\Phi^{-1}(u_{ms}))^{n_{md1}} (1 - p_B^*(\beta_1, \alpha, X, \sigma\Phi^{-1}(u_{ms})))^{N_{md} - n_{md1}} \end{array} \right) \tag{18}$$

where S is the number of draws used for each market. SML estimates are only consistent if, as M and $D_m \to \infty$, $S \to \infty$ as well (Gouriéroux and Monfort (1996), p. 43) so I use 50 independent simulations draws per market.[32]

The α parameters are identified by whether commercials are more clustered in markets with high values of the X^α characteristics. The β_1 parameters are identified by whether commercials are more clustered in markets where most stations choose :53-:57 and whether this varies with the X^β characteristics. σ is identified from how much variation in the degree of clustering in different markets cannot be explained by the X^α variables. The X_m^β and X_m^α variables and ξ_m^α are constant over time so the parameters are identified from "between-market" variation. It is assumed that the X variables are exogenous to station timing decisions.

The X_m^β variables are a constant and the $COMMUTE$ variable described in Section 5.3. I use various combinations of X_m^α variables in addition to a constant. $MARKETRANK$ was also described in Section 5.3. $NUMBER$ is the average number of rated (non-zero listening share listed by Arbitron) contemporary music stations which are home to the market during the 4 ratings quarters in 2001. I use the average value because the very limited amount of within-market variation in the number of rated stations between ratings quarters is almost entirely due to whether small stations are rated by Arbitron and whether BIAfn classifies them in a contemporary music category. This variation should not affect the incentive of larger stations to coordinate with each other and given that ξ_m^α is assumed to be constant over ratings quarters this sort of variation would give strong predictions for how the degree of clustering should vary within-markets over time which would almost certainly be rejected

[32] It is appropriate to mention some further details of the estimation procedure. I use a nested estimation procedure where for each value of the parameters β_1, α and σ I maximize (18) with respect to the market-specific λ_m. (18) may have multiple local maxima so I start the estimation procedure from multiple starting points using 10 simulation draws per market and then, having established reasonable starting values, restart the estimation using 50 simulation draws per market. For both stages of the estimation procedure I try both derivative-based and non-derivative based methods to maximize (18). Multiple local maxima appear only to present a significant problem when there are several covariates. If $X_m^\alpha \alpha + \xi_{ms}^\alpha < 0$ I assume that the symmetric equilibrium is played even though there could be asymmetric equilibria. It is very rare for $X_m^\alpha \alpha + \xi_{ms}^\alpha < 0$ at the estimated parameters.

by the data.[33] The mean of $NUMBER$ is 12.1 stations (standard deviation 5.0). The markets with the most stations are Salt Lake City (24), Chicago, Pittsburgh and Wilkes-Barre/Scranton (all 22). These markets have the most stations because of demographics (Spanish-language stations which are common in many large markets are not included in the contemporary music categories) and the absence of large markets nearby. $OUTLISTENING$ measures the proportion of contemporary music listening which is to stations which are outside of the market. This is also averaged across ratings quarters in 2001. The mean of $OUTLISTENING$ is 0.20 (0.27) and it varies from 0 (in 20 markets) to 0.91 in Morristown, NJ. HHI measures ownership concentration among the rated contemporary music radio stations which are home to the market. This is also averaged across ratings quarters because there were few within-market mergers in 2001 and within-market variation is therefore largely due to the number of stations counted as being in the market. Ownership is identified using the transactions history of each station listed in BIAfn's database and this is used to calculate each firm's share of rated contemporary music stations. HHI is the sum of the squares of these shares. The average of HHI is 0.28 (0.15). It has a maximum of 1 in Reading, PA where there is only one home to the market contemporary music station and minimum values in Nashville, TN (0.11) and Columbus, OH (0.12).

Table 7 presents the estimation results using data from 4-5 pm with action 1 defined as having a commercial in the :53-:57 time interval. The same simulation draws for each market are used in each specification to make the results more comparable across columns. The covariates, apart from the constants, are normalized to have means equal to zero and standard deviations equal to one. The coefficients therefore represent the effect of a one standard deviation increase in the variables. Standard errors are, as before, calculated using a resampling bootstrap procedure with 25 replications.

In column (1) X_m^β includes a constant and $COMMUTE$ while X_m^α only includes a constant. The incentive to coordinate can vary across markets because of unobserved heterogeneity. Both of the β_1 parameters are very small and insignificantly different from zero, indicating earlier commercials are not systematically more attractive in markets where commuting times are longer or shorter than

[33] Sweeting (2004a) examines the effect of within-market variation in characteristics on the overlap of commercials using linear regressions. He finds that within-market variation has less effect on overlap than between-market variation in the same characteristics although the coefficients typically have the same sign.

Table 7: Results with Market-Specific λ_m Equilibrium Selection Probabilities and Market Characteristics for 4-5 pm

	(1)	(2)	(3)	(4)	(5)
β_1					
Constant	4.5e-6(1.0e-4)	-3.6e-5(9.6e-5)	-8.2e-5(5.3e-5)	-7.9e-5(7.5e-5)	-7.7e-5(7.6e-5)
$COMMUTE$	1.3e-6(1.1e-4)	-1.1e-5(1.1e-4)	5.7e-5(6.1e-5)	6.6e-5(9.1e-5)	6.9e-5(9.4e-5)
α					
Constant	2.0773(0.0133)	2.0846(0.0080)	2.0800(0.0124)	2.0897(0.0114)	2.0898(0.0099)
$MARKETRANK$	-	0.0579(0.0123)	-	0.0458(0.0089)	0.0436(0.0078)
$NUMBER$	-	-	-0.0446(0.0107)	-	-
HHI	-	-	-	0.0390(0.0198)	0.0374(0.0178)
$OUTLISTENING$	-	-	-	-	4.1e-6(0.0075)
σ	0.0816(0.0157)	0.0671(0.0143)	0.0817(0.0151)	0.0668(0.0126)	0.0659(0.0143)
Simulated Log Likelihood	-20025.1	-20003.2	-20018.2	-19999.0	-19998.8
Market-days	7,656	7,656	7,656	7,656	7,656
Station-days	30,332	30,332	30,332	30,332	30,332

