cemmap
centre for microdata methods and practice

Global estimation of finite mixture and misclassification models with an application to multiple equilibria

Yingyao Hu
Ruli Xiao

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP32/18

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

Global estimation of finite mixture and misclassification models with an application to multiple equilibria

Yingyao Hu* Ruli Xiao†

May 26, 2018

Abstract

We show that the identification results of finite mixture and misclassification models are equivalent in a widely-used scenario except an extra ordering assumption. In the misclassification model, an ordering condition is imposed to pin down the precise values of the latent variable, which are also of researchers' interests and need to be identified. In contrast, the identification of finite mixture models is usually up to permutations of a latent index. This local identification is satisfactory because the latent index does not convey any economic meaning. However, reaching global identification is important for estimation, especially, when researchers use bootstrap to estimate standard errors, which may be wrong without a global estimator. We provide a theoretical framework and Monte Carlo evidences to show that imposing an ordering condition to achieve a global estimator innocuously improves the estimation of finite mixture models. As a natural application, we show that games with multiple equilibria fit in our framework and the global estimator with ordering assumptions provides more reliable estimates.

Keywords: Finite mixture, misclassification, global estimation, identification, bootstrap, multiple equilibria.

*Department of Economics, Johns Hopkins University, 3100 Wyman Park Dr, Baltimore, MD 21211; email: y.hu@jhu.edu.

†Department of Economics, Indiana University, 100 S Woodlawn Ave, Bloomington, IN 47405; email:rulixiao@iu.edu.

1 Introduction

Mixture structures arise with the presence of a latent variable, which could be a variable measured with error or unobserved heterogeneity of different sources such as heterogeneous preferences, unobserved heterogeneity within/across markets, different types of beliefs, and multiple equilibria in games. Both finite mixture and misclassification models can be reformulated into similar mixture structures and are widely used in economic applications such as labor economics, industrial organization, and so forth. For example, Keane and Wolpin (1997) consider unobserved type-specific endowments; Hu et al. (2013) control for auction-level unobserved heterogeneity; and Xiao (2014) controls for the presence of multiple equilibria in games. See Hu (2017) for a survey of applications using measurement error and Compiani and Kitamura (2016) for review of finite mixture models.

This paper shows that the identification results of two types of models are equivalent in a widely-used scenario without an ordering assumption. Specifically, both literatures of finite mixture and misclassification models recover the unobserved component-specific distributions through joint distribution of observables and some versions of rank conditions. Moreover, they share a prevalent label swapping issue. The two literatures address this issue differently in accordance with their respective interpretation of the latent variable. In particular, since the latent variable in misclassification models usually carries economic implications, extra conditions are imposed to pin down the precise value of the latent variable. In contrast, the unobserved component in finite mixture models does not convey any economic meaning, so precise location of the unobserved component is not necessary. Consequently, misclassification models reach global identification while finite mixture models reach local identification.

A problem arises, however, when researchers try to use bootstrap to approximate the standard errors of the estimators. Without an appropriate ordering condition, the estimator is a local one in the sense that multiple estimators can generate the same values for the chosen criteria function; thus, it is not straightforward which local estimator should be chosen for each bootstrap resampling. This paper illustrates the problem using both theory and Monte Carlo simulations, and thus advocates imposing an assumption to pin down the order of the latent components by which a global estimator may be obtained, as in misclassification models.

The existing literature on finite mixture models is beginning to realize the importance

and necessity of pinning down the component order when standard error is estimated through resampling. For instance, Kasahara and Shimotsu (2009) suggest that one may determine the component ordering by using the marginal distribution of the component to uniquely pin down the order. Hall et al. (2003) also point out similar treatment. To this end, finite mixture models are very similar to misclassification models. Bonhomme et al. (2016b) also realize that the label swapping issue presents a challenge for inference methods based on resampling algorithms such as bootstrap. However, as of yet, no formal investigation has been made regarding this concern. This paper therefore provides a formal investigation into the potential problem as a caveat to empirical researchers working with finite mixture models who may attempt to use bootstrap or jackknife to approximate estimator variance.

As a natural application, we apply the global estimator to games with multiple equilibria. Games generally admit multiple equilibria, which sometimes is important in explaining various aspects of economic data. Thus, allowing multiple equilibria in game applications is important. Since the labeling of equilibria again does not convey any economic meaning, we can label them in any orders and wouldn't affect the estimation and interpretation of the game payoffs. As a result, imposing the ordering condition is harmless, nonrestrictive and useful in estimation. We look at a game where radio stations choose timing to air their commercials, where multiple equilibria rationalize the clustering patterns of commercial timings in the data. We further see that imposing the ordering condition improves estimation of the standard error via bootstraps.

The remainder of this paper is organized as follows. Section 2 lays out the common framework and shows that the identification results of finite mixture and misclassification models are equivalent in a widely-used scenario except an extra ordering assumption. Section 3 provides a simple theoretical illustration of the problem when the label swapping issue persists. Section 4 employs Monte Carlo simulations to illustrate the problem in finite samples. Section 5 provides an empirical illustration in game with multiple equilibria. Section 6 concludes.

2 A Common Framework

Both finite mixture and misclassification models can be represented through an equation associating observables with unknowns, as follows:

$$f_X = \sum_T f_{X|T} f_T, \tag{1}$$

where f is a probability density or mass function, X represents the observables in the data, and $T \in \{t_1, t_2, \ldots, t_K\}$ can be either the unobserved component in the finite mixture model or the latent true variable in the misclassification model.

A vast literature studies identification and estimation in the two areas. Specifically, several studies (see Hu (2008) and Allman et al. (2009)) focus on the case where there are multiple measurements, i.e., $X = \{X_1, X_2, X_3\}$, which satisfy the following conditional independence condition:

$$X_1 \perp X_2 \perp X_3 \,|\, T. \tag{2}$$

This conditional independence assumption leads to the following representation:

$$f_{X_1 X_2 X_3|T} = f_{X_1|T} f_{X_2|T} f_{X_3|T}. \tag{3}$$

For simplicity, we assume that the cardinality of the unobserved component K is known and is the same as the cardinality of X_i, $i = 1, 2, 3$.

