Social Interactions under Incomplete Information with Heterogeneous Expectations and Multiple Equilibria

Chao Yang*

A Very Preliminary Version

February 12, 2017

Abstract

Socially interacted behaviors under incomplete information can be modeled as equilibrium outcomes of a simultaneous move game. Parameter identification and estimation can be based on the equilibrium expected outcomes. When there are asymmetric information on the exogenous characteristics, the equilibrium expectations are heterogeneous, varying with both individual's traits and the private information used to make predictions. When there are multiple equilibria, the set of equilibrium expectations are a set of functionals defined on the space of private information, which has not been fully characterized in the previous literature. Utilizing the intersection theory of differential topology and functional fixed point theorems, we find that when all exogenous characteristics are public information and only the idiosyncratic shocks are privately known, the set of equilibria is composed of a finite number of vectors and can be computed via the homotopy continuation method. When some exogenous covariates are private information, the equilibrium set is compact in a Banach space and can be approximated by a finite number of equilibria. Thus, it can be numerically computed using basis functions, Gauss-Legendre quadrature, and the homotopy method. Attaching a probability mass function to this approximated set, a computationally-feasible approximation of the complete sample likelihood is derived. Estimation is achievable by either maximizing the likelihood function or using simulated moments. This paper supplements the economic theory on games with multiple equilibria and extends the all-solution method for estimation of discrete choice games to a general framework incorporating discrete and continuous choices, bounded and unbounded outcomes, as well as different types of incomplete information. This method is especially useful for the model with peer effects, where the dimension of the equilibrium conditional expectation functionals can be reduced. We analyze the binary choice models in detail. Monte Carlo experiments show that our estimation method performs well. In addition, large estimation biases can occur if imposing equilibrium uniqueness, either the assumed unique equilibrium is computed through contraction mapping or Newton's method.

JEL-Classification: C31, C35, C57

Keywords: Social Interactions, Multiple Equilibria, Incomplete Information, Intersection Theory, Gaussian Quadrature

*Shanghai University of Finance and Economics, RM 420, School of Economics, No. 777 Guoding Road, Shanghai, China, 200433, *Tel.* +86(21) 6590-2704. *Email*: yang.chao@mail.shufe.edu.cn.

1 Introduction

Estimating models for social interactions with possible multiple equilibria is a challenging issue both theoretically and empirically. The distribution of outcomes is influenced by not only the unknown parameters, but also the underlying equilibrium which is actually played but not observed. Without further specifications, with any given parameter values, the sample likelihood or moment conditions are still indeterminate and cannot be used for estimation. Bajari et al.(2010a) propose a two-step algorithm to estimate a discrete choice game under incomplete information. They first derive nonparametric estimates for players' choice probabilities and then use them to estimate other structural parameters. When there are a large number of repetitions of the same game, under the assumption that the same equilibrium is played for all the repetitions, the individual choice probabilities can be estimated consistently. However, in empirical studies of social interactions, especially interactions among friends, it is frequent to work with cross-section data sets. For data sets with individual information over years, there is also a problem with the evolution of social relations, which makes it unrealistic to assume that the same equilibrium is played repeatedly. This method, nonetheless, can still be used for some special model structures or under some additional assumptions. For example, when individual outcomes are influenced by a global equilibrium aggregate, Bisin et al.(2011) first estimate the equilibrium aggregate and then recover other parameters. Leung(2015) focuses on one particular type of equilibria, where individuals with the same observable characteristics play the same strategy. With a large number of independent groups and/or a large number of agents playing the same strategy, repetitions are derived. However, the method by Bajari et al.(2010a) would be invalid with the presence of unobserved group heterogeneity which cannot be fully explained by observed characteristics. For social interaction models with potentially multiple equilibria, without assuming repetitions of the same equilibrium, this paper proposes using a parametric stochastic rule to specify a probability distribution of equilibrium selection. Although this method requires solving or approximating equilibria, it can be applied to a general model framework and data generating processes, including both discrete and continuous choices, bounded and unbounded outcomes, and incomplete information about idiosyncratic shocks and exogenous characteristics. In a related paper, Bisin et al(2011) choose both the parameter values and equilibria to maximize the sample likelihood function. That method is implicitly built on a particular selection distribution. That is, exactly one equilibrium is chosen with probability one. However, that specification cannot lead to economic implications on equilibrium selection. Moreover, if the likelihood function is of a complicated form, it might be difficult to maximize likelihood when choosing both parameter values and equilibria.

Using a stochastic rule to complete a game with multiple equilibria is not new in the literature of game estimation. Bajari et al.(2010b),(2010c) use this method to identify and estimate discrete choice games under both complete and incomplete information. However, it is not straightforward to make this approach applicable to various empirical studies in social interactions under incomplete information. Socially interacted behaviors can be either discrete or continuous, bounded or unbounded. Moreover, in Bajari et al.(2010b),(2010c), only the idiosyncratic shocks are private information. In their recent research, Yang and Lee(2017) point out that there can also be incomplete information about some exogenous characteristics for social interactions. For example, when analyzing peer effects in class performance, class and individual characteristics such as grades, locations, genders, SAT scores, and IQ scores are often used as exogenous covariates. It would be unrealistic to assume that individual SAT scores and IQ scores are public information. Nonetheless, incorporating various types of behaviors and information structures makes it more difficult to specify the probability distribution of equilibrium selection and compute the likelihood of the complete model.

The difficulty in specification comes from the equilibrium set. The structure of Nash equilibria for a game with complete information has been well understood. So is that for a finite-player finite-action game under incomplete information. However, the existence of a pure strategy equilibrium in a game with private information when the number of possible actions and types are not finite has been an open question for a long time. Following the pioneering work by Milgrom and Weber(1985) and Radner and Rosenthal(1982), Khan and Sun(1995) and Kan and Zhang(2014) provide existence conditions when the set of actions is compact. Although their conditions apply to general abstract private information, in many empirical applications, it is unsatisfactory to restrict the values of outcomes to be bounded. Moreover, as our model is based on a reduced-form Bayesian Nash Equilibrium (BNE), their conditions about the payoff functions may not be directly applied. Therefore, we characterize the equilibrium set and derive conditions for existence and equilibrium properties specific to our framework.

It is shown that, in terms of the distribution of outcomes, a BNE is equivalent to an equilibrium conditional expectation of individual choices, which are functions of the private information used to make predictions. Therefore, the equilibrium conditional expectations are used to represent BNEs. Particularly, if all exogenous characteristics are public information and only the idiosyncratic shocks are privately known, conditional expectation functions reduce to vectors in an Euclidean space satisfying a system of nonlinear equations. By the transversality theorem and the intersection theory in differential topology, it is shown that under certain regularity conditions, there are a finite number of equilibria. Another result is a sufficient condition for equilibrium uniqueness, which is weaker than the condition derived from contraction mapping by Yang and Lee (2017). When some exogenous characteristics are private information and they have a continuum support, the equilibrium conditional expectations are generally functions. They are embedded into a Banach space of functions, which is related to the classical L^p spaces for integrable functions. By the Schauder fixed point theorem, sufficient conditions are derived, which ensure that the set of equilibria is nonempty and compact. As a result, the set of equilibria can be approximated by a finite number of equilibria. With a finite number of elements in the (possibly approximated) equilibrium set, a probability mass function for equilibrium selection can be specified based on a parametric selection rule. That completes the model.

By the strategy of "identification at infinity" and techniques in spatial econometrics, parameters can be identified for the linear model with continuous choices, the model with binary choices, and the Tobit model. Challenges in estimation come from computation. When all exogenous characteristics are public information, solving for the set of equilibria is equivalent to getting all solutions to a system of nonlinear equations. According to Garcia and Zangwill(1981), it is possible to get all solutions via a homotopy continuation method under the regularity and path-finiteness condition, when the system can be extended to complex spaces in an analytic way. It is verified that those conditions hold for a couple of models with normal shocks. There are also discussions about the application of another related homotopy algorithm, used by Borkovski et al.(2010a) and (2010b). There is also a brief discussion about group unobservables, peer effects, and a deterministic selection rule. For models with peer effects, it suffices to focus on conditional expectations about group average outcomes. Hence, an equilibrium can be represented by a vector-valued function with less coordinate functions than that in the general model framework.

Particularly, this paper discusses about two types of binary choice models in detail. In the Type I model, agents take one of two actions, 0 and 1. The utility for choice 0 is normalized to 0 for every agent. When an agent chooses 1, however, her utility depends on the number of agents who are associated with her and also choose 1. Two key features of this model is that ex post, the agents who choose 0 are not affected by others and have no effects on the utilities of agents who choose 1. The entry game is a case in point. The Type II model does not have these two properties. Like the model discussed in Brock and Durlauf(2001), in the Type II model, the utility an agent can get depends on the difference between her choices and those of the agents who she is associated with. With normal idiosyncratic shocks, it is more likely to have multiple equilibria in the Type II model than it is in the Type I model. Additionally, by comparing different estimation methods in the Monte Carlo experiments, it is found that assuming equilibrium uniqueness can bring in biases when the intensity of social interactions is large and there are multiple equilibria in the data generating process.

The paper proceeds as follows. The model framework is introduced in Section 2. Sections 3 and 4 contain a detailed analysis of the set of equilibria, identification, and estimation for two different types of information structures. In Section5, we discuss in detail how to achieve dimension reduction when considering the influence from group peers. Section 6 focuses on equilibrium set characteristics and Monte Carlo experiments of two types of binary choice models. Section 7 concludes. Technical proofs are put in Appendices B through F. In Appendix H, there are extensions on group unobservables and discussions about deterministic rules of equilibrium selection.

2 Models

2.1 A Model Framework

The discussion of multiple equilibria is in the framework for social interactions under a general form of incomplete information analyzed by Yang and Lee (2017). Consider a group of n socially related agents. Their relations are represented by an $n \times n$ matrix W_n. For any $i, j = 1, \cdots, n$,

$W_{n,ij} \geq 0$. $W_{n,ii} = 0$ for all $i = 1, \cdots, n$. For any $i \neq j$, $W_{n,ij} > 0$ if i connects with j; and $W_{n,ij} = 0$ otherwise. Take a game with n players for example. As the payoffs of any two agents are interdependent, $W_{n,ij} = 1$ for any $i \neq j$; and $W_{n,ii} = 0$. This social relation matrix can also be used in the model for peer effects in a social group, where the behavior of an agent is related to those of all the other group members. If W_n is used to represent the spatial relations of geographic regions or local governments, $W_{n,ij}$ is usually negatively correlated with the distance between i and j. In that case, W_n is symmetric. If we use W_n to represent friendship networks, W_n will be symmetric if only mutual friendship are considered. However, if the network is directed, it is possible that i considers j as one of her close friends while j does not regard i as her good friends. Then W_n may be asymmetric.