Notes: Standard errors in parentheses. Log-likelihoods do not include binomial coefficients which are constants and independent of the parameters.

average. This is true for all of the specifications in the table. The average incentive to coordinate across markets is 2.0773, which implies equilibrium values of (p_A^*, p_B^*) of (0.6646,0.3354). Multiple equilibria cannot be supported if $X_m^\alpha \alpha + \xi_m^\alpha$ is less than 2, and with an estimated standard deviation of ξ_m^α of 0.0816, the probability that a market has $X_m^\alpha \alpha + \xi_m^\alpha$ less than 2 is approximately 0.17. The remaining columns examine whether market characteristics explain the heterogeneity in the incentive to coordinate.

Column (2) includes $MARKETRANK$ in X_m^α and the positive coefficient shows that there is a greater incentive to coordinate (more clustering) in markets with lower market population. The $MARKETRANK$ coefficient is highly significant. A one standard deviation decrease in $MARKETRANK$ from its mean, which is equivalent to moving from El Paso, TX to San Jose, CA, changes the equilibrium values of (p_A^*, p_B^*) from (0.6716,0.3284) to (0.5989,0.4011) if $\xi_m^\alpha = 0$. The reduction in the estimate of σ is consistent with $MARKETRANK$ explaining some of the heterogeneity in the degree of clustering across markets. Column (3) replaces $MARKETRANK$ with $NUMBER$. The negative and significant coefficient indicates that the incentive to coordinate is lower in markets with more stations, although the estimate of σ is the same as in column (1). This suggests that variation in the incentive to coordinate is determined more by factors which vary with market size than the number of rated contemporary music stations in the market which, as noted above, is larger in Salt Lake City (ranked 34) and Wilkes-Barre/Scranton (67) than New York City (1) and Los Angeles.(2).

Column (4) includes $MARKETRANK$ and HHI in X_m^{α}. The positive coefficient on HHI, which is significant at the 5% level, indicates that there is greater incentive to coordinate in markets with more concentrated ownership. This is consistent with the analysis in Section 5.4 where I showed that stations would coordinate more if they internalized how their individual timing decisions affected the audience of other stations because common owners should internalize these externalities. With $MARKETRANK$ equal to its mean, a one standard deviation decrease in HHI (from 0.28 to 0.13) changes the equilibrium values of (p_A^*, p_B^*) from (0.6730,0.3237) to (0.6348,0.3652) if $\xi_m^{\alpha} = 0$. Column (5) includes $MARKETRANK$, HHI and $OUTLISTENING$ in X_m^{α}. If markets with more stations have weaker incentives to coordinate then we might also expect that the ability of listeners to receive stations in other markets would also weaken the incentive. However, the $OUTLISTENING$ coefficient is almost identical to zero.

6 Conclusion

Many models with interesting interactions between agents have multiple equilibria. In much of the applied literature multiplicity has been seen as only creating estimation problems. Common responses to this perceived problem such as changing the model to guarantee uniqueness or assuming that only one equilibrium is played in the data are unsatisfactory if it is plausible, as it surely often is, that the data contains observations from different equilibria. The central idea in this paper is that not only is it possible to estimate models where there are multiple equilibria both in the model and in the data but that the existence of multiple equilibria can actually help to identify the parameters. The intuition is that multiple equilibria only arise when interactions are important and the existence of multiple equilibria in the data allows us to rule out parameters which cannot support all of the equilibria. I illustrate this idea using a simple game in which the parameters are only separately identified if there are multiple equilibria both in the game and in the data.

The game is used to study the timing of commercial breaks by music radio stations. A station, which sells the audience of its commercial breaks to advertisers, may have an incentive to play its commercials at the same time as other stations in its market to reduce the number of listeners who avoid its commercials by switching to music on other stations. However, the fact that stations do tend to play commercials at the same time (Figure 1) could also be explained by common factors, such as Arbitron's methodology for estimating station ratings, which make certain times attractive for all stations to play commercials. I use the possibility that stations may coordinate on playing commercials at different times in different markets, together with the assumption that common factors are the same across markets, to identify the incentive to coordinate. I find evidence of multiple equilibria, allowing the incentive to coordinate to be identified, during drivetime. The estimated incentive to coordinate has only modest effects on Nash equilibrium timing strategies but implies that commercials would overlap almost perfectly if each station internalized how the timing of its commercials affects the audience of other stations. I find that markets tend to stay in the same equilibrium over time and that commercials overlap more in smaller markets and in markets with more concentrated station ownership.

Two issues merit further discussion. The first issue is how the overlap of commercials affects welfare. The externalities in the coordination game suggest that advertising time would become more valuable if commercials overlapped more than they do at present. Stations would extract some of this value through higher prices to advertisers. Increased listenership to commercials is one possible reason why increases in ownership concentration in local radio markets have been associated with small increases in advertising prices (Brown and Williams (2002)). A large increase in revenues might indirectly benefit listeners by encouraging station entry, which would increase programming variety, and by encouraging investments in station quality. The free-rider problem means that an individual listener ignores these effects when switching stations. However, a welfare calculation would also take into account listeners' disutility from hearing commercials they do not value and which they are unable to pay to avoid.