Identification of Finite Mixture models In the finite mixture model, the unobserved component is finite, while the observables in the data can be discrete or continuous. Identification is similar for both continuous and finite observable scenarios by using a three-way array or table and relying on a rank condition. For example, Allman et al. (2009) follow the fundamental algebraic result in Kruskal (1977) to provide conditions for identifying the mixture structures. In particular, in the scenario where X_i has finite state space,[1] they first define a three-dimensional array (tensor) $[\tilde{M}_1, M_2, M_3]$ whose (u, v, w) element is

$$\begin{aligned} [\tilde{M}_1, M_2, M_3]_{u,v,w} &\equiv \sum_j \pi_j p_j^1(u) p_j^2(v) p_j^3(w) \\ &= \Pr(X_1 = u, X_2 = v, X_3 = w), \end{aligned}$$

[1]Note that Allman et al. (2009) do not assume that the X_is are identically distributed conditional on the true T or have the same state space. Here for illustration purpose, we assume that X_i has the same state space but is not necessary identically distributed, conditional on the true T.

where M_i is of sizes $K \times K$ with the jth row defined as $p_j^i = \Pr(X_i = \cdot | T = t_j)$, i.e., for $i = 1, 2, 3$,

$$M_i \equiv \begin{pmatrix} f_{X_i|T}(t_1|t_1) & f_{X_i|T}(t_2|t_1) & ... & f_{X_i|T}(t_K|t_1) \\ f_{X_i|T}(t_1|t_2) & f_{X_i|T}(t_2|t_2) & ... & f_{X_i|T}(t_K|t_2) \\ ... & ... & ... & ... \\ f_{X_i|T}(t_1|t_K) & f_{X_i|T}(t_2|t_K) & ... & f_{X_i|T}(t_K|t_K) \end{pmatrix},$$

π is the marginal probability distribution of T such that $\pi = (\pi_j) \in (0,1)^K$ with $\sum_j \pi_j = 1$, and $\tilde{M}_1 \equiv diag(\pi) M_1$. Therefore the identification boils down to whether we can recover M_is and π using information on tensor $[\tilde{M}_1, M_2, M_3]$. Note that $[\tilde{M}_1, M_2, M_3]$ is invariant to simultaneously permuting the rows of all the M_is and π. Thus, the identification is subject to the label swapping problem.

The identification relies on a rank condition associated with the matrix's Kruskal rank defined in Kruskal (1977). Specifically, the Kruskal rank of a matrix is defined as the largest number I such that every set of I rows of the matrix are independent. Consequently, the Kruskal rank of matrix M is never greater than the regular rank, i.e., $rank_K(M) \leq rank(M)$. Moreover, if matrix M is of full row rank, its Kruskal rank is the same as its regular rank, i.e., $rank_K(M) = rank(M)$. We summarize the identification result in Allman et al. (2009) (Corollary 2) in the following theorem.

Theorem 1 *(Allman et al. (2009)) Consider the model described above. Suppose all entries of π are positive. For each $i = 1, 2, 3$, let $I_i = rank_K(M_i)$. If*

$$I_1 + I_2 + I_3 \geq 2K + 2, \tag{4}$$

the tensor $[\tilde{M}_1, M_2, M_3]$ uniquely determines M_1, M_2, M_3, and π, up to label swapping.

That is, the M_is and π are identified up to a permutation of its support $\{t_1, t_2, \ldots, t_K\}$. Note that the identification does not require the M_is to be full rank.

They further apply this identification result to the scenario of continuous X_i, where the mixture structure of Equation (1) also applies to the corresponding cumulative density function. Let μ_j^i denote the c.d.f. of variable X_i for component j, i.e., $\mu_j^i(x) = \Pr(X_i \leq x | T = t_j)$. Thus, the mixture structure can be represented as

$$\Pr(X_1 <= x_1, X_2 <= x_2, X_3 <= x_3) = \sum_j \pi_j \mu_j^1(x_1) \mu_j^2(x_2) \mu_j^3(x_3). \tag{5}$$

We summarize their identification result (Theorem 8) in the following corollary.

Corollary 1 *(Allman et al. (2009)) Consider the mixture structure described above. If the measures* $\{\mu_j^i\}_{t_K \geq j \geq t_1}$ *are linearly independent, the parameters* π *and the* $\mu_j^i s$ *are strictly identifiable from the joint distribution of* $\{X_1, X_2, X_3\}$*, up to label swapping.*

The identification is achieved by choosing cut points to partition the continuous state space into K exclusive intervals. The idea is that, by partitioning the state space, we can construct a three-way array as in the discrete case. The linear independence of the measures $\{\mu_j^i\}_{t_K \geq j \geq t_1}$ is equivalent to the fact that the corresponding matrix M_i is full row rank. Consequently, the rank condition in Equation (4) is satisfied because I_1=I_2=I_3=K and $K \geq 2$.

The rank condition required in identification using three-way array is less restrictive than a full row rank condition. However, it does not provide a closed-form expression for the identified components. Consequently, we cannot follow the identification procedure to recover the identified mixture components. In fact, the rank condition is comparable to a traditional identification argument that a local identification is feasible if the number of restrictions is larger than or equal to that of unknowns.

Note that the finite mixture literature does not impose additional assumptions for addressing the label swapping problem. There is no need to provide a unique ordering for these mixing components since they do not convey any economics meanings. Consequently, identification with label swapping is not an issue.

Identification of Misclassification Models In the misclassification or the measurement error literature, T represents the latent true variable so it conveys economic meaning itself. For example, in the literature on the returns to education, self-reported education levels might has measurement error such as those who do not go to college may report that their have college degrees. In this case, T would represent different levels of education. Consequently, pinning down the precise value of T is very important.

In the misclassification literature, one of the measurements does not have to have a cardinality at least the same as the cardinality of the latent true variable, as summarized in the so-called 2.1 measurement model in Hu (2017). For easy exposition, we assume that the cardinality of X_1 and X_2 are K, while X_3 is allowed to contain as little information as a binary variable. Using an eigenvalue-eigenvector decomposition, Hu (2008) imposes extra conditions to pin down the precise location of T and achieve the nonparametric global identification of the model. In particular, we introduce the following matrix

representation.

$$\begin{aligned} M_i &= \big[\Pr(X_i = t_k|T = t_j)\big]_{j,k}, \qquad i = 1, 2, 3 \\ A(x_3) &\equiv \big[\Pr(X_1 = t_j, X_2 = t_k, X_3 = x_3)\big]_{j,k}, \\ A &\equiv \big[\Pr(X_1 = t_j, X_2 = t_k)\big]_{j,k}, \\ \Omega &\equiv diag\big(\pi_1, ..., \pi_K\big), \\ D(x_3) &\equiv diag\big(Pr(X_3 = x_3|T = t_1), ..., \Pr(X_3 = x_3|T = t_K)\big). \end{aligned}$$

Consequently, we have the following two matrix representations:

$$\begin{aligned} A &= M_1^T \Omega M_2, \\ A(x_3) &= M_1^T D(x_3) \Omega M_2. \end{aligned}$$

With a full rank condition, i.e., M_1 and M_2 are invertible, we have

$$A(x_3)A^{-1} = M_1^T D(x_3)(M_1^T)^{-1}.$$

Note that the observed matrix $A(x_3)A^{-1}$ in Hu (2008) is allowed to be asymmetric while symmetry plays an important role in Bonhomme et al. (2016a).[2] For the decomposition to be unique, the eigenvalues must vary with the latent index t_j for some value of x_3. Notice that the eigenvectors do not change with x_3. One sufficient condition for distinctive eigenvalues is as follows:

Assumption 1 *(Distinctive eigenvalues) For any $t_j \neq t_k$, there exist an x_3 such that*

$$\Pr(X_3 = x_3|T = t_j) \neq \Pr(X_3 = x_3|T = t_k).$$

Assumption 1 implies that for any two rows of the M_3 matrix defined above, there exists at least a column, i.e., a value x_3, that the probabilities in the two rows differ from each other. This assumption is less restrictive than the usual assumption that the exists a column so that probabilities of every row, i.e., corresponding to every t_j, differs from each other. Moreover, this assumption is empirically testable since matrix $A(x_3)A^{-1}$ can be estimated directly from the data.