Let y_i^* denote the latent variable. The observed outcome, or agent's behavior, y_i, depends on y_i^* in the following way:

$$y_i = h_i(y_i^*), \tag{2.1}$$

where $h_i(\cdot) : \Re \to \Re$ is a real-valued function, which can be linear or nonlinear. In the general setting, we allow the form of this function to vary across agents. In applications, we usually have $h_i(\cdot) = h_j(\cdot) = h(\cdot)$. The latent variable for i is related to her expectations about the outcomes for other agents as

$$y_i^* = u(X_i) + \lambda \sum_{j \neq i} W_{i,j} E[y_j | X_{J_i}^p, Z] - \epsilon_i. \tag{2.2}$$

According to (2.2), the value of y_i^* depends on three parts. The first part, $u(X_i)$, represents the direct effects of exogenous covariates, $X_i = (X^g, X_i^c, X_i^p)$. We consider the group features, X^g; some commonly known individual characteristics, X_i^c, and some personal features which may be privately known, X_i^p. The third part is the idiosyncratic shock, represented by ϵ_i. Those shocks are i.i.d. and independent of all the exogenous characteristics and social relations. Their identical distribution is characterized by a pdf function, $f_\epsilon(\cdot)$, with its cdf, $F_\epsilon(\cdot)$. Assume that ϵ_i is known by individual i herself, but not by other agents or econometricians. The second part represents the interaction effects from socially associated agents. There are two features in this formulation. First, y_i^* is affected by agent j only if i connects with j; i.e., $W_{n,ij} \neq 0$. Second, j influences i through i's expectation on j's true outcome. In the model, i's expectations are made on the basis of public information about social relations in W_n, group features, X^g, commonly known individual characteristics, X_j^c's, and her private information about exogenous characteristics, X_j^p's. The information structure can be fully described by specifying the subset of agents whose X_j^p's are known to an agent. Given a finite number of agents, this is achievable using vectors. For each i, we define an $n \times 1$ vector, J_i, such that $J_i(j) = 1$ if i knows X_j^p; and $J_i(j) = 0$ otherwise, for each $1 \leq j \leq n$. As a result, information structure in a group of n agents is represented by an $n^2 \times 1$ vector, $\mathfrak{J} = (J_1^{'}, \cdots, J_n^{'})^{'}$. For every i, we define by $X_{J_i}^p$, the vector formed by X_j^p's, which are known to i, $X_{J_i}^p = (X_j^{p'} : J_i(j) = 1)'$. Suppose that X_j^p is of dimension k_p. Then the dimension of $X_{J_i}^p$ is $N_i = (\sum_{j=1}^n J_i(j))k_p$.[1] To simplify notation, we summarize all the publicly known variables in one vector,

$$Z = (X^g, X_1^{c'}, \cdots, X_n^{c'}, W_{n,11}, \cdots, W_{n,1n}, \cdots, W_{n,n1}, \cdots, W_{n,nn}, J_1^{'}, \cdots, J_n^{'})^{'}. \tag{2.3}$$

Then we sum up i's information set used to make predictions by two random vectors, one about private information, $X_{J_i}^p$, and the other about public information, Z.[2] The parameter, λ, represent the intensity of social interactions. If $\lambda > 0$, the social interaction effect is positive. If $\lambda < 0$, outcomes are negatively related. The case of $\lambda = 0$ represents absence of social interactions.

The model, (2.1) and (2.2), is general enough to include different types of outcomes, such as the continuous outcomes and binary choices in Yang and Lee(2017) and the tobit model i Yang, Qu and Lee(2016). There are also other applications.

1. (Linear Model with Continuous Choices) If $h_i(d) = d$ for all i and $d \in \Re$, we have that

$$y_i = u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} E[y_j | X_{J_i}^p, Z] - \epsilon_i. \tag{2.4}$$

2. (Binary Choice Model I) If $h_i(d) = I(d > 0)$, for all i and $d \in \Re$, where $I(\cdot)$ is the indicator, we have that

$$y_i = I(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} E[y_j | X_{J_i}^p, Z] - \epsilon_i > 0). \tag{2.5}$$

[1]For example, if 1 only knows X_1^p, X_2^p and X_3^p, $J_1(j) = 1$ for $j = 1, 2, 3$; and $J_1(j) = 0$ for $j > 3$, and $X_{J_1}^p = (X_1^{p'}, X_2^{p'}, X_3^{p'})^{'}$. We assume that such an information structure J is common knowledge. However, the realizations of those random variables are private information. In this example, although it is publicly known that agent 1 knows her own features and those of agents 2 and 3, the realizations of X_1^p, X_2^p and X_3^p may be unknown to agent 4.

[2]As it is explained in Yang and Lee (2017), because the idiosyncratic shocks, ϵ_i's are independent of each other and they are also independent of the exogenous covariates and social relations, adding the realization of ϵ_i to her information set does not change i's predictions on others' outcomes.

3. (Binary Choice Model II) Based on the assumption that agents drive utilities from taking actions similar to their friends and/or neighbors, Brock and Durlauf(2001) consider another model for binary choices, $y_i = 1 \ or \ -1$, according to $h_i(d) = 2I(d > 0) - 1$, and

$$y_i = 2I(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} E[y_j | X^p_{J_i}, Z] - \epsilon_i > 0) - 1. \tag{2.6}$$

4. (Tobit Model with Homogeneous Cutoff Points) If all negative outcomes are censored, i.e, $h_i(d) = dI(d \geq 0)$ for all i and $d \in \Re$, we have that

$$y_i = \max \left\{ u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} E[y_j | X^p_{J_i}, Z] - \epsilon_i, 0 \right\}. \tag{2.7}$$

5. (Tobit Model with Heterogeneous Cutoff Points) If $h_i(d) = I(d > v(X^g, X^c_i))$ for all $d \in \Re$ for $i = 1, \cdots, n$, we have that

$$y_i = I(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} E[y_j | X^p_{J_i}, Z] - \epsilon_i > v(X^g, X^c_i))(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} E[y_j | X^p_{J_i}, Z] - \epsilon_i). \tag{2.8}$$

6. (Two-sided Censored Outcomes) If $h_i(d) = dI(c_1 < d < c_2)$ for all i and $d \in \Re$ and some parameters, $c_1 < c_2$, we get

$$y_i = (u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} E[y_j | X^p_{J_i}, Z] - \epsilon_i)(c_1 < u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} E[y_j | X^p_{J_i}, Z] - \epsilon_i < c_2). \tag{2.9}$$

7. (Ordered Multiple Choices) If $h_i(d) = \sum_{k=0}^{K} k(c_k < d < c_{k+1})$ for all i and $d \in \Re$, where $K > 1$ is a fixed integer, $c_0 = -\infty$, $c_1 < \cdots < c_K$, $c_{K+1} = \infty$, we derive the following model:

$$y_i = \sum_{k=0}^{K} k(c_k < u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} E[y_j | X^p_{J_i}, Z] - \epsilon_i < c_{k+1}). \tag{2.10}$$

8. (Investment Decisions with Cobb-Douglas Production Functions) At last, consider interactions in investment among competing firms or contiguous local governments. y_i^* denotes the latent investment. y_i represents the output. Assume that output is influenced by technology A, the labor input L_i, as well as the capital investment. Since investment cannot be negative, the actual investment is $\max\{y_i^*, 0\}$. Assume that A can be estimated from other data sources. L_i's are public information and exogenously given. The potential

investment y_i^* is still determined by (2.2). With a Cobb-Douglas production function,

$$y_i = h_i(y_i^*) = AL_i^{1-\iota}(\max\{y_i^*, 0\})^\iota, \tag{2.11}$$

where $o < \iota < 1$.

Various information structures can be discussed in this framework.

1. (Publicly-known Characteristics) If all exogenous covariates are public information, $J_i = 1_n$, for all $i = 1, \cdots, n$, where 1_n is an $n \times 1$ vector of 1's.

2. (Self-known characteristics) If X_i^p is revealed just to agent i, for any $i = 1, \cdots, n$, $J_i(i) = 1$ and $J_i(j) = 0$ for all $j \neq i$.

3. (Socially-known Characteristics) If for any two agents i and j, i knows X_j^p if and only if i connects to j, for all i, we have that $J_i(i) = 1$, $J_i(j) = I(W_{n,ij} > 0)$ for all $i \neq j$.

It is shown in Yang and Lee (2017) this model can be built on the basis of a simultaneous move game under incomplete information. Under generic conditions, there is a one-to-one correspondence between a Bayes Nash equilibrium and a consistent equilibrium conditional expectation functional. To facilitate analysis in later sections, the definition of the equilibrium expectations are a little bit different from that in Yang and Lee (2017). As a result, we put the details in Appendices A and B.

Assumption 2.1 *$f_\epsilon(\epsilon) > 0$ for all $-\infty < \epsilon < \infty$. That is, the support for all ϵ_i's is $\Re$.*

Assumption 2.2 *For any real number a, $H_i(a) = E_\epsilon[h_i(a - \epsilon)] < \infty$ and is differentiable with respect to a.*

Yang and Lee (2017) find a sufficient condition for the existence of a unique equilibrium.

Proposition 2.1 *Under the condition that $\max_i \sup_a |\frac{dH_i(a)}{da}| < \infty$, if*

$$|\lambda| \, \|W_n\|_\infty \max_i \sup_a |\frac{dH_i(a)}{da}| < 1, \tag{2.12}$$

where $\|W_n\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n W_{n,ij}$, there is one unique equilibrium in the model.

Proof. See the Appendix in Yang and Lee (2017). ■

With a unique equilibrium, the model, (2.1) and (2.2), will be complete. Parameters can estimated using standard likelihood or moment conditions. According to (2.12), in order to ensure a unique equilibrium, the possible range for the intensity of social interactions, λ, depends on the number of links and the derivative of the functions, $H_i(\cdot)$'s. If they are large, the range for λ will be very narrow. If W_n represents the relations among n players in a game, $\|W_n\|_\infty = n-1$, which increases with the group population. For the case of peer effects, $\|W_n\|_\infty = 1$, which is constant. Actually, in the literature of social interactions, it is conventional to row-normalize W_n so that $\|W_n\|_\infty = 1$, which helps alleviate the problem. For example, with row-normalization, it is shown in Yang and Lee (2017) and Yang, Qu and Lee (2016) that $|\lambda| < 1$ is sufficient for the existence of a unique equilibrium for the continuous choice model, the binary choice model and the Tobit model with zero cutoffs. However, for some models, such as the Tobit model with heterogeneous cutoff points and the model of ordered multiple choices listed in this paper, $\max_i \sup_a |dH_i(a)/da|$ can be very large. See also Yang (2014) for a similar case in the sample selection model. Moreover, by imposing $|\lambda| < 1$, possible strong interactions are excluded.