The second issue is whether the idea that multiple equilibria can help to identify parameters applies in models other than the type of game considered here where players have an incentive to choose the same action. In many settings, especially in Industrial Organization, a more natural assumption may be that agents want to choose different actions to avoid competition. For example, if firm A provides a high quality product then firm B may want to have a low quality product to soften price competition. Can multiple equilibria in the data help to provide identification in this setting? This question deserves detailed consideration in its own right but a simple example suggests that they can. Suppose that A and B make simultaneous product quality choices in a set of independent markets (an example might be two gasoline retailers deciding whether each outlet should be full-serve or self-serve) and that, in the data, we always observe one high quality firm and one low quality firm. If firm A is the high quality firm in every market then this pattern could be explained either by strategic differentiation or by A having a relative cost advantage in producing a high quality product. On the other hand, if we observe A as the high quality firm in some markets and B as the high quality firm in other markets then if we treat these different outcomes as representing multiple equilibria in the quality choice game and are willing to assume away perfect negative correlation in firm costs then this may provide convincing evidence that strategic differentiation is important.

References

[1] Abernethy, Avery M. (1991), "Differences Between Advertising and Program Exposure for Car Radio Listening", *Journal of Advertising Research*, 31(2), 33-42

[2] Ackerberg, Daniel A. and Gautam Gowrisankaran (2002), "Quantifying Equilibrium Network Externalities in the ACH Banking Industry", mimeo, University of California at Los Angeles

[3] Aguirregabiria, Victor and Pedro Mira (2004), "Sequential Estimation of Dynamic Discrete Games", mimeo, Boston University

[4] Andrews, Donald W.K. and Steven T. Berry (2003), "On Placing Bounds on Parameters of Entry Games in the Presence of Multiple Equilibria", mimeo, Yale University

[5] Arbitron Company and Edison Media Research (1999), "Will Your Audience Be Right Back After These Messages?", available on-line at www.arbitron.com

[6] Arbitron Company and Edison Media Research (2003), "The National In-Car Study: Fighting for the Front Seat", available on-line at www.arbitron.com

[7] Arbitron Company (2003), "Radio Market Report. Audience Estimates in the New England County Metropolitan Area, ADI and TSA for Boston: Fall 2002 Issue", New York, NY: Arbitron Company

[8] Bajari, Patrick, C. Lanier Benkard and Jonathan Levin (2003), "Estimating Dynamic Models of Imperfect Competition", mimeo, Stanford University

[9] Bajari, Patrick, Han Hong and Stephen Ryan (2004), "Identification and Estimation of Discrete Games of Complete Information", mimeo, Duke University

[10] Beran, Rudolf (1988), "Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements", *Journal of the American Statistical Association*, 83(403), 687-697

[11] Berry, Steven T. (1992), "Estimation of a Model of Entry in the Airline Industry", *Econometrica*, 60(4), 889-917

[12] BIA Financial Network, Inc. (2002), *MediaAccess Pro version 3.0*, Chantilly, VA: BIA Financial Network, Inc.

[13] Bjorn, Paul A. and Quang H. Vuong (1985), "Simultaneous Equation Models for Dummy Endogenous Variables: a Game Theoretic Formulation with an Application to Labor Force Participation", Social Science Working Paper No. 537, California Institute of Technology

[14] Bresnahan, Timothy F. and Peter C. Reiss (1991), "Entry and Competition in Concentrated Markets", *Journal of Political Economy*, 99(5), 977-1009

[15] Brock, William A. and Steven N. Durlauf (2001), "Interactions-Based Models", in *Handbook of Econometrics* Volume 5, ed. by J.J. Heckman and E. Leamer, New York: New Holland

[16] Brown, Keith and George Williams (2002), "Consolidation and Advertising Prices in Local Radio Markets", mimeo, Federal Communications Commission

[17] Brydon, Alan (1994), "Radio Must Prove Its Merit as an Advertising Medium", *Campaign*, 3(21), 25-6

[18] Chen, Hanfeng and Jiahua Chen (2001), "The Likelihood Ratio Test for Homogeneity in Finite Mixture Models", *Canadian Journal of Statistics*, 29(2), 201-215

[19] Chernoff, Herman and Eric Lander (1995), "Asymptotic Distribution of the Likelihood Ratio Test that a Mixture of Two Binomials is a Single Binomial", *Journal of Statistical Planning and Inference*, 43, 19-40

[20] Ciliberto, Federico and Elie Tamer (2004), "Market Structure and Multiple Equilibria in Airline Markets", mimeo, Northwestern University

[21] Cooper, Russell (2004), "Estimation and Identification of Structural Parameters in the Presence of Multiple Equilibria", *Les Annales d'Economie et de Statistique*, 66, 1-26

[22] Davis, Donald R. and David E. Weinstein (2004), "A Search for Multiple Equilibria in Urban Industrial Structure", National Bureau of Economic Research Working Paper No. 10252

[23] Dempster, Arthur P., Nan M. Laird and Donald B. Rubin (1977), "Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society Series B (Methodological)*, 39(1), 1-38

[24] Dick, Steven J. and Walter McDowell (2003), "Estimating Relative Commercial Zapping Among Radio Stations Using Standard Arbitron Ratings", mimeo, University of Miami at Coral Gables

[25] Duflo, Esther and Emmanuel Saez (2002), "Participation and Investment Decisions in a Retirement Plan: The Influence of Colleagues' Choices", *Journal of Public Economics*, 85(1), 121-148

[26] Ellison, Glenn D. and Edward L. Glaeser (1997), "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach", *Journal of Political Economy,* 105(5), 889-927

[27] Elmore, Ryan and Shaoli Wang (2003), "Identifiability and Estimation in Multinomial Mixture Models", Pennsylvania State University, Department of Statistics Technical Report No. 03-04

[28] Epstein, Gil S. (1998), "Network Competition and the Timing of Commercials", *Management Science*, 44(3), 370-387

[29] Feng, Ziding D. and Charles E. McCulloch (1996), "Using Bootstrapped Likelihood Ratios in Finite Mixture Models", *Journal of the Royal Statistical Society Series B*, 58(3), 609-617

[30] Glaeser, Edward L., Bruce Sacerdote and Jose A. Scheinkman (1996), "Crime and Social Interactions", *Quarterly Journal of Economics*, 111(2), 507-548

[31] Goolsbee, Austan and Peter J. Klenow (2002), "Evidence on Learning and Network Externalities in the Diffusion of Home Computers", *Journal of Law and Economics*, 45(2), 317-43