The label swapping issue is also prevalent in the identification of misclassification models. Hu (2008) provides a set of flexible ordering conditions to pin down the value of

[2]Bonhomme et al. (2016a) also use an eigenvalue-eigenvector decomposition technique to obtain identification in the setting of a finite mixture model. They assume M_1, M_2, and M_3 are the same.

the latent true variable to achieve global identification. These conditions are also quite intuitive in various economic contexts. Specifically, Assumption 2 provides a few examples of conditions to fix the order of the eigenvalues or the eigenvectors.

Assumption 2 *(Ordering) one of the following conditions holds:*

1) $f_{X_1|T}(.|t)$ *has a unique mode at* t;

2) $f_{X_1|T}(.|t)$ *has a median (min, max, or a known quantile) at* t;

3) there exist t_i *such that* $f_{X_1|T}(t_i|t)$ *is increasing in* t*;*

4) $E[\omega(X_3)|T = t]$ *is increasing in* t *for a known function* $\omega(.)$.

The function in condition 4 may be user-specific, such as $\omega(X_3) = X_3$ or $\omega(X_3) = I(X_3 = x_3)$, where $I(.)$ is the indicator function. We summarize the global identification result of misclassification models as follows:

Theorem 2 *(Hu (2008)) Consider a structure described as Equation (1). Suppose that matrix* M_1 *and* M_2 *have a full rank, and that Assumptions 1 and 2 are satisfied. Then,* M_1*,* M_2*,* M_3*, and* π *are uniquely identified.*

Note that the condition that matrix M_1 and M_2 have a full rank implies that the Kruskal ranks of matrices M_1 and M_2 are equal to the regular matrix rank K. In the scenario with a binary X_3, Assumption 1, which guarantees distinctive eigenvalues, holds if and only if the so-called Kruskal rank of M_3 is equal to 2. Therefore, the total Kruskal rank equals $2K + 2$. We further prove the equivalence of the two rank conditions for a general discrete X_3 in the following theorem.

Sufficient and Necessary Conditions for Identification This paper connects the identification results of finite mixture models and those of misclassification models in a widely-used scenario in empirical research. Below we first present a set of sufficient and necessary conditions for identification of the models, then we provide detailed discussions. Our main results are summarized as follows:

Theorem 3 *Consider a structure described as Equation (1). Suppose that* M_1 *and* M_2 *have a full rank, i.e.,* $I_1 = I_2 = K$*. The following four statements are equivalent:*

1. *(Nontrivial Kruskal rank) The Kruskal rank* I_3 *of* M_3 *satisfies* $I_3 \geq 2$*;*

2. *(Distinctive eigenvalues) Assumption 1 holds, i.e., for any $t_j \neq t_k$, there exist an x_3 such that*

$$\Pr(X_3 = x_3|T = t_j) \neq \Pr(X_3 = x_3|T = t_k).$$

3. *(Distinctive eigenvalues) there exist a function $\omega(.)$ such that*

$$E[\omega(X_3)|T = t_j] \neq E[\omega(X_3)|T = t_k]$$

for any $t_j \neq t_k$

4. *(Uniqueness) M_1, M_2, M_3, and π are uniquely identified, up to label swapping.*

Proof: We show that these four statements are equivalent in three steps.

Step 1: Statement 1 holds if and only if statement 2 does. First, we show statement 2, i.e., Assumption 1, implies statement 1, i.e., $I_3 \geq 2$. For a general discrete X_3 with support $\{t_1, t_2, \ldots, t_K\}$[3] Note that M_3 is defined as

$$M_3 = \begin{pmatrix} f_{X_3|T}(t_1|t_1) & f_{X_3|T}(t_2|t_1) & ... & f_{X_3|T}(t_K|t_1) \\ f_{X_3|T}(t_1|t_2) & f_{X_3|T}(t_2|t_2) & ... & f_{X_3|T}(t_K|t_2) \\ ... & ... & ... & ... \\ f_{X_3|T}(t_1|t_K) & f_{X_3|T}(t_2|t_K) & ... & f_{X_3|T}(t_K|t_K) \end{pmatrix}.$$

We can show that the Kruskal rank of M_3 is at least 2 if and only if for any $t_j \neq t_k$ there exists a $x_3 \in \{t_1, t_2, ..., t_K\}$ such that $f_{X_3|T}(x_3|t_j) - f_{X_3|T}(x_3|t_k) \neq 0$. For any two rows with $t_j \neq t_k$, we consider the following matrix of dimensions $2 \times K$

$$M_{3.2} \equiv \begin{pmatrix} f_{X_3|T}(t_1|t_j) & f_{X_3|T}(t_2|t_j) & ... & f_{X_3|T}(t_K|t_j) \\ f_{X_3|T}(t_1|t_k) & f_{X_3|T}(t_2|t_k) & ... & f_{X_3|T}(t_K|t_k) \end{pmatrix}.$$

Without loss of generality, let $x_3 = t_m$. Define $\mathbf{1} = (1, 1, ..., 1)^T$ and $e_m = (0, ..., 0, 1, 0, ..., 0)^T$, where 1 is at the m-th coordinate. We consider

$$M_{3.2} \times (e_m \quad \mathbf{1}) = \begin{pmatrix} f_{X_3|T}(x_3|t_j) & 1 \\ f_{X_3|T}(x_3|t_k) & 1 \end{pmatrix}.$$

Therefore, the rank of $M_{3.2}$ equals 2 if $f_{X_3|T}(x_3|t_j) - f_{X_3|T}(x_3|t_k) \neq 0$. That means the Kruskal rank of M_3 is larger than or equal to 2.

[3]Note that X_3 does not have to have the same cardinality as T. Theorem 3 holds for a general X_3 with a cardinality not less than 2.

Second, we show that $I_3 \geq 2$ implies Assumption 1. If the Kruskal rank of M_3 is at least 2, the regular rank of matrix $M_{3.2}$ for any $t_j \neq t_k$ equals 2. That means there must exist a column, say m, in $M_{3.2}$ such that $f_{X_3|T}(t_m|t_j) - f_{X_3|T}(t_m|t_k) \neq 0$. Assumption 1 then holds with $x_3 = t_m$ for the given $t_j \neq t_k$.

Step 2: Statement 2 holds if and only if statement 4 does. First, we show statement 2 implies statement 4. We have shown that statement 2 implies statement 1, i.e., $I_3 \geq 2$. Since $I_1 + I_2 + I_3 \geq 2K + 2$, Allman et al. (2009) shows the unique identification up to label swapping, i.e., statement 4. Next, we show that if statement 2 does not hold, neither does statement 4. If the assumption of distinctive eigenvalues fails, it indicates that at least two different rows of matrix M_3 are the same. The eigenvalue-eigenvector decomposition as in Hu (2008) implies that if the distinctive eigenvalues assumption fails, e.g., $f_{X_3|T}(.|t_j) = f_{X_3|T}(.|t_k)$, any convex combination of the eigenvectors, e.g., $f_{X_1|T}(.|t_j)$ and $f_{X_1|T}(.|t_k)$, corresponding to the same eigenvalue is an eigenvector. Therefore, the eigenvectors in M_1 is not uniquely identified, i.e., statement 4 does not hold.