This paper investigates model estimation without imposing Assumption (2.12). It is shown that the method of random equilibrium selection in Bajari et al (2010b) and (2010c) can be extended to this general framework for social interactions. Suppose that an equilibrium ξ^e is selected from the set of equilibria, $\mathfrak{E}(X, W_n)$, according to probability measure,

$$\mu_e(\gamma(\cdot; X, W_n), \alpha), \tag{2.13}$$

where $\gamma(\cdot; X, W_n)$ is a vector-valued criterion function whose coordinates correspond to different criteria, such as Pareto efficiency and maximal entry rate. α is a parameter vector, attaching weights to different criteria. Then the full likelihood for $y = (y_1, \cdots, y_n)'$ can be written as

$$L(y; X, W_n) = \int_{\mathfrak{E}(X,W_n)} \prod_i^n f(y_i|\xi^e, X, W_n) d\mu_e(\gamma(\xi^e; X, W_n), \alpha), \tag{2.14}$$

where we apply the independence of y_i's owing to the independence across the i.i.d. idiosyncratic shocks.

A practical specification depends on characterization of the set of equilibria. Although it is well established that there are a finite number of equilibria for the finite-player discrete choice game analyzed by Bajari et al.(2010a),(2010b), and (2010c), as far as I know, there are no conditions in the literature that can be directly and easily applied to our model framework

(See Khan and Sun (2002) for a survey of related theories). Therefore, we investigate the set of equilibria in this model and derive conditions specific to our framework. As it turns out that the equilibrium sets have different characteristics under different information structures, according to whether exogenous characteristics are private information or not, we discuss those two scenarios separately.

3 Estimation with Publicly Known Characteristics

We begin our discussions with the case that all exogenous covariates are public information and only the idiosyncratic shocks are privately known. In this case, there are just two types of exogenous characteristics, X^g and X_i^c's. Since all expectations are based on public information, every agent other than i will form the same expectations on i's behavior. Therefore, an equilibrium conditional expectation function ξ^e reduces to an $n \times 1$ vector, $\xi^e = (\xi_1^e, \cdots, \xi_n^e)'$. In this case, (A.7) can be rewritten as

$$\xi_i^e = H_i(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} \xi_j^e), \tag{3.1}$$

for all $i = 1, \cdots, n$. Given exogenous covariates, X, and social matrix, W_n, define $S : \Re^n \to \Re^n$ such that for all $i = 1, \cdots .n$,

$$S(\xi; X, W_n)_i = H_i(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} \xi_j^e) - \xi_i. \tag{3.2}$$

$\xi^e \in \Re^n$ is an equilibrium conditional expectation vector if and only if $S(\xi^e; X, W_n) = 0$. Therefore, for a group of n agents with publicly-known exogenous covariates, X, and social relations, W_n, the set of equilibria, $\mathfrak{E}(X, W_n)$, can be describes as the set of solutions to this system of nonlinear equations. That is,

$$\mathfrak{E}(X, W_n) = \{\xi \in \Re^n : S(\xi; X, W_n) = 0\}. \tag{3.3}$$

3.1 Characterization of the Equilibrium Set

In this section, we characterize the set of equilibria, $\mathfrak{E}(X, W_n)$, through inspecting solutions to $S(\xi; X, W_n) = 0$. Solving equations is one of the central topics in mathematics. There are a myriad of well-established results about it. For this particular problem, we employ the oriented intersection theory to analyze solutions to the equation system, (3.2), aiming to derive conditions specific to this model framework. The applications of differential topology for equilibrium characterizations is not new in economic studies. For example, Debreu(1970) and Dierker(1972) used these theories to analyze the set of competitive equilibria in an economy. The key idea of this approach is to deform $S(\cdot; X, W_n)$ and connect it to a function with a simpler form in a "smooth" way, which is called a homotopy. Implications about the set of zeros of $S(\cdot; X, W_n)$ are then derived from the set of zeros of that simpler function. Garcia and Zangwill(1981) provide an intuitive introduction and basic results for this method. In this paper, we utilize some more general results from the textbook by Guillemin and Pollack(1974) and construct homotopies tailored to our model framework. In this way, several properties of the set of equilibria and, especially, a new sufficient condition for the existence of a unique equilibrium can be derived. We present our main results and leave technical proofs to Appendix C.

Define the function $T(\cdot; X, W_n) : \Re^n \to \Re^n$ such that

$$T(\xi; X, W_n)_i = H_i(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij}\xi_j), \tag{3.4}$$

for all $i = 1, \cdots, n$. We can see that ξ is a solution to $S(\xi; X, W_n) = 0$ if and only if ξ is a fixed point of T. The rate that the Euclidean norm of $\|T(\xi)\|_E$ explodes relative to $\|\xi\|_E$ is crucial for the existence of an equilibrium.

Assumption 3.1 *For any group, X, W_n, there is a real number $b < 1$, such that*[3]

$$\lim_{\|\xi\|_E \to \infty} \|T(\xi; X, W_n)\|_E / \|\xi\|_E = b. \tag{3.5}$$

Under Assumption 3.1 and an easy-to-satisfy regularity condition, the conclusions in Proposition 3.1 hold.[4]

Proposition 3.1 *Under Assumptions 2.1, 2.2, C.1, for any regular social group, (X, W_n), if, in addition, Assumption 3.1 holds, there is $r_0(X, W_n) > 0$, such that all equilibria are within the open ball $B(0, r_0(X, W_n)) = \{\xi \in \Re^n : \|\xi\|_E < r_0(X, W_n)\}$ and the number of equilibria is finite.*

Proof. See Appendix C. ■

[3]The subscript, "E", denotes the Euclidean norm.

[4]See Appendix C for detailed discussions about the meaning and implications of the regularity condition. Additionally, it is shown that for models where $u(\cdot)$ is linear in X_i^c with a non-zero slope, if $dH_i(a)/da \neq 0$ for any i and $a \in \Re^1$, almost all groups, (X, W_n), satisfy the regularity condition.

Finiteness comes from regularity. Regularity implies that all equilibria are isolated. As a result, in the closed ball $B[0, r_0(X, W_n)]$, which is compact and contains $B(0, r)$, the number of equilibria is finite. Moreover, we can derive a new condition for the existence of a unique equilibrium.

Proposition 3.2 *Under Assumptions 2.1, 2.2, C.1, and 3.1, for a regular group, the total number of equilibia is odd. In addition, if the Jacobian determinant, det($DS(\xi; X, W_n)$), does not change its sign in the ball $B(0, r_0(X, W_n))$ which contains all equilibria, there is a unique equilibrium.*

Proof. See Appendix C. ■

Recalling that in Yang and Lee(2017), the sufficient condition for a unique equilibrium is (2.12). The following lemma shows that condition (2.12) is stronger than the condition in Proposition 3.2.

Lemma 3.1 *When* (2.12) *holds, sgn(det($DS(\xi; X, W_n)$)) $= (-1)^n$ for all $\xi \in \Re^n$.*

Proof. See Appendix C. ■

The above characterizations of the equilibrium set hinges on Assumption 3.1. It is easy to see that this condition is satisfied when $H_i(\cdot)$ are uniformly bounded. Therefore, the above results apply for models with binary choices, two-sided censored outcomes and ordered multiple choices. It also holds for unbounded (piecewisesly) continuous choices with the magnitude of $H_i(a)$ increases with the magnitude of $a \in \Re^1$, as long as the increasing rate is not very big. A case in point is the example of investment choices. Since the elasticity coefficient in the Cobb-Douglas production function, ι, is between 0 and 1, using Jensen's inequality, we have that

$$
\begin{aligned}
|T(\xi)_i| &= AL_i^{1-\iota} \int (\max \left\{ u(X_i) + \lambda \sum_{j \neq i} W_{n,ij}\xi_j - \epsilon_i, 0 \right\}^{\iota} f_\epsilon(\epsilon_i) d\epsilon_i \\
&\leq AL_i^{1-\iota} [\int \max \left\{ u(X_i) + \lambda \sum_{j \neq i} W_{n,ij}\xi_j - \epsilon_i, 0 \right\} f_\epsilon(\epsilon_i) d\epsilon_i]^{\iota} \\
&= AL_i^{1-\iota} [H_T(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij}\xi_j)]^{\iota},
\end{aligned}
$$

where $H_T(a) = aF_\epsilon(a) - \int_{c<a} cf_\epsilon(c)dc$ is the $H(\cdot)$ function corresponding to the Tobit model.

When $E[\epsilon] < \infty$, $H_T(a)/|a| \leq 1$ when $|a|$ is sufficiently large. For any i,

$$\frac{|T(\xi)_i|}{\|\xi\|_E} \leq AL^{1-\iota}\Big(\frac{H_T(u(X_i) + \lambda\sum_{j\neq i} W_{n,ij}\xi_j)}{|u(X_i) + \lambda\sum_{j\neq i} W_{n,ij}\xi_j|}\Big)^{\iota} \cdot \Big(\frac{|u(X_i) + \lambda\sum_{j\neq i} W_{n,ij}\xi_j|}{\|\xi\|_E}\Big)^{\iota}\frac{\|\xi\|_E^{\iota}}{\|\xi\|_E}.$$

When $|\lambda|\|W_n\|_\infty < \infty$, the right hand side goes to zero as $\|\xi\|_E$ goes to infinity, for $0 < \iota < 1$.[5]Therefore, we derive a condition which can guarantee the existence of a pure strategy BNE with unbounded piecewisely continuous choices and non-compact private shocks. Our condition, Assumption 3.1, therefore, is complementary to sufficient conditions about the existence of an equilibrium for games with private information under general game settings (See Khan and Sun(2002) for a research survey and Khan and Zhang(2014) for a recent improvement). However, in some models, to ensure Assumption 3.1 to hold, we need to impose restrictions on the range of λ, which may be stringent sometimes. The linear model with continuous choices and the Tobit model are such examples. We discuss how to characterize equilibria in those models.