[32] Gouriéroux, Christian and Alain Monfort (1996), *Simulation-Based Econometric Methods*, Oxford, UK: Oxford University Press

[33] Greene, William H. (1997), *Econometric Analysis*, 3^{rd} edition, Upper Saddle River, N.J.: Prentice Hall

[34] Keith, Michael C. (1987), *Radio Programming*, Boston, MA: Focal Press

[35] Keith, Michael C. (2000), *The Radio Station*, 5^{th} edition, Boston, MA: Focal Press

[36] Kim, Byung Soo (1984), "Studies of Multinomial Mixture Models", unpublished PhD thesis, University of North Carolina at Chapel Hill

[37] Kooreman, Peter (1994), "Estimation of Econometric Models of Some Discrete Games", *Journal of Applied Econometrics,* 9(3), 255-68

[38] Lemdani, Mohamed and Odile Pons (1996), "Likelihood Ratio Tests in Mixture Models", *C.R. Acad. Sci. Paris Série I*, 322, 399-404

[39] Lemdani, Mohamed and Odile Pons (1995), "Tests for Genetic Linkage and Homogeneity", *Biometrics*, 51(3), 1033-1041

[40] Lemdani, Mohamed and Odile Pons (1997), "Likelihood Ratio Tests for Genetic Linkage", *Statistics and Probability Letters*, 33, 15-22

[41] Liang, Kung-Yee and Paul J. Rathouz (1999), "Hypothesis Testing Under Mixture Models: Application to Genetic Linkage Analysis", *Biometrics*, 55(1), 65-74

[42] Lynch, Joanna R. and Greg Gillespie (1998), *Process and Practice of Radio Programming*, Lanham, MD: University Press of America

[43] MacFarland, David T. (1997), *Future Radio Programming Strategies*, 2^{nd} edition, Mahwah, N.J.: Erlbaum

[44] Manski, Charles F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies*, 60(3), 531-42

[45] Margolin, Barry H., Byung Soo Kim and Kenneth J. Risko (1989), "The Ames Salmonella/ Microsome Mutagenicity Assay: Issues of Inference and Validation", *Journal of the American Statistical Association*, 84(407), 651-661

[46] Mazzeo, Michael J. (2002), "Product Choice and Oligopoly Market Structure", *RAND Journal of Economics*, 33(2), 221-242

[47] McDowell, Walter and Steven J. Dick (2003), "Switching Radio Stations While Driving: Magnitude, Motivation and Measurement Issues", *Journal of Radio Studies*, 10(1), 46-62

[48] McLachlan, Geoffrey J. and Thriyambakam Krishnan (1997), *The EM Algorithm and Extensions*, New York, NY: John Wiley

[49] McLachlan, Geoffrey J. and David Peel (2000), *Finite Mixture Models*, New York, NY: John Wiley

[50] Media Daily News (2004), "SMG Strikes TV Commercial Ratings Deal, Move Could Alter TV Advertising Market", June 14 2004, New York, NY: Media Post Communications

[51] Mogelonsky, Marcia (1995), "Coping with Channel Surfers", *American Demographics*, 17(12), 13-15

[52] Moro, Andrea (2003), "The Effect of Statistical Discrimination on Black-White Wage Inequality: Estimating a Model with Multiple Equilibria", *International Economic Review*, 44(2), 467-500

[53] Pakes, Ariel, Michael Ostrovsky and Steven T. Berry (2003), "Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples)", mimeo, Harvard University

[54] Pesendorfer, Martin and Philip Schmidt-Dengler (2003), "Identification and Estimation of Dynamic Games", National Bureau of Economic Research Working Paper No. 9726

[55] Radio Advertising Bureau (2002), *Media Facts: A Guide to Competitive Media*, Irving, TX

[56] Redner, Richard (1981), "Note on the Consistency of the Maximum Likelihood Estimate for Nonidentifiable Distributions", *Annals of Statistics*, 9(1), 225-228

[57] Salami, Michael K. (1986), "Solution and Estimation of Linear Rational Expectations Models", *Journal of Econometrics*, 31(1), 41-66

[58] Sargent, Thomas J. and Neil Wallace (1987), "Inflation and the Government Budget Constraint", in *Economic Policy in Theory and Practice*, ed. by A. Razin and E. Sadka, New York, NY: St. Martin's Press

[59] Seim, Katja (2002), "An Empirical Model of Firm Entry with Endogenous Product-Type Choices", mimeo, Stanford University

[60] Sweeting, Andrew T. (2004a), "Coordination, Differentiation and the Timing of Radio Commercials", Northwestern University Center for the Study of Industrial Organization Working Paper No. 50

[61] Sweeting, Andrew T. (2004b), "Music Variety, Station Listenership and Station Ownership in the Music Radio Industry", Northwestern University Center for the Study of Industrial Organization Working Paper No. 49

[62] Tamer, Elie T. (2003), "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria", *Review of Economic Studies*, 70(1), 147-165

[63] Teicher, Henry (1963), "Identifiability of Finite Mixtures", *Annals of Mathematical Statistics*, 34(4), 1265-1269

[64] Wang, Peiming (1994), "Mixed Regression Models for Discrete Data", unpublished PhD thesis, University of British Columbia

[65] Warren, Steve (2001), *The Programming Operations Manual*, San Marcos, TX: Warren Consulting

[66] Zhou, Wen (2000), "The Magnitude, Timing, and Frequency of Firm Choice: Essays on Commercial Breaks and Price Discounts", unpublished Ph.D. thesis, Duke University

A Proofs of Propositions 1 and 2

Proposition 1. *All parameter vectors (β_1, α) where (β_1, α) support only one equilibrium are not separately identified.*