Step 3: Statement 2 holds if and only if statement 3 does. First, we show statement 3 implies statement 2. This can be shown by contradiction. Suppose statement 2 do not hold. That means the distributions $f_{X_3|T}(.|t_j)$ is the same as $f_{X_3|T}(.|t_k)$. Then $E[\omega(X_3)|T = t_j] = E[\omega(X_3)|T = t_k]$ for any function ω, which is contradictory to statement 3.

Next, we show that statement 2 implies statement 3. Define a vector

$$Df_{j,k} \equiv \begin{pmatrix} f_{X_3|T}(t_1|t_j) - f_{X_3|T}(t_1|t_k) \\ f_{X_3|T}(t_2|t_j) - f_{X_3|T}(t_2|t_k) \\ ... \\ f_{X_3|T}(t_K|t_j) - f_{X_3|T}(t_K|t_k) \end{pmatrix}.$$

Assumption 1 in statement 2 guarantees that $Df_{j,k} \neq 0$ for all j and k. Therefore, there exists a vector $W = (w_1, w_2, ..., w_K)'$ such that W is not orthogonal to either of $Df_{j,k}$ for any $j \neq k$. [4] That is

$$W' \times Df_{j,k} \neq 0, \quad j \neq k, j, k = 1, 2, ..., K.$$

[4]The existence of such a vector W can be shown by contradiction. Suppose such a vector W does not exist. Then for any W, there exists a pair (j, k) with $k \neq j$ such that $W' \times Df_{j,k} = 0$. Note that the left hand side the that equation, i.e., $W' \times Df_{j,k}$, is continuous in W. Thus, for that given pair of (j, k), there must exist K linearly independent W^1,...,W^K such that $[(W^1)', ..., (W^K)']' \times Df_{j,k} = 0$. As a result, we have $Df_{j,k} = 0$. A contradiction.

Therefore, we can define a function of X_3 as

$$\omega(X_3) = \sum_{i=1,2,...,K} w_i \times I(X_3 = t_i)$$

which satisfies for any $t_j \neq t_k$,

$$E[\omega(X_3)|T = t_j] - E[\omega(X_3)|T = t_k] = W' \times Df_{j,k} \neq 0.$$

■

Although we focus on the case where M_1 and M_2 are full rank, we are able to provide a set of sufficient and necessary conditions for identification up to label swapping, while Allman et al. (2009) provide sufficient conditions for the identification of a more general case. Statement 3 provides a transparent explanation of the distinctive-eigenvalue assumption. Notice that the eigenvalue-eigenvector decomposition still holds with eigenvalue matrix $D(x_3)$ replaced with:

$$D(\omega) = diag\big(E[\omega(X_3)|T = t_1], ..., E[\omega(X_3)|T = t_K]\big).$$

Specifically,

$$A(\omega)A^{-1} = M_1^T D(\omega)(M_1^T)^{-1},$$

where

$$A(\omega) \equiv \big[E[\omega(X_3)|X_1 = t_j, X_2 = t_k] \Pr(X_1 = t_j, X_2 = t_k)\big]_{j,k}.$$

Statement 3 guarantees that the eigenvalues $E[\omega(X_3)|T = t_j]$ in $D(\omega)$ are distinctive. Therefore, the eigenvector matrix M_1 corresponding to each t_j is uniquely identified up to the label t_j. That means the eigenvector matrix M_1 is uniquely identified up to label swapping.

It is natural to compare these results for the two models. Allman et al. (2009) provide local identification results for a more general setting with less intuitive conditions. The general setting may include the case where each of the measurements can have a smaller support than that of the latent true variable. For example, when the latent true variable has 5 possible values, identification is still feasible when researchers only observe 4 possible values in each of the three measurements. That is a scenario not considered in Hu (2008) or the current paper. Meanwhile, their conditions are based on the abstract Kruskal rank, which may be difficult to test and interpret with economic models. In addition, it is not clear how to extend the Kruskal rank condition to the continuous case. Finally, it is well known that the local identification results cause problems in estimation.

Hu (2008) presents global identification results with conditions which are more intuitive and applicable to empirical research. In particular, the ordering assumption may have useful economic implications. Additionally, the regular rank condition is directly testable from the data and can be intuitively extended to the injectivity condition the continuous case as in Hu and Schennach (2008). More importantly, Assumption 2 is innocuous and very intuitive, and can transform a local identification into a global identification, which is very helpful in estimation. Therefore, one should impose Assumption 2 even in the estimation of finite mixture models.

3 Global Estimation and Bootstrap

With the model identified, one can use a minimum Hellinger distance estimator (MHD) defined by minimizing the distance of the joint distribution directly from the data $\hat{f}$ and predicted by the models f, respectively. The MHD estimator for finite mixture models can be represented as:

$$\left(f_{M_1|T}, f_{M_2|T}, f_{M_3|T}, f_T\right) = \arg \min_{f_{M_1|T}, f_{M_2|T}, f_{M_3|T}, f_T} \|\hat{f}^{1/2}_{M_1M_2M_3} - (\Sigma f_{M_1|T} f_{M_2|T} f_{M_3|T} f_T)^{1/2}\|, \tag{6}$$

and the MHD estimator for misclassification models can be represented as:

$$\begin{aligned} \left(f_{M_1|T}, f_{M_2|T}, f_{M_3|T}, f_T\right) \quad = \quad & \arg \min_{f_{M_1|T}, f_{M_2|T}, f_{M_3|T}, f_T} \|\hat{f}^{1/2}_{M_1M_2M_3} - (\Sigma f_{M_1|T} f_{M_2|T} f_{M_3|T} f_T)^{1/2}\|, \\ & s.t. \quad \text{Assumption 2 holds} \end{aligned} \tag{7}$$

where $\|\cdot\|$ represents the L_2 norm. Since the finite mixture model is identified up to a permutation of T, the estimator is a local estimation in the sense that there are $K!$ minima of the criterion function and these minima all lead to the same value for the criteria function. The estimator for the misclassification model is a global one because it directly pins down which minima is the correct one. This may not seem to be a problem because the permutations of the T types do not matter economically. However, such identification up to a permutation makes the bootstrap method invalid because it is not clear which local minimum the estimator reaches in each bootstrap draw. Therefore, we argue that it is still better off to impose Assumption 2 in the estimation of the finite mixture model, i.e., treating it as a misclassification model.

Remark 1: This label swapping issue is a problem for more than just minimum distance estimation. It is a prevalent problem due to the identification strategy, and thereby affects every estimator.