First, for continuous choices, $H_i(a) = a$ for all $i = 1, \cdots, n$ and $a \in \Re$. $S(\cdot; X, W_n) = 0$ is actually a linear equation system:

$$S(\xi; X, W_n) = u + \lambda W_n\xi - \xi = 0, \tag{3.6}$$

where $u = (u(X_1), \cdots, u(X)_n)'$ and $\xi = (\psi_1, \cdots, \psi_n)'$. For a regular group, $DS(\xi; X, W_n) = \lambda W_n - I_n$ is non-singular. Therefore, (3.6) has one and only one solution in that case. That is to say, for a regular group, in the linear model of socially interacted continuous choices, there is one and only one equilibrium.

For the Tobit model with homogeneous cutoffs which are normalized to be equal to 0, $H_i(a) = H(a) = aF_\epsilon(a) - \int_{c<a} cf_\epsilon(c)dc$, is unbounded and strictly increasing. $S(\cdot; X, W_n) = 0$ is a system of nonlinear equations:

$$S(\xi; X, W_n)_i = H(u(X_i) + \lambda\sum_{j\neq i} W_{n,ij}\xi_j) - \xi_i, \tag{3.7}$$

for $i = 1, \cdots, n$. Define $p_i = E[I(y_i > 0)]$, for all $i = 1, \cdots, n$. There is a one-to-one correspondence between p and ξ, for $\xi_i = E[y_i] = H(F_\epsilon^{-1}(p_i))$ holds for all $i = 1, \cdots, n$. Then (3.7) can

[5]When $\|\xi\|_E \to \infty$, $|\xi_j| \to \infty$ for at least one j. If $|\xi_k| < \infty$ for all k with $W_{n,ik} > 0$, $(\frac{H_T(u(X_i)+\lambda\sum_{j\neq i} W_{n,ij}\xi_j)}{\|\xi\|_E})^{\iota}$ goes to 0, as $\|\xi\|_E$ goes to infinity. Otherwise, $(\frac{H_T(u(X_i)+\lambda\sum_{j\neq i} W_{n,ij}\xi_j)}{|u(X_i)+\lambda\sum_{j\neq i} W_{n,ij}\xi_j|})^{\iota}$ is bounded by 1 as $\|\xi\|_E$ is large enough. As $|u(X_i) + \lambda\sum_{j\neq i} W_{n,ij}\xi_j| \leq |u(X_i)| + |\lambda|\|W_n\|_\infty\|\xi\|_E$, when $|\lambda|\|W_n\|_\infty < \infty$, $\Big(\frac{|u(X_i)+\lambda\sum_{j\neq i} W_{n,ij}\xi_j|}{\|\xi\|_E}\Big)^{\iota}$ is also bounded.

be written as

$$F_\epsilon(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} H(F_\epsilon^{-1}(p_j)) - p_i = 0. \tag{3.7$'$}$$

for $i = 1, \cdots, n$. This is a system of equations about p. Because $p \in [0,1]^n$ and $F_\epsilon(\cdot)$ is bounded, we can apply Propositions 3.1 and 3.2, for the Tobit model and find an odd number of equilibria within a ball $B(0, r)$ with $r > 1$. Similarly, if the cutoffs are heterogeneous and are modeled as $v(X^g, X_i^c)$, we have that $\xi_i = E[y_i] = H(F_\epsilon^{-1}(p_i) + v(X^g, X_i^c))$. Then the counterpart to (3.7) is

$$F_\epsilon\Big(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} H(F_\epsilon^{-1}(p_j) + v(X^g, X_i^c)) - v(X^g, X_i^c)\Big) - p_i = 0. \tag{3.7$'$}$$

3.2 Selection Rule and Complete Likelihood

When there are a finite number of equilibria for a group, to attach a probability of equilibrium selection is to attach a probability mass to each point in this set. Following Bajari et al.(2010b) and (2010c), the probability masses are associated with some selection criteria and parameters. To be specific, let $\gamma(\xi^e, X, W_n) = (\gamma_1(\xi^e, X, W_n), \cdots, \gamma_L(\xi^e, X, W_n))'$ be a vector composed of 0's and 1's representing equilibrium properties. For example, $\gamma_1(\xi^e, X, W_n) = 1$ if ξ^e is Pareto dominated by another equilibria; and $\gamma_1(\xi^e, X, W_n) = 0$ otherwise. $\gamma_2(\xi^e, X, W_n) = 1$ if the equilibrium expected utility is maximal in ξ^e. For the model with binary choices, we can make $\gamma_3(\xi^e, X, W_n) = 1$, if the number of agents who choose 1 is bigger in ξ^e than that in any other equilibria. Similarly, for the Tobit model, we can make $\gamma_4(\xi^e, X, W_n) = 1$, if the number of agents whose behaviors are not censored is maximized at ξ^e. Let $\alpha \in \Re^L$ denote the weight. Suppose that given a set of equilibia, $\mathfrak{E}(X, W_n)$, an equilibrium is picked randomly according to the following random rule: $\xi^{e,l}$ is selected if

$$\alpha' \gamma(\xi^{e,l}, X, W_n) + \epsilon_l^s \geq \alpha' \gamma(\xi^{e,l'}, X, W_n) + \epsilon_{l'}^s, \tag{3.8}$$

for any $\xi^{e,l'} \in \mathfrak{E}(X, W_n)$ and $l \neq l'$, where ϵ_l^s's are i.i.d. equilibrium-specific shocks with type-I extreme value distribution. Therefore, the propabiity that $\xi^{e,l}$ is selected is

$$\rho(\xi^{e,l}; \mathfrak{E}(X, W_n), \alpha) = \frac{\exp(\alpha' \gamma(\xi^{e,l}, X, W_n))}{\sum_{\xi \in \mathfrak{E}(X,W_n)} \exp(\alpha' \gamma(\xi^{e,l}, X, W_n))}. \tag{3.9}$$

Then the complete likelihood function for an outcome $y = (y_1, \cdots, y_n)'$ in a social group is as follows,

$$L(y; X, W_n) = \sum_{\xi^e \in \mathfrak{E}(X, W_n)} \rho(\xi^e; \mathfrak{E}(X, W_n), \alpha) \prod_{i=1}^{n} f(y_i | \xi^e), \tag{3.10}$$

which is the basis for identification and estimation.

3.3 Identification

Our analysis about identification is based on the parametric and distributional assumptions below. First, we assume that the payoff function, $u(\cdot)$, is linear in covariates. Since all exogenous characteristics are public information in this section, we only need to consider X^g and X^c.

Assumption 3.2 $u(X_i) = \beta_{0,0} + X^{g\prime}\beta_{0,1} + X_i^{c\prime}\beta_1$ *for all* $i = 1, \cdots, n$.

Assumption 3.3 *The pdf for the i.i.d. idiosyncratic shocks,* ϵ_i*'s, is* $f_\epsilon(\cdot; \sigma)$ *with a known function form and an unknown parameter,* $\sigma > 0$.

Assumption 3.4 $X_i^c \in \Re^L$ *and* ϵ_i*'s have full support.*

For interactions within one group, the group characteristics, X^g, is absorbed by the constant term. Therefore, we suppress $\beta_{0,1} = 0$ now.

Definition 3.1 *Given social relations,* $W = \overline{W}_n$, $(\alpha, \beta, \lambda, \sigma)$ *and* $(\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\lambda}, \widetilde{\sigma})$ *are observationally equivalent at* $\overline{W}_n$*, if they imply the same distribution of observables, namely,*

$$F_{Y,X|\overline{W}_n}(\cdot, \cdot; \alpha, \beta, \lambda, \sigma) = F_{Y,X|\overline{W}_n}(\cdot, \cdot; \widetilde{\alpha}, \widetilde{\beta}, \widetilde{\lambda}, \widetilde{\sigma}). \tag{3.11}$$

$(\alpha, \beta, \lambda, \sigma)$ *is identifiable at* $\overline{W}_n$*, if any* $(\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\lambda}, \widetilde{\sigma}) \neq (\alpha, \beta, \lambda, \sigma)$ *is not observationally equivalent to* $(\alpha, \beta, \lambda, \sigma)$ *at* $\overline{W}_n$.

Different functions $h_i(\cdot)$'s correspond to different types of behaviors. In this paper, we concentrate on identifying model parameters for three of them, linear model for continuous choices, binary choices, and the Tobit model for censored outcomes. Those three models are representative in terms of the relationships between the observed outcomes and the latent variables. For the linear model with continuous choices, because there is a unique equilibrium for almost all groups, we can identify β, σ and λ from socially interacted outcomes without worrying about equilibrium selection.[6] For the binary choice model and the Tobit model, however, given the possibility of equilibrium multiplicity, the distribution of outcomes will depend on the probabilities of equilibrium selection and the distribution of outcomes in an equilibrium. In a nonparametric setting, Aguirregabiria and Mira (2013) identify structural parameters via exclusion restrictions for a game with discrete choices. Corresponding to the way that social interactions are specified, we adopt the strategy of identification at infinity and combine it with techniques in spatial econometrics, when necessary.

Proposition 3.3 *In the linear model of continuous choices, for a group with social relations $\overline{W}_n$, denote $\widetilde{X}_i = (1, \sum_{j\neq i} \overline{W}_{n,ij}, X_i^{c'}, \sum_{j\neq i} \overline{W}_{n,ij} X_j^{c'})'$,*
and $\widehat{X}_i = (1, X_i^{c'}, \sum_{j\neq i} \overline{W}_{n,ij} X_j^{c'})'$.
Under Assumptions 3.2, 3.3, and 3.4, $\beta_{0,0}$, σ and λ can be identified, if

$$\min_{1\leq i\leq n} \min eig E[\widetilde{X}_i \widetilde{X}_i'] > 0; \tag{3.12}$$

or

$$\min_{1\leq i\leq n} \min eig E[\widehat{X}_i \widehat{X}_i'] > 0, \tag{3.12$'$}$$

when $\overline{W}_n$ is row-normalized, where $\min eig(\cdot)$ is the minimal eigenvalue of the corresponding matrices. Since there is only one equilibrium in the linear model for almost all groups, α is not identified in this case.

Proof. See Appendix D. ■

For the payoff parameters, β, λ and σ, in the binary choice model and the Tobit model, we make the following assumption on the model coefficients.

Assumption 3.5 *$\beta_{1,l} \neq 0$ for some $l \in \{1, \cdots, L\}$. Without loss of generality, suppose that $l = 1$.*

Assumption 3.6 *The i.i.d. idiosyncratic shocks, ϵ_i's, are distributed according to a mean-scale family with a pdf, $f_\epsilon(c) = (1/\sigma) f_s((c - \mu_\epsilon)/\sigma)$, where $f_s(\cdot)$ is some known standard distribution pdf, μ_ϵ and σ are the location and scale parameters. Normalize $\mu_\epsilon = 0$ and $\sigma = 1$.*

[6] As a consequence, the parameter for the selection rule, α, is not identifiable in this case.