Proof. If (β_1, α) support only one equilibrium then the single equilibrium choice probability p^* satisfies

$$p^*(\beta_1, \alpha) = \frac{e^{\beta_1 + \alpha p^*}}{e^{\beta_1 + \alpha p^*} + e^{\alpha(1-p^*)}} \tag{19}$$

and the probability of an (N_m,n_{1m}) observation is

$$\Pr(n_{1m}|\beta_1, \alpha, N_m) = \frac{N_m!}{n_{1m}!N_m - n_{1m}!} p^*(\beta_1, \alpha)^{n_{1m}}(1 - p^*(\beta_1, \alpha))^{N_m - n_{1m}} \tag{20}$$

I show that (β_1, α) are not separately identified by naming $(\beta_1', \alpha') \neq (\beta_1, \alpha)$ pairs which also support one equilibrium choice probability $p^{*\prime}(\beta_1', \alpha')$ with $p^{*\prime}(\beta_1', \alpha') = p^*(\beta_1, \alpha)$, implying that $\Pr(n_{1m}|\beta_1, \alpha, N_m) = \Pr(n_{1m}|\beta_1', \alpha', N_m)$. I name particular (β_1', α') pairs but there is a continuum of pairs which would work.

If $\alpha > 0$, consider $\alpha' = 0$ and $\beta_1' = \beta_1 - \alpha + 2\alpha p^*$. As α'=0 there is only a single equilibrium $p^{*\prime}(\beta_1', \alpha')$ which satisfies

$$p^{*\prime}(\beta_1', \alpha') = \frac{e^{\beta_1'}}{1 + e^{\beta_1'}} = \frac{e^{\beta_1 - \alpha + 2\alpha p^*}}{1 + e^{\beta_1 - \alpha + 2\alpha p^*}} = \frac{e^{\beta_1 + \alpha p^*}}{e^{\beta_1 + \alpha p^*} + e^{\alpha(1-p^*)}} \tag{21}$$

so $p^{*\prime}(\beta_1', \alpha') = p^*(\beta_1, \alpha)$ and (β_1, α) are not separately identified.

If $\alpha = 0$, $p^*(\beta_1, \alpha)$ is given by

$$p^*(\beta_1, \alpha) = \frac{e^{\beta_1}}{1 + e^{\beta_1}} \tag{22}$$

Consider $\alpha' = 1$ and $\beta_1' = \beta_1 + 1 - 2p^*$. As a necessary condition for there to be multiple equilibria is that the maximum slope of a station's reaction function ($\frac{1}{2}\alpha$) is at least 1, there is a single equilibrium.[34] (β_1', α') support a single equilibrium choice probability $p^{*\prime}$ with

$$p^{*\prime}(\beta_1', \alpha') = \frac{e^{\beta_1' + \alpha' p^{*\prime}}}{e^{\beta_1' + \alpha' p^{*\prime}} + e^{\alpha'(1-p^{*\prime})}} = \frac{e^{\beta_1 + 1 - 2p^* + p^{*\prime}}}{e^{\beta_1 + 1 - 2p^* + p^{*\prime}} + e^{(1-p^{*\prime})}} \tag{23}$$

If $p^{*\prime}(\beta_1', \alpha') = p^*(\beta_1, \alpha)$ then the RHS of (23) simplifies to

$$p^{*\prime}(\beta_1', \alpha') = \frac{e^{\beta_1}}{1 + e^{\beta_1}} \tag{24}$$

which verifies that $p^{*\prime}(\beta_1', \alpha') = p^*(\beta_1, \alpha)$ and (β_1, α) are not separately identified. ■

Proposition 2. *Parameter vectors (β_1, α) which support two distinct stable equilibria are separately identified if Conditions 1 and 2 hold.*

Proof. If (β_1, α) support two distinct equilibria the probability of an (N_m, n_{1m}) observation is

$$\Pr(n_{1m}|\beta_1, \alpha, \lambda, N_m) = \frac{N_m!}{n_{1m}!(N_m - n_{1m})!} \begin{pmatrix} \lambda p_A^*(\beta_1, \alpha)^{n_{1m}}(1 - p_A^*(\beta_1, \alpha))^{N_m - n_{1m}} + \\ (1 - \lambda) p_B^*(\beta_1, \alpha)^{n_{1m}}(1 - p_B^*(\beta_1, \alpha))^{N_m - n_{1m}} \end{pmatrix} \tag{25}$$

[34] Recall from Section 2.1 that all equilibria are symmetric so that there is an equilibrium where $R(p) = p$ (the reaction function crosses the 45^0 line). The reaction function is continuous on $[0, 1]$. Taken together these facts imply that for there to be two distinct equilibria the slope of the reaction function must be greater than 1 at some point.

with $p_A^*(\beta_1,\alpha) \neq p_B^*(\beta_1,\alpha)$. I proceed in two stages: first, I apply a well-known result to show that (p_A^*, p_B^*, λ) are separately identified under the stated conditions and second, I show that (β_1, α) are identified if p_A^* and p_B^* are identified.

First stage: (25) is the pmf of a binomial mixture distribution with two components. The binomial probabilities for the two components are p_A^* and p_B^*, and the mixing proportion parameter is λ. Proposition 4 of Teicher (1963) and the lemma of Margolin et al. (1989) give sufficient conditions for p_A^*,p_B^* and λ to be identified. Applying these results, (p_A^*, p_B^*, λ), with $p_A^* \geq p_B^*$, are separately identified if $p_A^* \neq p_B^*$ ((β_1,α) support more than one equilibrium), $0 < \lambda < 1$ (Condition 1) and some of the observed groups contain at least 3 members (Condition 2). Therefore (p_A^*, p_B^*, λ) are separately identified.[35]