Remark 2: Some may argue that we do not need to worry about this problem if one can derive the variance-covariance matrix for the estimator theoretically. However, some applications, especially applications such as dynamic discrete choice models or dynamic games, rely heavily on a sequential estimation approach to estimate structural parameters while also requiring controls for unobserved heterogeneity. In those applications, deriving the variance matrix is very challenging and maybe infeasible. Thus, bootstrap is a popular alternative for standard deviation approximation. The label swapping issue again causes similar problems in these scenarios.

A global estimator is always preferable if possible. The rest of this section illustrates the problem one might encounter if using bootstrap to estimate the standard error for the finite mixture model without imposing the ordering condition, which is a local estimator. Let θ denote all unknown parameters, i.e., $\theta_0 = \left(f_{M_1|T}, f_{M_2|T}, f_{M_3|T}, f_T\right)$. For ease of exposition, assume that T is binary so the minimization has two local minima. In particular, one minima orders the components estimated as t_1 and t_2, respectively, while the other minima orders the components estimated as t_2 and t_1, respectively. In particular, minimum 1 is denoted as

$$\theta_0^1 \equiv \left(\{f_{M_1|T=t_1}, f_{M_1|T=t_2}\}, \{f_{M_2|T=t_1}, f_{M_2|T=t_2}\}, \{f_{M_3|T=t_1}, f_{M_3|T=t_2}\}, \{f_{T=t_1}, f_{T=t_2}\}\right),$$

and minimum 2 is denoted as

$$\theta_0^2 \equiv \left(\{f_{M_1|T=t_2}, f_{M_1|T=t_1}\}, \{f_{M_2|T=t_2}, f_{M_2|T=t_1}\}, \{f_{M_3|T=t_1}, f_{M_3|T=t_1}\}, \{f_{T=t_2}, f_{T=t_1}\}\right).$$

We further use $\hat{\theta}_n^1$ and $\hat{\theta}_n^2$ to denote the estimators for the two minima, respectively, given a sample. We first state the asymptotic property of the estimator in the following lemma; refer to the existing literature for rigorous proof, e.g., Tang and Karunamuni (2013).

Lemma 1 *(**Asymptotic property**) With some regularity conditions, the minimal distance estimator is consistent and asymptotically normally distributed when the sample size goes to infinity. That is,*

$$\begin{aligned}\sqrt{n}(\hat{\theta}^1 - \theta_0^1) &\rightarrow N(0, \Sigma_1),\\ \sqrt{n}(\hat{\theta}^2 - \theta_0^2) &\rightarrow N(0, \Sigma_2),\end{aligned}$$

where Σ_1 and Σ_2 are the variance matrices of minima $\hat{\theta}^1$ and $\hat{\theta}^2$, respectively. Ideally, if there was an easy way to estimate the variance-covariance matrix Σ_1 or Σ_2, it would not be a problem to ignore the ordering of the unobserved components. However, direct estimation of the variance-covariance matrix may not always be feasible. An alternative method would be to treat the sample in hand as a population and resample the data to approximate the variance-covariance matrix. Specifically, an approximation of the variance-covariance matrix can be obtained by a sample of bootstrap estimators denoted as $\hat{\theta}_b, b = 1, ..., B$, which is obtained by sampling n observations with replacement from the original data and recomputing $\hat{\theta}$ for each new sample. After B times resampling, one can approximate the asymptotic covariance matrix of the estimator $\hat{\theta}$ with

$$Est.Asy.\Sigma = \frac{1}{B-1}\sum_b [\hat{\theta}_b - \bar{\theta}][\hat{\theta}_b - \bar{\theta}]',$$

where $\bar{\theta} = \frac{1}{B}\sum_b \hat{\theta}_b$ represents the mean of the estimates across the B bootstrap samples (Efron et al. (1979)). In the scenario with multiple minima, the variance matrix is correctly approximated with an appropriate ordering condition. Consequently, we can approximate the variance matrix for both minima using

$$Est.Asy.\Sigma^1 = \frac{1}{B-1}\sum_b [\hat{\theta}_b^1 - \bar{\theta}^1][\hat{\theta}_b^1 - \bar{\theta}^1]' \quad (8)$$

$$\text{and} \qquad Est.Asy.\Sigma^2 = \frac{1}{B-1}\sum_b [\hat{\theta}_b^2 - \bar{\theta}^2][\hat{\theta}_b^2 - \bar{\theta}^2]', \quad (9)$$

respectively.

However, the lack of an ordering condition could generate a mixture of the estimator in bootstrap. The local minimum estimated in each resampling may not be the same one, and the probability of getting different minima is unknown. For illustration purposes, let us assume that we reach minimum $\hat{\theta}_b^1$ and $\hat{\theta}_b^2$ with probability λ and $1-\lambda$, respectively. That is,

$$\hat{\theta}_b = \begin{cases} \hat{\theta}_b^1, \text{ with a probability of } \lambda, \\ \hat{\theta}_b^2, \text{ with a probability of } 1-\lambda. \end{cases}$$

Essentially the estimator obtained from resampling b is a mixture of the two minima. This mixture feature will cause problems when one uses the variance matrix from the bootstrap to approximate the asymptotic variance matrix. In particular,

$$Est.Asy.\Sigma = \frac{1}{B-1}\sum_b [\hat{\theta}_b - \bar{\theta}][\hat{\theta}_b - \bar{\theta}]', \quad (10)$$

which basically is to approximate the population variance-covariance matrix of the mixture

$$\begin{aligned} var(\hat{\theta}_b) &= E[var(\hat{\theta}_b|\text{minimum})] + var[E(\hat{\theta}_b|\text{minimum})] \\ &= \lambda\Sigma_1 + (1-\lambda)\Sigma_2 + \lambda(1-\lambda)[\theta^1 - \theta^2][\theta^1 - \theta^2]'. \end{aligned}$$

The second equality holds because

$$E(\hat{\theta}_b|minimum) = \begin{cases} \theta^1, \text{ if minimum is 1 }, \\ \theta^2, \text{ if minimum is 2 }. \end{cases}$$

Consequently, the variance-covariance matrix approximated by bootstrap resampling without an ordering condition differs from the targeted population variance matrix, which is either Σ_1 or Σ_2, depending on the original estimate. Specifically, the bias from the approximation can be computed as

$$\begin{aligned} var(\hat{\theta}_b) - \Sigma_1 &= (1-\lambda)\big[\Sigma_2 - \Sigma_1 + \lambda[\theta^1 - \theta^2][\theta^1 - \theta^2]'\big], \text{ if minimum is 1,} \\ \text{and} \quad var(\hat{\theta}_b) - \Sigma_2 &= \lambda\big[\Sigma_1 - \Sigma_2 + (1-\lambda)[\theta^1 - \theta^2][\theta^1 - \theta^2]'\big], \text{ if minimum is 2.} \end{aligned}$$

It is not straightforward to check whether the variance from bootstrap without ordering over-estimates or under-estimates the targeted population matrix. That is, are $var(\hat{\theta}_b) - \Sigma_1$ and $var(\hat{\theta}_b) - \Sigma_2$ negative definite, positive definite, or neither?