Lemma 3.2 *Under Assumptions 3.2, 3.4, and 3.5, for any i and $\omega_{-i} \in \{0,1\}^{n-1}$, there is a subset $\mathcal{X}^c(\omega) \subseteq \Re^{nL}$ such that $P(X \in \mathcal{X}^c(\omega)) > 0$,*

and $\lim_{|X^c_{j,1}| \to \infty, j \neq i, X^c \in \mathcal{X}^c(\omega)} P(y_{-i} = \omega | X^c) = 1$.

Proof. See Appendix D. ■

Proposition 3.4 *In the model of binary choices, for a group with social relations $\overline{W}_n$, for any $\omega \in \{0,1\}^{n-1}$, denote $\widetilde{X}(\omega)_i = (1, X_i^{c'}, \sum_{j \neq i} \overline{W}_{n,ij}\omega_j)'$. Under Assumptions 3.2, 3.4, 3.5, and 3.6, β and λ can be identified, if for some non-zero vector $\omega_0 \in \{0,1\}^{n-1}$, there is $D_0 > 0$, such that*

$$\inf_{D \geq D_0} \min eig E[\widetilde{X}(\omega)_i \widetilde{X}'(\omega)_i | X^c \in \mathcal{X}(\omega)^c, |X^c_{j,1}| \geq D, j \neq i] > 0. \tag{3.13}$$

Proof. See Appendix D. ■

As for the Tobit model, we still use the technique of "identification at infinity" for parameters of the payoff function. However, since uncensored outcomes are continuous, we cannot fix any uncensored outcomes as a result of dominant strategy. Nonetheless, when no outcomes are censored, the distribution of interacted outcomes will be similar to that of continuous choices in a linear model, where there is only one equilibrium. Therefore, we can identify parameters for payoffs and shock distributions separately from the parameters for equilibrium selection. Additionally, different from the binary choice model where those parameters are identified up-to-scale, in the Tobit model, σ can be identified based on the following relationship about the average (individual) outcomes and censoring rate found by Yang, Qu and Lee (2016):

$$E[y_i|X^c] = E[I(y_i > 0)|X^c] F_\epsilon^{-1}(E[I(y_i > 0)|X^c]; \sigma) - \int_{c < F_\epsilon^{-1}(E[I(y_i>0)|X^c];\sigma)} c f_\epsilon(c) dc. \tag{3.14}$$

Because (3.14) holds for any individual under every equilibrium, it can be used to identify σ regardless equilibrium multiplicity. To utilize this relationship, we impose the assumption below:

Assumption 3.7 *$f_\epsilon(\cdot;\sigma)$ is differentiable with respect to σ, $\lim_{c \to \infty} c(dF_\epsilon(c;\sigma)/d\sigma) = 0$, and $\frac{\partial F_\epsilon(c;\sigma)/\partial c}{f_\epsilon(c;\sigma)}$ is strictly monotonic with respect to c.*

Lemma 3.3 *Under Assumptions 3.2, 3.4, and 3.5, there is a subset $\mathcal{X}^c_1 \subseteq \Re^{nL}$ such that $P(X \in \mathcal{X}^c_0) > 0$ and $\lim_{|X^c_{j,1}| \to \infty, 1 \leq j \leq n, X^c \in \mathcal{X}^c_1} P(y_j = 1, 1 \leq j \leq n | X^c) = 1$.*

Proof. See Appendix D. ■

Proposition 3.5 *In the Tobit model, for a group with social relations $\overline{W}_n$, denote*
$\widetilde{X}_i = (1, \sum_{j\neq i}\overline{W}_{n,ij}, X_i^{c'}, \sum_{j\neq i}\overline{W}_{n,ij}X_j^{c'})'$ *and* $\widehat{X}_i = (1, X_i^{c'}, \sum_{j\neq i}\overline{W}_{n,ij}X_j^{c'})'$.
Under Assumptions 3.2, 3.4, and 3.5, $\beta_{0,0}$ and λ can be identified, if there is $D_0 > 0$ with

$$\inf_{D\geq D_0}\min_{1\leq i\leq n}\min eigE[\widetilde{X}_i\widetilde{X}_i'|\mathcal{X}_1^c, |X_{j,1}^c| \geq D, 1 \leq j \leq n] > 0; \tag{3.15}$$

or

$$\inf_{D\geq D_0}\min_{1\leq i\leq n}\min eigE[\widehat{X}_i\widehat{X}_i'|\mathcal{X}_1^c, |X_{j,1}^c| \geq D, 1 \leq j \leq n] > 0; \tag{3.15$'$}$$

when $\overline{W}_n$ is row-normalized. If, in addition, Assumption 3.7 is satisfied, σ is identified.

Proof. See Appendix D. ■

After parameters for the payoff function $u(\cdot)$ are identified, we identify α given (β, σ, λ). Suppose that for (β, σ, λ), for group (X, W_n), there are D equilibria. Denote their values under the criterion $\gamma(\cdot; X, W_n)$ by $\gamma(\xi^d; X, \overline{W}_n) \in \Re^q$. Stacking them together, we get an $n \times q$ matrix: $\Gamma(X^c, \overline{W}_n; \beta, \sigma, \lambda) = (\gamma'(\xi^1; X, \overline{W}_n, \cdots, \gamma'(\xi^D; X, \overline{W}_n))'$.

Proposition 3.6 *In the binary choice (Tobit) model, with social relations, $\overline{W}_n$, under the assumptions in Proposition 3.4(Proposition 3.5), α can be identified if*
$E[\Gamma'(X^c, \overline{W}_n; \beta, \sigma, \lambda)\Gamma(X^c, \overline{W}_n; \beta, \sigma, \lambda)|X^c]$ *has full column rank for any X^c.*

Proof. See Appendix D. ■

When there are different groups, $\beta_{0,1}$ can be identified from variations across groups.

3.4 Computation and Estimation

With a parametric selection probability distribution, it is natural to derive parameter estimates by maximizing the complete likelihood function, (3.10). However, that requires computation of all the equilibria, which is a challenging issue both theoretically and numerically. According to Garcia and Zangwill(1981), that ambitious goal is achievable for a class of problems by the homotopy continuation method (simply, the homotopy method). To be specific, in complex spaces, $\overline{S} : \mathfrak{C}^n \to \mathfrak{C}^n$, is analytic. Define a homotopy: $\{\widetilde{c} : S(\widetilde{c}) = 0\}$. we construct the following homotopy:

$$\overline{R}_i(\widetilde{c}, t) = (1 - t)(\widetilde{c}_i^{q_i} - 1) + t\overline{S}_i(\widetilde{c}), \tag{3.16}$$

for $i = 1, \cdots, n$ and $0 \leq t \leq 1$. q_i is a positive integer. For $t = 0$, $\overline{R}_i(\widetilde{c}) = \widetilde{c}_i^{q_i} - 1$, which is a polynomial with q_i solutions. For $t = 1$, $\overline{R}_i(\widetilde{c}) = \overline{S}_i(\widetilde{c})$. Separate the real and imaginary parts of variables and functions. That is, $\widetilde{c} = (\widetilde{c}^R, \widetilde{c}^I)$ and $\overline{R}_i^*(\widetilde{c}, t) = (\overline{R}_i^R(\widetilde{c}, t), \overline{R}_i^I(\widetilde{c}, t))$. The function $\overline{R}^*$ then has $2n$ coordinates. Re-parametrize the system by ω, such that $\widetilde{c} = \widetilde{c}(\omega)$ and $t = t(\omega)$.[7] Then we get $\overline{R}_i^R(\widetilde{c}, t) = 0$ and $\overline{R}_i^I(\widetilde{c}, t) = 0$ for all $i = 1, \cdots, n$. Denote this system as $\overline{R}^* = (\overline{R}_1^R, \overline{R}_1^I, \cdots, \overline{R}_n^R, \overline{R}_n^I)$. Taking derivatives, we get that

$$D\overline{R}^*(y)Dy = \frac{\partial \overline{R}^*}{\partial \widetilde{c}} D\widetilde{c}(\omega) + \frac{\partial \overline{R}}{\partial t} Dt(\omega) = D\overline{R}^*(\widetilde{(c)}, t) \begin{pmatrix} D\widetilde{c}' & Dt(\omega) \end{pmatrix}' = 0, \tag{3.17}$$

where $y = (\widetilde{c}', t)'$. It is shown by Garcia and Zangwill(1981) that the above system can be solved through "basic differential equations" (BDE):

$$y_i'(\omega) = (-1)^i det D\overline{R}^*_{-i}(y), \tag{3.18}$$

for $i = 1, \cdots, 2n+1$, staring at $(\widetilde{c}_0, t_0)$'s with $\overline{R}^*(\widetilde{c}_0, t_0) = 0$. $D\overline{R}^*_{-i}(y)$ is the Jacobian of $D\overline{R}^*(y)$ with the i-th column removed. The above calculation is possible if $D\overline{R}^*(\widetilde{c}, t)$ is of full row rank for all $(\widetilde{c}, t)$ with $\overline{R}(\widetilde{c}, t) = 0$, which is called the "regularity" condition. If in addition, the homotopy is also "path finite"(that it, for any t, any $\widetilde{c}$ satisfying $\overline{R}^*(\widetilde{c}, t)$ never goes to infinity), solving the BDEs can guarantee getting all solutions to $S(\widetilde{c}) = 0$. Similar to the previous discussions about regular groups, the regularity condition is satisfied in general.[8] A sufficient condition for path-finiteness is that the following limit

$$\lim_{\|\widetilde{c}_i\| \to \infty} \frac{\overline{S}(\widetilde{c})_i}{\widetilde{c}^{q_i} - 1} \tag{3.19}$$

is not a pure real negative number, for all i. Especially, the path-finite condition is satisfied when the above limit is equal to zero.

To apply this theory, we need first to extend $S(\xi; X, W_n)$ from the real line to complex spaces and make sure its extension is analytic. This extension, actually, is important, in order to ensure that all solutions to $S(\xi; X, W_n) = 0$ can be solved in this way. That is because for analytic functions in complex spaces, we can make sure $det(\partial \overline{R}^*(\widetilde{c}, t)/\partial \widetilde{c}) \geq 0$.[9] It then follows from (3.18) that $t(\omega)$ is monotonic. That is, on any path, t is monotonic and never turns back. As a result, stemming backwardly from any points in $\left\{\widetilde{c} : \overline{R}^*(\widetilde{c}, 1) = \overline{S}(\widetilde{c}) = 0\right\}$, a path goes

[7] This re-parametrization is important. With multiple equilibria, any fixed t may corresponds to different $\widetilde{c}$'s.