Second stage: it is sufficient to show that there is a unique (β_1, α) which can support a pair (p_A^*, p_B^*) with $p_A^* \neq p_B^*$ as equilibrium choice probabilities. Suppose not and that $(\beta_1', \alpha') \neq (\beta_1, \alpha)$ also supports (p_A^*, p_B^*) with $p_A^* \neq p_B^*$ as equilibrium choice probabilities. From (4), four equations must hold

$$p_A^* = \frac{e^{\beta_1 + \alpha p_A^*}}{e^{\beta_1 + \alpha p_A^*} + e^{\alpha(1-p_A^*)}} \qquad p_A^* = \frac{e^{\beta_1' + \alpha' p_A^*}}{e^{\beta_1' + \alpha' p_A^*} + e^{\alpha'(1-p_A^*)}} \tag{26}$$

$$p_B^* = \frac{e^{\beta_1 + \alpha p_B^*}}{e^{\beta_1 + \alpha p_B^*} + e^{\alpha(1-p_B^*)}} \qquad p_B^* = \frac{e^{\beta_1' + \alpha' p_B^*}}{e^{\beta_1' + \alpha' p_B^*} + e^{\alpha'(1-p_B^*)}} \tag{27}$$

and manipulating each of these equations gives

$$\ln\left(\frac{p_A^*}{1-p_A^*}\right) = \beta_1 + \alpha(2p_A^* - 1) \qquad \ln\left(\frac{p_A^*}{1-p_A^*}\right) = \beta_1' + \alpha'(2p_A^* - 1) \tag{28}$$

$$\ln\left(\frac{p_B^*}{1-p_B^*}\right) = \beta_1 + \alpha(2p_B^* - 1) \qquad \ln\left(\frac{p_B^*}{1-p_B^*}\right) = \beta_1' + \alpha'(2p_B^* - 1) \tag{29}$$

Combining the equations in (28) and (29) gives

$$\alpha' - \alpha = \frac{\beta_1 - \beta_1'}{2p_A^* - 1} \quad \text{and} \quad \alpha' - \alpha = \frac{\beta_1 - \beta_1'}{2p_B^* - 1} \tag{30}$$

implying $p_A^* = p_B^*$. This contradicts $p_A^* \neq p_B^*$. ■

B Estimation and Testing

B.1 EM Algorithm Estimation of the Basic Model with Multiple Equilibria

This Appendix describes how the EM Algorithm can be used to estimate the parameters of the basic model. The equilibrium probabilities of choosing action 1 are $p_A^*(\beta_1,\alpha)$ and $p_B^*(\beta_1,\alpha)$ with $p_A^*(\beta_1,\alpha) \geq p_B^*(\beta_1,\alpha)$. I use the algorithm to estimate p_A^*, p_B^* and λ. The estimates of p_A^* and p_B^* are then used to provide estimates of β_1 and α. The log-likelihood in terms of p_A^*, p_B^* and λ, ignoring the binomial coefficient, is

$$\ln L = \sum_{m=1}^{M} \ln \left\{ \begin{array}{c} \lambda {p_A^*}^{n_{1m}} (1 - p_A^*)^{N_m - n_{1m}} \\ +(1-\lambda) {p_B^*}^{n_{1m}} (1 - p_B^*)^{N_m - n_{1m}} \end{array} \right\} \tag{31}$$

[35] These conditions are also necessary. If $p_A^* = p_B^*$ then one cannot identify λ because the two components of the mixture are identical (indeed λ is not really defined). If $\lambda = 0$ then one cannot identify p_A^* (no observations come from this component so there is no information on its binomial probability) and if $\lambda = 1$ then one cannot identify p_B^*. If all groups have only 1 or 2 members then the only possible $(n_{1m}, N_m - n_{1m})$ outcomes are $(2,0)$, $(1,1)$ and $(0,2)$. It is easy to show that there is more than one combination of (p_A^*, p_B^*, λ) which give the same probabilities for these outcomes. The intuition is that as the probabilities sum to 1 there are only two equations and three unknowns.

where M is the number of independent repetitions of the game.

The first-order conditions from maximizing (31) with respect to λ, p_A^* and p_B^* are

$$\sum_{m=1}^{M} \frac{\tau_m}{\lambda} - \frac{(1-\tau_m)}{(1-\lambda)} = 0 \tag{32}$$

$$\sum_{m=1}^{M} \left(\frac{n_{1m}\tau_m}{p_A^*} - \frac{(N_m - n_{1m})\tau_m}{(1-p_A^*)} \right) = 0 \tag{33}$$

$$\sum_{m=1}^{M} \left(\frac{n_{1m}(1-\tau_m)}{p_B^*} - \frac{(N_m - n_{1m})(1-\tau_m)}{(1-p_B^*)} \right) = 0 \tag{34}$$

where τ_m is the conditional probability (given that data and parameters) that equilibrium A is played in repetition m,

$$\tau_m = \frac{\lambda {p_A^*}^{n_{1m}}(1-p_A^*)^{N_m-n_{1m}}}{\lambda {p_A^*}^{n_{1m}}(1-p_A^*)^{N_m-n_{1m}} + (1-\lambda){p_B^*}^{n_{1m}}(1-p_B^*)^{N_m-n_{1m}}} \tag{35}$$

The EM Algorithm exploits the fact that the solution to (32)-(34) is also a solution to iterating a two-step "Expectation(E)-Maximization(M)" procedure.[36] The E-step takes the conditional expectation of the complete data log-likelihood by replacing indicator variables for the equilibrium being played (the Z_m^As in (6)) with $\widehat{\tau}_m$ which is τ_m evaluated at the current values of the parameters $(\widehat{p_A^*}, \widehat{p_B^*}, \widehat{\lambda})$. Ignoring the binomial coefficient, this expectation is

$$E(\ln L^C) = \sum_{m=1}^{M} \left(\begin{array}{c} \widehat{\tau}_m \left[\ln \widehat{\lambda} + n_{1m} \ln \widehat{p_A^*} + (N_m - n_{1m}) \ln(1-\widehat{p_A^*}) \right] + \\ (1-\widehat{\tau}_m) \left[\ln(1-\widehat{\lambda}) + n_{1m} \ln \widehat{p_B^*} + (N_m - n_{1m}) \ln(1-\widehat{p_B^*}) \right] \end{array} \right) \tag{36}$$