We further simplify the framework to investigate further the problem of using the variance from bootstrap to approximate the targeted population variance matrix. Suppose the components have the same distributions, i.e.,

$$f_{M_1T} = f_{M_2|T} = f_{M_3|T} = \begin{pmatrix} p_1 & p_2 \\ 1-p_1 & 1-p_2 \end{pmatrix},$$

and

$$f_T = (\pi, 1-\pi)\,.$$

Consequently, $\theta^1 = (p_1, p_2, \pi)$, and $\theta^2 = (p_2, p_1, 1-\pi)$; $\hat{\theta}^1 = (\hat{p}_1, \hat{p}_2, \hat{\pi})$ and $\hat{\theta}^2 = (\hat{p}_2, \hat{p}_1, 1-\hat{\pi})$ are estimators for θ^1 and θ^2, respectively. We also assume

$$\Sigma_1 = var(\hat{\theta}^1) \equiv \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{1\pi} \\ \sigma_{12} & \sigma_2^2 & \sigma_{2\pi} \\ \sigma_{1\pi} & \sigma_{2\pi} & \sigma_\pi^2 \end{pmatrix}.$$

Then

$$\Sigma_2 = var(\hat{\theta}^2) = \begin{pmatrix} \sigma_2^2 & \sigma_{12} & \sigma_{2\pi} \\ \sigma_{12} & \sigma_1^2 & \sigma_{1\pi} \\ \sigma_{2\pi} & \sigma_{1\pi} & \sigma_\pi^2 \end{pmatrix}.$$

Without loss of generality, we assume that $\sigma_2^2 > \sigma_1^2 > 0$. We compute the difference between the variance from bootstrap without ordering and the targeted population variance through:

$$\begin{aligned} var(\hat{\theta}_b) - \Sigma_1 &= (1-\lambda)\big[\Sigma_2 - \Sigma_1 + \lambda[\theta^1 - \theta^2][\theta^1 - \theta^2]'\big], \\ &= (1-\lambda)\Bigg[\begin{pmatrix} \sigma_2^2 & \sigma_{12} & \sigma_{2\pi} \\ \sigma_{12} & \sigma_1^2 & \sigma_{1\pi} \\ \sigma_{2\pi} & \sigma_{1\pi} & \sigma_\pi^2 \end{pmatrix} - \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{1\pi} \\ \sigma_{12} & \sigma_2^2 & \sigma_{2\pi} \\ \sigma_{1\pi} & \sigma_{2\pi} & \sigma_\pi^2 \end{pmatrix} \\ &\quad +\lambda \begin{pmatrix} (p_1-p_2)^2 & -(p_1-p_2)^2 & (p_1-p_2)(2\pi-1) \\ -(p_1-p_2)^2 & (p_1-p_2)^2 & -(p_1-p_2)(2\pi-1) \\ (p_1-p_2)(2\pi-1) & -(p_1-p_2)(2\pi-1) & (2\pi-1)^2 \end{pmatrix}\Bigg] \\ &= (1-\lambda)\begin{pmatrix} \sigma_2^2 - \sigma_1^2 + \lambda(p_1-p_2)^2 & -\lambda(p_1-p_2)^2 & \Sigma_{12\pi} \\ -\lambda(p_1-p_2)^2 & \sigma_1^2 - \sigma_2^2 + \lambda(p_1-p_2)^2 & -\Sigma_{12\pi} \\ \Sigma_{12\pi} & -\Sigma_{12\pi} & \lambda(2\pi-1)^2 \end{pmatrix}, \end{aligned}$$

where $\Sigma_{12\pi} = \sigma_{2\pi} - \sigma_{1\pi} + \lambda(p_1 - p_2)(2\pi - 1)$. The variance from bootstrap without an ordering condition tends to over-estimate the targeted population variance. First, the bootstrap variance without ordering over-estimates the variance of the type probability (σ_π^2) because $\lambda(2\pi-1)^2 \geq 0$. Second, since $\sigma_2^2 > \sigma_1^2$, variance from bootstrap without ordering also over-estimates the variance of the first element (σ_1^2) because $\sigma_2^2 - \sigma_1^2 + \lambda(p_1 - p_2)^2$. However, bootstrap without the ordering could result in an under-estimation of variance if $\sigma_1^2 - \sigma_2^2 + \lambda(p_1 - p_2)^2 < 0$. The intuition here is that, without the ordering condition, different draws in the bootstrap procedure could reach different minima so that the variance computed in this fashion is a mixture of variance from different minima, which is a weighted average of the variance of different minima ($\lambda\Sigma_1 + (1-\lambda)\Sigma_2$) with extra noise ($\lambda(1-\lambda)[\theta^1 - \theta^2][\theta^1 - \theta^2]'$).

In practice, the process of approximating the variance matrix from bootstrap without an ordering condition is very complicated. The probability of getting a given local minimum in each draw of bootstrap could vary, which thereby complicates understanding how the lack of an ordering condition affects the approximation. Moreover, different

softwares might have different algorithms so that probability may also vary with software. This difficulty increases with the cardinality of the latent components since the number of minima increases exponentially.

4 Monte Carlo Evidences

This section illustrates the problem of variance estimation using bootstrap without an ordering condition through Monte Carlos simulations. Consider a simple example where T is binary, i.e., $T \in \{1, 2\}$, and all M_1, M_2, M_3 are also binary, and the mixture distributions are the same for all M_1, M_2, M_3. Since the latent component is binary, there exist two minima. Both minimize the moment conditions in Equation (6). If we follow the misclassification models, we can impose a condition that the diagonal elements in the mixture matrix are bigger than the off-diagonal elements, resulting in a global minima. We consider the following scenario:

$$f_{M_1T} = f_{M_2|T} = f_{M_3|T} = \begin{pmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{pmatrix} \quad \& \quad f_T = (0.2, 0.8)\,,$$

with $\theta^1 = \{0.8, 0.1, 0.2\}$ and $\theta^2 = \{0.1, 0.8, 0.8\}$ as the two minima.

To illustrate the invalidity of the bootstrap method without ordering conditions, we simulate the data, estimate the parameters $\theta = \{p_1, p_2, \pi\}$ from minimizing distance between the left and right-hand sides of the equations, and approximate the variance by resampling with replacement 1000 times. In the scenario without an ordering condition, we introduce a selection rule of getting local minimum 1 with a probability of 0.5, i.e., $\lambda = 0.5$. We provide histograms for the estimates of bootstrap with and without ordering conditions for sample size=4000 in Figure (1). We also provide the variance of the estimators for sample sizes=500, 1000, 1500, and present the results in Table (1).

From the results, we can clearly see that estimates without an ordering condition are inconsistent and biased. Moreover, the standard error from bootstrap with an ordering condition is much smaller than without an ordering condition but with random selection.

5 Multiple Equilibria

This section illustrates the importance of the ordering condition in estimation of games with multiple equilibria. The application uses a simultaneously move game to characterize

the timing decisions for broadcasting commercials by radio stations with contemporary music formats (Contemporary Hit Radio (CHR)/Top 40, Country, Rock etc.).