[8] This result, as well as that for regular groups in equilibrium analysis, are a direct result of the Sard's theorem. See the textbook written by Guillemin and Pollack(1974) for details.

[9] This result follows from the Cauchy-Riemann equations.

down directly to a zero for $\overline{R}^*(\widetilde{c},0)$ and never goes up. That ensures that every zero of $\overline{S}(\widetilde{c})$ can be connected through a path to one of the zeros of the homotopic function, $\overline{R}^*(\widetilde{c},0)$. If we restrict to the real line, however, $det(\partial\overline{R}^*(\widetilde{c},t)/\partial\widetilde{c})$ can be either positive or negative or zero. In that case, paths can reverse back and some zeros of $\overline{S}(\widetilde{c})$ may not be reached. One example can be found in Garcia and Zangwill (1981).

This extension is crucial, nonetheless, not straightforward for some economic applications. Bajari et al.(2010c) show that when the power of the polynomials, q_i's, are sufficiently large, the homotopy method can be used to derive all solutions for the multinomial choice model. For the framework in this paper, instead, it is relatively easy to extend $S(\xi; X, W_n)$ for many classes of models. Inspecting the models we listed in the paper, most of the $H_i(\cdot)$'s are defined by integrals. Noticing that $G(\widetilde{a}) = \int_{a_0}^{a} g(\widetilde{c})d\widetilde{c}$ is analytic, when $g(\cdot)$ is analytic in a simply-connected region, we can derive analytic extensions for the models listed in this paper. For example, when the idiosyncratic shocks are normally distributed, its density, $f_\epsilon(\cdot)$, is analytic. For its CDF, we can manipulate as $F_\epsilon(a) = \int_{c<a} f_\epsilon(c)dc = \int_{a_0}^{a} f_\epsilon(c)dc + F_\epsilon(a_0)$, for some real number $a_0 > 0$. The extension that $\widehat{F}_\epsilon(\widetilde{a}) = \int_{a_0}^{\widetilde{a}} f_\epsilon(\widetilde{c})d\widetilde{c} + F_\epsilon(a_0)$ is analytic.[10] Since sums, products, and composites of analytic functions are analytic, we can derive analytic extensions for other models in a similar way. For the model of investment decisions under Cobb-Douglas technology, there is a power function, which can be multivalued in the complex plane. We choose one sheet in that case.

It is not hard to check the sufficient condition for "path-finiteness" in the specific model structure of this paper. As the diagonal entries of the social relation matrix, W_n, are all zeros, for any i, $\overline{S}_i(\widetilde{c})$ does not depend on $\widetilde{c}_i$. Hence, the limit in (3.19) will be zero when $q_i = 2$.

Solving the "basic differential equations" requires computing the Jacobinas, which can increase computation burden. As a result, Garcia and Zangwill(**?**) propose to combine the homotopy continuation method with Newton's method, which can facilicate computation.

In practice, there is a computation algorithm, the homotopy algorithm, with a Fortran code suite, HOMPACK90, provided by Watson et al. (1987) and (1997). That algorithm, is related but not the same as the homotopy continuation method we discussed above. Due to Watson et al. (1987), the homotopy algorithm is a global convergent algorithm and is based on the theory that almost all starting points can lead the a zero of a function, or equivalently, a fixed

[10] The whole space, $\mathfrak{C}^2$ is simply-connected. The integral here then does not depend on the path.

point of a function. Therefore, there is not guarantee that it can lead to all solutions to an equation. However, it is also noted by Borkovsly et al. (2010a) and (2010b), with discretization in computation, the homotopy algorithm can alleviate the problem that the function we are using is non-analytic. In practice, we may compare the performance of both methods.

With G independent groups, the log likelihood of the whole sample can be written as:

$$\begin{aligned} &\log L(Y_1, \cdots, Y_G|\beta, \lambda, \sigma, \alpha) \\ =&\sum_{g=1}^{G} \log(\sum_{\xi^e \in \mathfrak{E}(X_g, W_g)} \frac{\exp(\alpha' \gamma(\xi^e; X_g, W_g))}{\sum_{\widetilde{\xi}^e \in \mathfrak{E}(X_g, W_g)} \exp(\alpha' \gamma(\widetilde{\xi}^e; X_g, W_g))} \prod_{i=1}^{n_g} f(y_{i,g}|\xi^e)) \end{aligned} \tag{3.20}$$

The form of the sample likelihood function follows from two types of independence. First, those groups are independent of each other. Second, because the privately known idiosyncratic shocks are independent, within any group, given an equilibrium, the outcome of a group member is independent of those of other group members. With a large number of independent groups, we can apply conventional large sample theory about maximum likelihood estimation. In practice, as it is computationally intensive to compute all the equilibria, instead of maximizing the sample log likelihood, for any given parameter vector, we can compute the set of equilibrium and simulate the selection result and outcome distribution. Then we can calculate simulated moments and estimate parameters through maximizing the simulated moment conditions.

4 Estimation with Self-Known Characteristics

When some exogenous characteristics, X_i^p's, are not public information, conditional expectations, $\xi_{i,m}^e(\cdot)$ vary with the private information used to make predictions. This paper focuses on the special case that X_i^p is known only to i (and the econometricians). That is, for any i, $J_i(i) = 1$ and $J_i(j) = 0$ for all $j \neq i$. Equilibrium and estimation method for the general information structure can be analyzed in a similar way, notations and calculations will be more complicated though. We make an additional simplifying assumption, as Yang and Lee (2017) do.

Assumption 4.1 *The conditional distribution of* $X^p = (X_1^{p'}, \cdots, X_n^{p'})'$ *is exchangeable, if* X_i^p*'s have the same support* $\mathfrak{X}^p \in \Re^{k_p}$*, for any public information* $Z = z$ *and for any permutation,*

$\varpi : \{1, \cdots, n\} \to \{1, \cdots, n\}$, the conditional distribution of X^p given $Z = z$, $f_p(\cdot)$, satisfies

$$f_p(X_1^p = x_1^p, \cdots, X_n^p = x_n^p) = f_p(X_{\varpi(1)}^p = x_1^p, \cdots, X_{\varpi(n)}^p = x_n^p),$$

for any $x^p = (x_1^{p'}, \cdots, x_n^{p'})'$ in their support.

Under Assumption 4.1, fixing public information, $Z = z$, the conditional distribution, $f_p(X_i^p|X_j^p, Z = z)$ is invariant with i, j as long as $i \neq j$. So we just denote it by $f_p(\widetilde{x}|x)$. For any i, and $k, k' \neq i$, for any x in the support $\mathfrak{X}_p$,

$$\begin{aligned}\xi_i^e(X_k^p = x) =& E[H_i(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij}\xi_j^e(X_i^p))|X_k^p = x, Z = z] \\ =& E[H_i(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij}\xi_j^e(X_i^p))|X_{k'}^p = x, Z = z] \\ =& \xi_i^e(X_{k'}^p = x).\end{aligned}$$

That is to say, the conditional expectation $\xi_i^e(\cdot)$ depends only on the realization of self-known information. Any two agents other than i will make the same prediction on i's behavior whenever their own self-known features are the same. According to Appendix B, the conditional expectation about i's behaviors, $\xi_i^e : \mathfrak{X}^p \to \Re$, is a mapping from the support of self-known covariates to the space of possible outcomes. For conditional expectations about behaviors of all group members, the vector-valued function $\xi^e = (\xi_1^e, \cdots, \xi_n^e)$, satisfies:

$$\xi^e(x_1^p, \cdots, x_n^p)_i = \xi_i^e(x_i^p), \tag{4.1}$$

and

$$\xi_i^e(x) = \int_{\widetilde{x}} H_i(u(X^g, X_i^c, \widetilde{x}) + \lambda \sum_{j \neq i} W_{n,ij}\xi_j^e(\widetilde{x}))f_p(\widetilde{x}|x)d\widetilde{x}, \tag{4.2}$$

for all $i = 1, \cdots, n$ and $x \in \mathfrak{X}^p$.

Particularly, if X_i^p is independent of X_j^p for any $i \neq j$, $f_p(\widetilde{x}|x) = f_p(\widetilde{x})$. That is, information about X_j^p does not help to predict i's actions, given public information $Z = z$. In this case, the conditional expectation function ξ^e reduces to an $n \times 1$ vector with

$$\xi_i^e = \int_{\widetilde{x}} H_i(u(X^g, X_i^c, \widetilde{x}) + \lambda \sum_{j \neq i} W_{n,ij}\xi_j^e)f_p(\widetilde{x})d\widetilde{x}, \tag{4.3}$$

which is a system of nonlinear equations and can be analyzed in a way similar to that used when all exogenous characteristics are public information.

Hence, in the subsequent sub-sections, discussions are focused on the case that X_i^p and X_j^p are correlated for $i \neq j$. Since the domain of each ξ_i^e is the support of X_j^p's, whether X_j^p's are discrete or continuous random vectors will have different implications on the conditional expectation functions. We discuss equilibrium, estimation and identification with self-known characteristics separately for those two cases.

4.1 Discrete Private Characteristics

Suppose that X_i^p's are discretely distributed and have a finite number of mass points. Suppose that their common support is $\mathfrak{X}^p = \left\{x^1, \cdots, x^K\right\}$ for some $K < \infty$. The transitional probabilities are $P_{k,k'} = Pr(X_i^p = x^{k'} | X_j^p = x^k, Z = z)$. Then ξ^e can be represented by a vector,

$$\xi^e = (\xi_{1,1}^e, \cdots, \xi_{1,K}^e, \cdots, \xi_{n,1}^e, \cdots, \xi_{n,K}^e)', \tag{4.4}$$

where $\xi_{i,k}^e = \xi_i^e(x^k)$. The consistency condition (4.2) reduces to

$$\xi_{i,k}^e = \sum_{k'}^{K} P_{k,k'} H_i(u(X^g, X_i^c, x^{k'}) + \lambda \sum_{j \neq i} W_{n,ij} \xi_{j,k'}^e), \tag{4.5}$$

for $i = 1, \cdots, n$ and $k = 1, \cdots, K$. This is a finite dimension nonlinear system of equations, similar to the equilibrium condition without private information. Thus, the techniques in the previous section can be used to analyze equilibria and complete the model. Identification of model parameters can be proved analogously. Although the support of privately known characteristics, X_i^p's, is bounded, if the commonly known individual features, X_i^c's, have a full support, the method of "identification at infinity" can still be used.