The M-step involves maximizing (36) with respect to the parameters $\widehat{\lambda}, \widehat{p_A^*}$ and $\widehat{p_B^*}$. The new parameter estimates are

$$\widehat{\widehat{\lambda}} = \frac{\sum_{m=1}^{M} \widehat{\tau}_m}{M} \tag{37}$$

$$\widehat{\widehat{p_A^*}} = \frac{\sum_{m=1}^{M} \widehat{\tau}_m n_{1m}}{\sum_{m=1}^{M} \widehat{\tau}_m N_m} \tag{38}$$

$$\widehat{\widehat{p_B^*}} = \frac{\sum_{m=1}^{M} (1-\widehat{\tau}_m) n_{1m}}{\sum_{m=1}^{M} (1-\widehat{\tau}_m) N_m} \tag{39}$$

The E- and M-steps are iterated until the likelihood and the parameter estimates converge. The final estimates $\widehat{p_A^*}$ and $\widehat{p_B^*}$ are used to calculate $\widehat{\beta_1}$ and $\widehat{\alpha}$ by solving two equations

$$\ln\left(\frac{\widehat{p_A^*}}{1-\widehat{p_A^*}}\right) = \widehat{\beta_1} + \widehat{\alpha}(2\widehat{p_A^*} - 1) \tag{40}$$

$$\ln\left(\frac{\widehat{p_B^*}}{1-\widehat{p_B^*}}\right) = \widehat{\beta_1} + \widehat{\alpha}(2\widehat{p_B^*} - 1) \tag{41}$$

[36] Dempster et al. (1977) show that an EM iteration always increases the value of the likelihood so convergence is guaranteed when the likelihood is bounded above as it is here.

I then check whether $\widehat{\beta_1}$ and $\widehat{\alpha}$ support $\widehat{p_A^*}$ and $\widehat{p_B^*}$ as stable equilibria. If they do then $\widehat{\beta_1}$ and $\widehat{\alpha}$ are the ML estimates. If they do not then it is necessary to maximize (12) with respect to β_1 and α directly. This requires the calculation of the $p_A^*(\beta_1, \alpha)$ and $p_B^*(\beta_1, \alpha)$ which β_1 and α support as stable equilibria at each step of the maximization.

B.2 Testing for Multiple Equilibria Using a Bootstrap

B.2.1 Bootstrap Test for the Homogeneity of Mixtures

Chen and Chen (2001) propose using a parametric bootstrap to estimate the distribution of the LRTS in testing for the homogeneity of a binomial mixture model against the alternative of a two component mixture. This simulation exercise provides evidence that this approach gives a test of the correct size.

In each simulation data is generated from a single component binomial distribution $Binomial(n, p)$ where p is chosen as 0.52 (the proportion of stations choosing action 1 is close to 0.5 in the actual data) and n is 1 in 50 repetitions, 2 in 50 repetitions and so on up to a value of 10 in 50 repetitions. This data is used to estimate a single component binomial model (estimated parameter $\widehat{p}$) and a two component binomial model (estimated parameters $\widehat{\lambda}, \widehat{p_A}, \widehat{p_B}$) using the EM Algorithm described above. The LRTS is calculated to compare the two models. The estimate $\widehat{p}$ is used to generate 249 new sets of data from a single component model $Binomial(n, \widehat{p})$ with n varying as before. One and two component models are estimated for each set of data in order to estimate the distribution of the LRTS under the null of a single component. The jth-order statistic estimates the $\frac{j}{250}$th quantile of the distribution of the LRTS. This estimated distribution is used to assess whether the LRTS for the initial set of data is statistically significant at the 10%, 5% and 1% levels.

500 simulations are used. The LRTS is statistically significant at the 1% level in 6 simulations (1.2%), at the 5% level in 20 simulations (4.0%) and at the 10% level in 47 simulations (9.4%).

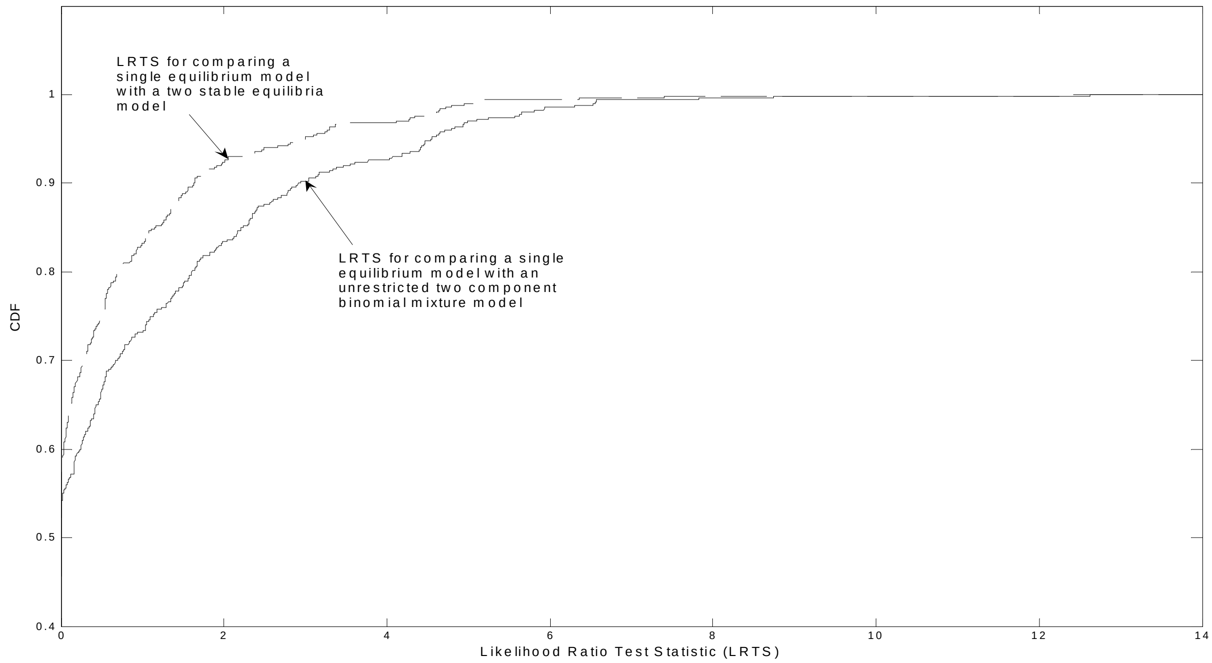
B.2.2 Conservative Assessment of the Significance of Multiple Equilibria