5.1 Data and Model

In reality, stations tend to play commercials at the same time (figure 2). One possible explanation is that some time intervals are more desirable for commercials. Note that Arbitron uses the same methodology to compute listenerships, which determines firms' willingness to pay for commercials. Suppose this indeed drives the clustering phenomenon, one then can expect that every market has the same pattern. This, however, is not the case (Figure 3). Thus, common factors cannot fully explain the clustering pattern of commercials. Another possible explanation is the presence of multiple equilibria. Stations coordinate to air their commercials at the same time to avoid listener switching, and different markets coordinate at different times, which indicates that they employ different equilibria. This rationalizes both the clustering pattern in general and the clustering in the different times across markets.

We model stations' decisions as to choose from two time blocks in every hour simultaneously, as in Sweeting (2009) and Xiao (2014). Specifically, we use information about whether commercials are being played at two particular time interval in each hour, and , denoted as option 0(: 48− : 52) and option 1(: 53− : 57), respectively. We assume further that stations are symmetric, and station i's payoff for placing a commercial in time block $t \in \{0, 1\}$ is defined as follows:

$$\begin{aligned} \pi(a_i = 1, a_{-i}) &= \alpha + \delta \frac{\sum_{j \neq i} I(a_j = 1)}{n - 1} + \epsilon_{i1}, \\ \pi(a_i = 0, a_{-i}) &= \delta \frac{\sum_{j \neq i} I(a_j = 0)}{n - 1} + \epsilon_{i0}, \end{aligned}$$

where α allows different average profit for stations when they play their commercials in timing 1, δ captures the coordination incentives, and ϵ's represent the idiosyncratic private profit shocks, which the Stations receive before they make their timing decisions. The ϵ's represent the fact that a station tends to play commercials at different times every day. This introduces variation due to the length of other programming, such as songs or travel news, can vary and be unpredictable. We assume ϵ_{it} to be independent across actions, players and markets. Furthermore, ϵ_{it} follows with a type one extreme value distribution.

Following the existing literature, we use the probability that firms choose time slots 0

and 1 to characterize the equilibrium conditions, denoted as p_0 and p_1, respectively. Since radio stations are assumed to be homogenous, we focus on symmetric equilibria. As a result, the equilibrium condition can be represented as:

$$\begin{aligned} p_1 &= \int I(\alpha + \delta p_1 + \epsilon_{i1} > \delta p_0 + \epsilon_{i0}) dF(\epsilon_i) \\ &= \frac{exp(\alpha + \delta p_1)}{exp(\delta p_0) + exp(\alpha + \delta p_1)}. \end{aligned} \tag{11}$$

The second equality holds due to the assumption that ϵ_i follows an extreme value distribution.

The data used in this paper are constructed using hourly airplay logs collected by Medabase 24/7 and extracted from airplay logs that stations play on a minute-by-minute basis[5]. In summary, there are 144 markets in total; the number of stations in each market varies from 3 to 15 with a mean of 5.7; each station has 236 observations, including 59 days (Table 2).

5.2 Estimation and Results

Note that we can pool markets with different number of radio stations for estimation because only the proportion of radio stations' timing instead of the number of firms enters the payoff function from the equilibrium condition. We thus use markets with at least three players to obtain a reasonable size of observations.

The number of equilibria is estimated to be two[6]. We then estimate the equilibrium CCPs using the proposed minimum distance estimation. To illustrate the problem of local estimators, we estimate the two equilibrium CCPs with and without imposing an ordering condition, as in the Monte Carlo experiment. We present the results in table 3. The standard deviation estimated from imposing the ordering condition is significantly smaller than that without the ordering condition. Note that because the labeling of equilibrium does not convey any economic meaning. The estimation of the payoff primitives are the same with or without imposing the ordering condition. Consequently, the ordering condition does not impose any restrictions on the payoff primitives. Both equilibrium CCPs satisfy the same equilibrium conditions. We skip the estimation of the payoff primitives here since it is not the focus of this paper.

[5]Thanks Andew Sweeting for sharing the data. Please refer to Sweeting (2009) for a detailed description of the data.

[6]Please refer to Xiao (2014) for estimating the number of equilibria.

6 Conclusion

This paper connects the identification results of finite mixture models and those of misclassification models in a widely-used scenario in empirical research. While existing studies provide sufficient identification conditions for a more general case, we present sufficient and necessary conditions for the identification of this widely-used case, up to label swapping of the latent values. In the misclassification model, an ordering condition is usually imposed to pin down the precise value of the latent variable, which are also of researchers' interests and need to be identified. In contrast, the identification of finite mixture models is usually up to label swapping. We argue that the ordering condition in misclassification models leads to global identification and should be imposed in estimation, especially, when researchers use bootstrap to estimate standard errors. That is to treat a finite mixture model as a misclassification model in estimation with an ordering condition. As an empirical application, games with multiple equilibria fit in our framework well and we show that the global estimator with ordering assumptions provides reliable estimates with real data.

References

Allman, Elizabeth S, Catherine Matias, and John A Rhodes, "Identifiability of parameters in latent structure models with many observed variables," *The Annals of Statistics*, 2009, pp. 3099–3132. 2, 2, 1, 1, 1, 2

Bonhomme, Stéphane, Koen Jochmans, and Jean-Marc Robin, "Estimating multivariate latent-structure models," *The Annals of Statistics*, 2016, *44* (2), 540–563. 2

_ , _ , and _ , "Non-parametric estimation of finite mixtures from repeated measurements," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2016, *78* (1), 211–229. 1

Compiani, Giovanni and Yuichi Kitamura, "Using mixtures in econometric models: a brief review and some new results," *The Econometrics Journal*, 2016, *19* (3). 1

Efron, B et al., "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, 1979, *7* (1), 1–26. 3

Hall, Peter, Xiao-Hua Zhou et al., "Nonparametric estimation of component distributions in a multivariate mixture," *The annals of statistics*, 2003, *31* (1), 201–224. 1

Hu, Yingyao, "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution," *Journal of Econometrics*, 2008, *144* (1), 27–61. 2, 2, 2, 2, 2

_ , "The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics," *Journal of Econometrics*, 2017, *200* (2), 154–168. 1, 2

_ **and Susanne M Schennach**, "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, 2008, *76* (1), 195–216. 2

_ **, D McAdams, and M Shum**, "Nonparametric identification of first-price auctions with non-separable unobserved heterogeneity," *Journal of Econometrics*, 2013, *174* (2), 186–193. 1

Kasahara, Hiroyuki and Katsumi Shimotsu, "Nonparametric identification of finite mixture models of dynamic discrete choices," *Econometrica*, 2009, *77* (1), 135–175. 1

Keane, Michael P and Kenneth I Wolpin, "The career decisions of young men," *Journal of political Economy*, 1997, *105* (3), 473–522. 1

Kruskal, Joseph B, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear algebra and its applications*, 1977, *18* (2), 95–138. 2