4.2 Continuous Private Characteristics

4.2.1 Equilibrium Set

If X_i^p's are continuous random variables, ξ^e is a function defined on a continuum set of points satisfying the functional equations (4.1) and (4.2). According to Appendix B, we can view ξ^e as a point in a Banach space, $(\Xi(W_n, \mathcal{J}), \|\cdot\|)$. Define an operator, $T : (\Xi(W_n, \mathcal{J}), \|\cdot\|) \to (\Xi(W_n, \mathcal{J}), \|\cdot\|)$ as

$$T(\xi)(x_1^p, \cdots, x_n^p)_i = \int_{\widetilde{x}} H_i(u(X^g, X_i^c, \widetilde{x}) + \lambda \sum_{j \neq i} W_{n,ij} \xi_j^e(\widetilde{x})) f_p(\widetilde{x} | x_i^p) d\widetilde{x}, \tag{4.6}$$

for all i and $(x_1^p, \cdots, x_n^p) \in \mathfrak{X}_p^n$. Then an equilibrium corresponds to a fixed point of this operator. We can prove the existence of an equilibrium by the Schauder fixed point theorem.[11] To apply this theorem, we impose the following two assumptions.

Assumption 4.2 *For all $i = 1, \cdots, n$, $H_i(\cdot)$ is differentiable. Additionally,*

$$\max_{1 \le i \le n} \sup_{c \in \Re} |\frac{dH_i(c)}{dc}| \|W_n\|_\infty < \infty. \tag{4.7}$$

Assumption 4.3 *There is $0 \le b < 1$ such that $\max_{\|\xi\| \to \infty} \|T(\xi)\|/\|\xi\| = b$.*

It is proved in Lemma E.4 that under Assumption 4.2, T is continuous.(Actually, T is a Lipschitz function when this condition holds.) In addition, it is shown by Lemma E.5 that with Assumption 4.3, there is $r_0 > 0$, such that for all ξ in the closed ball, $B[0, r_0] = \{\xi : \|\xi\| \le r_0\}$, $\|T(\xi)\| \le r_0$. That is to say, the images of all points in $B[0, r_0]$ is still in this ball. Moreover, if there is any equilibrium, it must be contained in $B[0, r_0]$. The ball, $B[0, r_0]$, is nonempty, closed, and convex. To apply the Schauder fixed point theorem, it suffices to show that $T(B[0, r_0])$ is contained in a compact subset of $B[0, r_0]$. However, that is not trivial, because the ball $B[0, r_0]$ is not compact in the function space, $(\Xi(W_n, \mathcal{J}), \|\cdot\|)$. To capture a compact set in this space, we begin with the relatively compact sets, for the closure of a relatively compact set is compact. As we have mentioned, $\xi \in (\Xi(W_n, \mathcal{J}), \|\cdot\|)$ if any only if each of its coordinate functions, ξ_i, belongs to the Lebesgue space, $L^1(\mathfrak{X}^p, \mathfrak{B}_X, \mu_p; \Re^1)$. Utilizing the characterization of relatively compact subsets in Lebesgue spaces by Dunford and Schwartz(1958), we derive necessary and sufficient conditions for relative compactness in $(\Xi(W_n, \mathcal{J}), \|\cdot\|)$. (See Proposition E.3 for the results of a general form of private information about X_i^p's.) On the basis of these discussions, a theorem about the set of equilibria is derived.

Proposition 4.1 *Under Assumptions 4.2 and 4.3, if in addition,*

$$\max_{1 \le i \le n} \int_{\mathfrak{X}^p} |T(\xi)_i(x + \widetilde{x}) f_p(x + \widetilde{x}) - T(\xi)_i(x) f_p(x)| dx \to 0, \tag{4.8}$$

as $\widetilde{x} \to 0$, uniformly for any $\xi \in B[0, r_0]$; and

$$\max_{1 \le i \le n} \int_{\mathfrak{X}^p - C_r} |T(\xi)_i(x)| f_p(x) dx \to 0, \tag{4.9}$$

[11] The theorem is cited as Proposition E.4 in Appendix E. See Bonsall (1962) for details. A brief introduction can be found at http://en.wikipedia.org/wiki/Schauder_fixed_point_theorem.

as $r \to \infty$, uniformly for all $\xi \in B[0, r_0]$, the set of equilibria, $\mathfrak{E}(X, W_n)$, is a nonempty and compact subset of $(\Xi(W_n, \mathcal{J}), \|\cdot\|)$ *and is contiained in the closed ball $B[0, r_0]$. In particular,* (4.8) *and* (4.9) *are satisfied, if*

1. *$H_i(\cdot)$'s are uniformly bounded, i.e.,* $\max_{1\le i\le n}\sup_{a\in\Re^1}|H_i(a)| \le \overline{B}'$ *for some $\overline{B}'$;*

2. *$E[X_i^p|Z = z] < \infty$, for all i; and*

3. *For some $\delta_0 > 0$, for each i, there is an function $g_i(x, \widehat{x})$ such that $\int_{\mathfrak{X}^p_{i,m}}\int_{\mathfrak{X}^p_{J_i}} |g_i(x, \widehat{x})| dx d\widehat{x} < \infty$, $f_{p,i}(x + \widetilde{x}, \widehat{x}) \le g_i(x, \widehat{x})$, a.e., for any $\widetilde{x}$ in the cube C_{δ_0}, where $f_{p,i}(\cdot, \cdot)$ is the joint density of X_i^p and X_j^p, $i \ne j$, conditional on public information $Z = z$.*[12]

This existence theorem requires some conditions about the behaviors, $H_i(\cdot)$'s, and the joint distribution of X_i^p's conditional on public information $Z = z$. Conditions (4.8) and (4.9) are about uniform convergence, which may be hard to verify. However, when $H_i(\cdot)$'s are uniformly bounded, we just need to verify some distribution conditions. This type of scenarios include models with bounded behaviors, such as the models for binary choices, ordered multinomial choices and two-sided censored choices. In that case, we can see that if conditional on public information $Z = z$, X_i^p's have a continuous joint distribution on a bounded support, the sufficient conditions about distribution are satisfied. When the support is unbounded, for some distributions, the joint density of X_i^p's can still be dominated by a integrable function. One example is the normal distribution. See Lemma E.6 in Appendix E. Therefore, when outcomes and/or behaviors are uniformly bounded and the joint distribution of X_i^p and X_j^p $(i \ne j)$ is normal conditional on public information $Z = z$, there is at least one equilibrium and the set of equilibria is a compact set contained in a closed ball $B[0, r_0]$.

The compactness of $\mathfrak{E}(X, W_n)$ is important. Given this result, for any small positive number, $\eta > 0$, there is a finite number of points in $\mathfrak{E}(X, W_n)$, $\{\xi^{e,1}, \cdots, \xi^{e,K}\}$, such that for any equilibrium ξ^e, we can pick one of those points, say $\xi^{e,k}$, with $\|\xi^e - \xi^{e,k}\| < \eta$.[13] This way, the finite set $\{\xi^{e,1}, \cdots, \xi^{e,K}\}$ can be viewed as an approximation of all equilibria with precision η.

[12] A cube, C_r, is the set in $\Re^{k_p}$ such that for any $\widetilde{x} \in C_r$, all its coordinates are within $[-r, r]$.

[13] See Dunford and Schwartz (1958) and Folland(1999) for a brief introduction on compactness and relative compactness.

This is the basis on which we specify the distribution of equilibrium selection in the current structure of incomplete information.

Conditions (4.8) and (4.9) are used to ensure that we can apply the Schauder fixed point theorem to the whole ball $B[0, r_0]$. If we restrict to equilibria with some special properties so that they are contained in a compact subset of $B[0, r_0]$, we may apply the Schauder fixed point theorem in this compact subset. Then (4.8) and (4.9) will be satisfied, some other possible equilibria will be excluded though. For the model of investment decisions with Cobb-Douglas production function, Assumption 4.3 is satisfied. To see this, via the Jensen's inequality and the Hölder's inequality,

$$\begin{aligned}
&\frac{\int_{\mathfrak{X}^p} |T(\xi)_i(x)| f_p(x)dx}{\|\xi\|} \\
=&AL_i^{1-\iota} \frac{\int_{\mathfrak{X}^p} \int \int \big(\max\Big\{ u(X^g, X_i^c, y) + \lambda \sum_{j\neq i} W_{n,ij}\xi_j(y) - \epsilon_i, 0 \Big\} \big)^\iota f_\epsilon(\epsilon_i) d\epsilon_i f_p(y|x) dy f_p(x) dx|}{\|\xi\|} \\
\leq& AL_i^{1-\iota} \frac{\int_{\mathfrak{X}^p} \int [H_T(u(X^g, X_i^c, y) + \lambda \sum_{j\neq i} W_{n,ij}\xi_j(y))]^\iota f_p(y|x) dy f_p(x) dx}{\|\xi\|} \\
\leq& AL_i^{1-\iota} \frac{[\int H_T(u(X^g, X_i^c, y) + \lambda \sum_{j\neq i} W_{n,ij}\xi_j(y)) f_p(x,y) dxdy]^\iota}{\|\xi\|} \\
=&AL_i^{1-\iota} \frac{\|\xi\|^\iota}{\|\xi\|} \Big[\int \frac{H_T(u(X^g, X_i^c, y) + \lambda \sum_{j\neq i} W_{n,ij}\xi_j(y))}{|u(X^g, X_i^c, y) + \lambda \sum_{j\neq i} W_{n,ij}\xi_j(y)|} \\
&\cdot \frac{|u(X^g, X_i^c, y) + \lambda \sum_{j\neq i} W_{n,ij}\xi_j(y)|}{\|\xi\|} f_p(x,y) dxdy \Big]^\iota \\
\leq& AL_i^{1-\iota} \frac{\|\xi\|^\iota}{\|\xi\|} \Big[\int \big(\frac{H_T(u(X^g, X_i^c, y) + \lambda \sum_{j\neq i} W_{n,ij}\xi_j(y))}{|u(X^g, X_i^c, y) + \lambda \sum_{j\neq i} W_{n,ij}\xi_j(y)|} \big)^p f_p(x,y) dxdy \Big]^{\frac{\iota}{p}} \\
&\cdot \Big[\int \big(\frac{|u(X^g, X_i^c, y) + \lambda \sum_{j\neq i} W_{n,ij}\xi_j(y)|}{\|\xi\|} \big)^q f_p(x,y) dxdy \Big]^{\frac{\iota}{q}},
\end{aligned} \tag{4.10}$$

for some $p, q \geq 1$ with $1/p + 1/q = 1$. Similar to our discussion in the previous section, if $|\lambda| \|W_n\|_\infty < \infty$, $\frac{\int_{\mathfrak{X}^p} |T(\xi)_i(x)| f_p(x)dx}{\|\xi\|}$ goes to zero as $\|\xi\|$ goes to infinity. Then $\lim_{\|\xi\|_E \to \infty} \frac{\|T(\xi)\|}{\|\xi\|} = 0$. However, other conditions may not hold in this case. Instead, if we focus on the case that X_i^p's have a compact support and all equilibrium conditional expectations, ξ^e's, are continuous, we may derive a similar characterization of the set of equilibria.[14]

Nonetheless, for some models with unbounded choices, such as the linear model of continuous choices and the Tobit model, in order to satisfy Assumption 4.3, we might have to impose strong conditions on λ and $\|W_n\|_\infty$. To avoid that, we employ other techniques to analyze those models.