The previous simulation exercise examined the use of a parametric bootstrap of the LRTS's distribution in comparing a single component binomial model against a two component binomial mixture model with no constraints on the components of the mixture p_A and p_B. The economic model in Section 2 does impose the constraint that the components must be supported as stable equilibria in the model. This simulation exercise confirms that imposing this constraint on the two component model (weakly) lowers the value of the LRTS (the one component model remains unconstrained). This implies that the testing procedure described in Section 4.3 will give conservative results, i.e., I am less likely to reject the null hypothesis of a single equilibrium.

In each simulation data is generated from the model with a single equilibrium where action 1 is chosen with probability $p^* = 0.52$ and there is 1 observed player in 50 repetitions, 2 players in 50 repetitions and so on up to 10 players in 50 repetitions. This data is used to estimate a single equilibrium model, a two equilibria model and an unconstrained two component binomial model where $\widehat{p_A}$ and $\widehat{p_B}$ do not have to be supported as stable equilibria. I calculate (a) an LRTS comparing the one equilibrium model and the two equilibria model and (b) an LRTS comparing the one equilibrium model and the unconstrained two component binomial model.

500 simulations are used. In every simulation the (a) comparison LRTS is less than or equal to the (b) comparison LRTS as expected. It is strictly less in 243 simulations. Figure 10 shows the cumulative distribution function of the LRTS for each comparison. The difference between the distribution functions suggests that my assessments of whether there is significant evidence of multiple equilibria may be quite conservative.

Figure 10: CDF of the LRTS Comparing the One Equilibrium Model and the Two Equilibria Model and the CDF of the LRTS Comparing the One Equilibrium Model and the Two Component Binomial Mixture Model



C Results from the Basic Model With Actions Defined by the Order of Songs and Commercial Breaks

Table 8 presents the results from estimating the basic model with action 1 defined as having a commercial break with one song remaining in the hour (action 0 two songs). The results are generally similar to those in Table 5. The main differences are that the LRTS for 4-5 pm is marginally insignificant at the 10% level (recall that my testing procedure gives conservative assessments of the significance of multiple equilibria) and the LRTS for 5-6 pm is much more significant than before. The equilibrium strategies for 5-6 pm and 6-7 pm also imply greater overlap of commercials than in Table 5.

Table 8: Basic Model Results with Timing Defined by the Order of Songs

(a) Afternoon Drivetime Hours

	3-4 pm	4-5 pm	5-6 pm	6-7 pm
One Equilibrium Model				
β_1 (assuming $\alpha = 0$)	-0.1390 (0.0372)	-0.0708 (0.0284)	-0.0540 (0.0339)	-0.1907 (0.0347)
Log-likelihood	-18722.4	-18948.8	-19136.0	-18468.2
Implied equilibrium p^*	0.465	0.482	0.487	0.452
Two Equilibria Model				
β_1	0.0002 (0.0014)	-0.0003 (0.0004)	0.0009 (0.0015)	0.0017 (0.0018)
α	2.0112 (0.0079)	2.0070 (0.0047)	2.0176 (0.0080)	2.0250 (0.0108)
λ	0.1994 (0.1456)	0.5310 (0.1133)	0.3106 (0.1575)	0.1080 (0.0760)
Log-likelihood	-18710.6	-18947.7	-19125.8	-18464.1
Implied equilibria p_A^*, p_B^*	0.568,0.440	0.519,0.441	0.591,0.438	0.609,0.433
Joint-payoff maximizing p^{JP}	0.979	0.021	0.980	0.980
Test for Multiple Equilibria				
LRTS	5.7	2.2	20.5	8.1
90^{th},95^{th}, 99^{th} percentiles of LRTS distribution	2.7,4.1,8.5	2.5,4.5,8.5	3.0,4.8,9.7	2.8,4.2,11.2
Number of market-days	7,447	7,574	7,563	7,470
Number of station-days	27,105	27,362	27,622	26,819

(b) Non-Drivetime Hours

	3-4 am	12-1 pm	9-10 pm	10-11 pm
One Equilibrium Model				
β_1 (assuming $\alpha = 0$)	-0.1478 (0.0416)	-0.0885 (0.0341)	0.1540 (0.0363)	0.1306 (0.0349)
Log-likelihood	-10116.8	-17503.6	-16968.4	-16039.9
Implied equilibrium p^*	0.463	0.478	0.538	0.533
Two Equilibria Model				
β_1	-0.1478 (0.0416)	-0.0885 (0.0341)	0.0003 (0.0007)	0.1306 (0.0349)
α	-	-	2.0075 (0.0060)	-
λ	-	-	0.3734 (0.1610)	-
Log-likelihood	-10116.8	-17503.6	16968.4	-16039.9
Implied equilibria p_A^*, p_B^*	0.463	0.477	0.561,0.466	0.533
Joint-payoff maximizing p^{JP}	-	-	0.980	-
Test for Multiple Equilibria				
LRTS	0	0	0.1	0
90^{th},95^{th}, 99^{th} percentiles of LRTS distribution	3.0,3.7,8.8	2.4,3.4,5.7	2.3,4.3,6.6	4.2,6.0,11.0
Number of market-days	6,086	7,365	7,268	7,127
Number of station-days	14,653	25,288	24,585	23,212

Notes: Standard errors in parentheses. Log-likelihoods do not include binomial coefficients which are constants and independent of the parameters.