Sweeting, A., "The strategic timing incentives of commercial radio stations: An empirical analysis using multiple equilibria," *The RAND Journal of Economics*, 2009, *40* (4), 710–742. 5.1, 5, 2, 3

Tang, Qingguo and Rohana J Karunamuni, "Minimum distance estimation in a finite mixture regression model," *Journal of Multivariate Analysis*, 2013, *120*, 185–204. 3

Xiao, Ruli, "Identification and estimation of finite action games with multiple equilibria," 2014. 1, 5.1, 6

Appendix

A Graphs and Tables

Figure 1: histogram

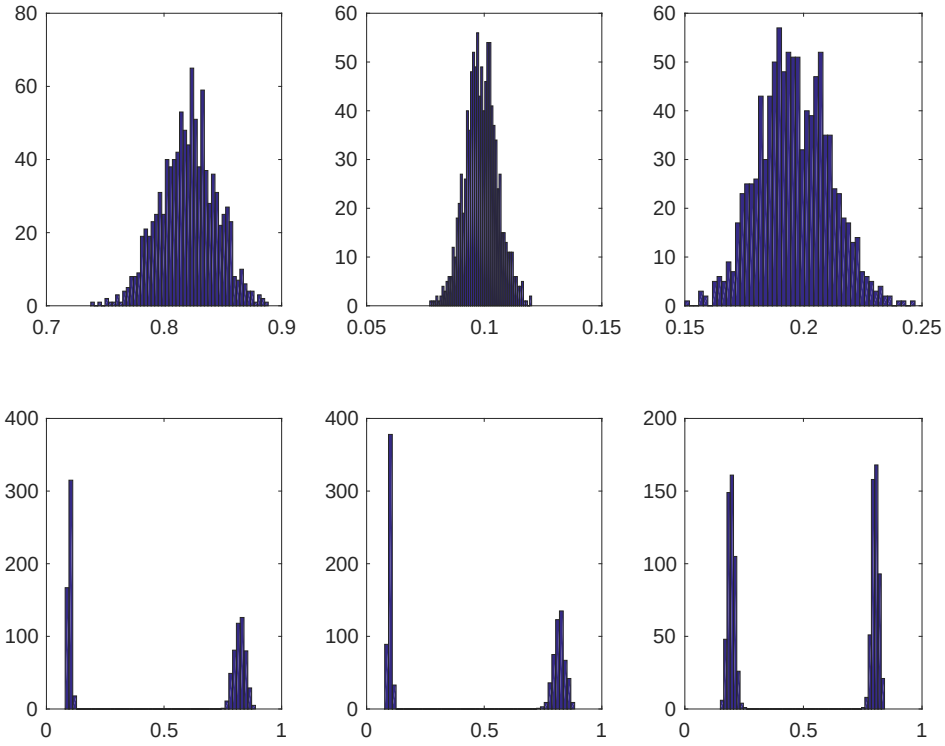

Table 1: Mean and Standard Deviation Estimated using Bootstrap

Sample Size	TRUE	est	bootstrap with order		bootstrap without order	
			mean	std error	mean	std error
500	0.8	0.75	0.75	0.05	0.42	0.33
	0.1	0.10	0.09	0.01	0.42	0.33
	0.2	0.21	0.21	0.03	0.50	0.29
1000	0.8	0.77	0.77	0.03	0.42	0.34
	0.1	0.09	0.09	0.01	0.43	0.34
	0.2	0.22	0.22	0.02	0.50	0.28
1500	0.8	0.82	0.82	0.02	0.46	0.36
	0.1	0.10	0.10	0.01	0.46	0.36
	0.2	0.20	0.20	0.01	0.50	0.30

Figure 2: Timing Patterns for Commercials across Markets (Sweeting (2009))

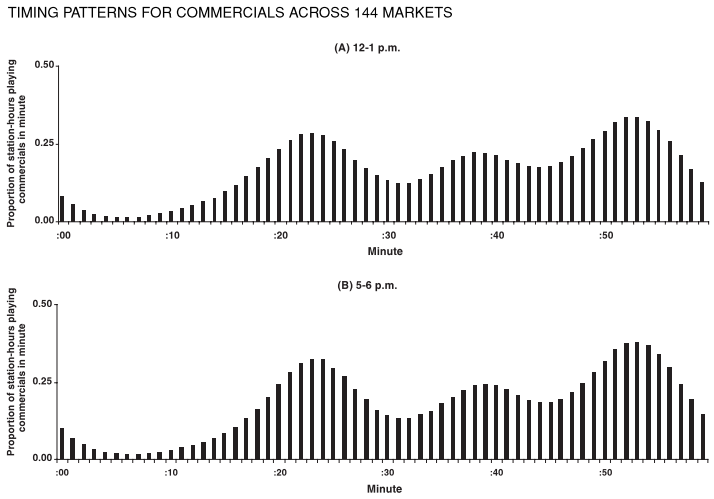

Table 2: Descriptive Statistics

Variable	Obs	Mean	std. Dev	Min	Max
No. Players	92766	5.641	2.054	3	15
Timing	92766	.499	.489	0	1
Day	92766	31.745	17.723	1	59

Figure 3: Timing Patterns for Commercials in Different Markets (Sweeting (2009))

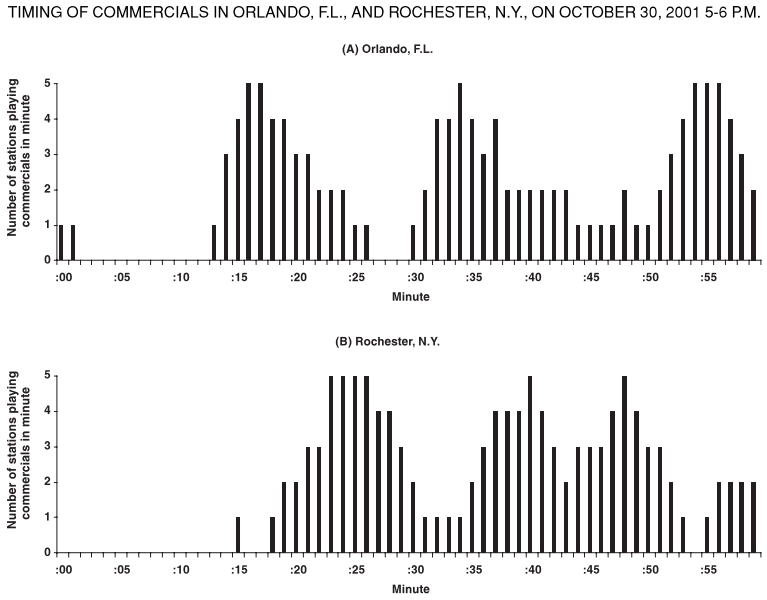

Table 3: Estimation of eq Strategy

	Ordering		No Ordering	
	estimates	std (bootstrap)	estimates	std (bootstrap)
p_1 (eq1)	0.602	0.065	0.602	0.120
p_1 (eq2)	0.420	0.056	0.420	0.123
λ(prob of eq1)	0.451	0.229	0.451	0.232