[14] In that case, we may use the sup-norm for each coordinate function, the ξ_i's, instead of the $\|\cdot\|_1$ norm.

First consider the linear model with continuous choices. In this case, the equilibrium condition for conditional expectations is:

$$\xi_i^e(x) = \int_{\widetilde{x}} u(X^g, X_i^c, \widetilde{x}) f_p(\widetilde{x}|x) d\widetilde{x} + \int_{\widetilde{x}} \lambda \sum_{j\neq i} W_{n,ij} \xi_j^e(\widetilde{x})) f_p(\widetilde{x}|x) d\widetilde{x}. \tag{4.11}$$

Reorganizing the above equation, we get that

$$\xi_i^e(x) + \int_{\widetilde{x}} (-\lambda) \sum_{j\neq i} W_{n,ij} \xi_j^e(\widetilde{x})) f_p(\widetilde{x}|x) d\widetilde{x} = \int_{\widetilde{x}} u(X^g, X_i^c, \widetilde{x}) f_p(\widetilde{x}|x) d\widetilde{x}. \tag{4.12}$$

This is an n-dimension Fredholm integration alternative, whose solutions are investigated by general Fredholm theory.[15] Especially, when $u(X^g, X_i^c, X_i^p) = v(X^g, X_i^c) + X_i^{p'}\beta$ and $E[X_j^p|X_i^p] = \mu + CX_i^p$, by Yang and Lee (2017), when $I_{nk_p} - \lambda(W_n \otimes C')$ and $I_n - \lambda W_n$ are both invertible, there is one and only one linear equilibrium conditional expectation, ξ^e.

For the Tobit model with heterogeneous cutoffs, define
$p_i(x) = F_\epsilon(u(X^g, X_i^c, x) + \lambda \sum_{j\neq i} W_{n,ij}\xi_j^e(x) - v(X^g, X_i^c))$,
for all i and $x \in \mathfrak{X}^p$. Then we have that $\xi_i(x) = \int H_i(F_\epsilon^{-1}(p_i(\widetilde{x})) + v(X^g, X_i^c)) f_\epsilon(\widetilde{x}|x) d\widetilde{x}$. The equilibrium condition can be rewritten as

$$p_i(x) = F_\epsilon(u(X^g, X_i^c, x) + \lambda \sum_{j\neq i} W_{n,ij} \int H_j(F_\epsilon^{-1}(p_j(\widetilde{y})) + v(X^g, X_i^c)) f_\epsilon(\widetilde{y}|x) d\widetilde{y} - v(X^g, X_i^c)).$$

As $F_\epsilon(\cdot)$ is bounded, we can apply Proposition 4.1 to analyze equilibrium set. If $v(X^g, X_i^c) = 0$ for all i, that is the case of the Tobit model with homogeneous cutoffs normalized to 0.

4.2.2 Equilibrium Set Approximation

Although we can approximate the whole equilibrium set by a finite number of equilibria for any level of precision, as functions defined on a continuum, it is not possible to derive the exact values of those functions at every point in their domains. In order to apply the stochastic selection rule, we approximate such a function. Four possible approximation approached are discussed here.

The first method uses the simple functions. In Appendix E, we show that $\xi = (\xi_1, \cdots, \xi_n) \in (\Xi(W_n, \mathcal{J}), \|\cdot\|)$ if and only if each of its coordinate function, ξ_i, is an element of the Lebesgue

[15] See Ruston (1986) for systematic discussions. Lax(2002) provides with a succinct discussions. Define an operator, T_F such that for any ξ, $T_F(\xi)(x_1, \cdots, x_n)_l = T_F(\xi)_l(x_l) = \int_{\widetilde{x}} (-\lambda) \sum_{j\neq i} W_{n,ij}\xi_j^e(\widetilde{x}) f_p(\widetilde{x}|x_l) d\widetilde{x}$,
for all l and $x_l \in \mathfrak{X}^p$. If $\mathfrak{X}^p$ is compact in $\Re^{k_p}$, T_F is a compact operator from $(\Xi(W_n, \mathcal{J}), \|\cdot\|)$ to a space of continuous functions on $\mathfrak{X}^p$. Then this integration has a solution if
$\widetilde{u}(x_1, \cdots, x_n) = (\int_{\widetilde{x}} u(X^g, X_i^c, \widetilde{x}) f_p(\widetilde{x}|x_1) d\widetilde{x}, \cdots, \int_{\widetilde{x}} u(X^g, X_i^c, \widetilde{x}) f_p(\widetilde{x}|x_n) d\widetilde{x})'$ is orthogonal to the null space of the transpose operator, T_F'.

space, $L^1(\mathfrak{X}^p, \mathfrak{B}_p, \mu_p; \Re^1)$. In such a function space, there is a special set of functions called the "simple functions". A function ξ from $\mathfrak{X}^p$ to $\Re^1$ is μ_p simple if ξ has only a finite set of values, $\{\nu_1, \cdots, \nu_K\}$; and for any ν_k, $\xi^{-1}(\nu_k)$ is an element of the σ-algebra, $\mathfrak{B}_p$. By Dunford and Schwartz (1958), the set of μ_p-measurable simple functions is dense in $L^1(\mathfrak{X}^p, \mathfrak{B}_p, \mu_p; \Re^1)$. That is, for any given level of precision, we can always find a simple function to approximate a function in $L^1(\mathfrak{X}^p, \mathfrak{B}_p, \mu_p; \Re^1)$. For any integer K, choose a finite partition of $\mathfrak{X}^p$, $U^{K,1}, \cdots, U^{K,K}$, where each $U^{K,k}$ is in $\mathfrak{B}_p$. For any $i = 1, \cdots, n$, define a simple function as $\xi_i^K(\widetilde{x}) = \sum_{k=1}^K \kappa_{i,K,k} I(\widetilde{x} \in U^{K,k})$. Then the equilibrium condition (4.2) can be approximated as

$$\sum_{k=1}^{K} \kappa_{i,K,k} I(\widetilde{x} \in U^{K,k}) \approx \int H_i(u(X^g, X_i^c, \widetilde{y}) + \lambda \sum_{j \neq i} W_{n,ij} \sum_{k=1}^{K} \kappa_{j,K,k} I(\widetilde{y} \in U^{K,k})) f_p(\widetilde{y} | \widetilde{x} \in U^{K,k}) d\widetilde{y}, \tag{4.13}$$

for all $i = 1, \cdots, n$ and $k = 1, \cdots, K$. Therefore,

$$\begin{aligned} \kappa_{i,K,k} &= \int H_i(u(X^g, X_i^c, \widetilde{y}) + \lambda \sum_{j \neq i} W_{n,ij} \sum_{k=1}^{K} \kappa_{j,K,k} I(\widetilde{y} \in U^{K,k})) f_p(\widetilde{y} | \widetilde{x} \in U^{K,k}) d\widetilde{y} \\ &= \sum_{k=1}^{K} P(U^{K,k'} | U^{K,k}) E[H_i(u(X^g, X_i^c, \widetilde{y}) + \lambda \sum_{j \neq i} W_{n,ij} \kappa_{j,K,k'}) | \widetilde{x} \in U^{K,k}, \widetilde{y} \in U^{K,k'}], \end{aligned} \tag{4.14}$$

where $P(U^{K,k'} | U^{K,k}) = P(X_j^p \in U^{K,k'} | X_i^p \in U^{K,k})$. As i runs over $1, \cdots, n$ and k runs over $1, \cdots, K$, there are nK equations for nK unknowns. This is similar to the previous analysis for the case of discretely distributed X_i^p's. What is different is that the conditional distribution, $f_p(\widetilde{y} | x_{K,k}^p)$, is not discrete. When all $H_i(\cdot)$'s and $f_p(\cdot)$ can be extended to the complex space in an analytic way, as the right-hand-side of (4.14) is a weighted sum of integrations over $H_i(\cdot)$, the above system can be extended to complex spaces. Multiple solutions for $\kappa_{j,K,k}$'s can then be computed by the homotopy method. The corresponding simple functions are then used as an approximation of the equilibrium set.

To approximate equilibrium conditional expectation functions by simple functions is, in essence, to discretize the domain, $\mathfrak{X}^p$. Rust(1987) uses this method to solve for the optimal engine replacement scheme in an optimization programming problem. Compared with his work, instead of a unique optimal scheme in Rust(1987), it is possible to get multiple solutions to (4.14), which makes the model in this paper more computationally intensive. Precision depends on the choice of partitions. Generally speaking, the finer the partition (increasing K), the more precise the approximation. However, it remains a problem how to choose the cutoffs for a partition.

Inspecting (4.2), the value of an equilibrium conditional expectation is determined by an integration, which can be approximated by the quadrature method(See Judd (1998) and Lee(2001) for details.). Employing this approximation of integration, we get the second method to approximate equilibria. Take the Gauss-Legendre quadrature as an example. Consider the simple case that the dimension for X_i^p's is $k_p = 1$ and $\mathfrak{X}^p$ is an interval, $[a, b]$. We have that

$$\begin{aligned}
\xi_i^e(x) &= \int_a^b H_i(u(x^g, x_i^c, \widetilde{x}) + \lambda \sum_{j \neq i} W_{n,ij} \xi_j^e(\widetilde{x})) f_p(\widetilde{x}|x) d\widetilde{x} \\
&\approx \sum_{k=1}^{K} \omega_k H_i(u(x^g, x_i^c, \frac{(\upsilon_k + 1)(b-a)}{2} + a) + \lambda \sum_{j \neq i} W_{n,ij} \xi_j^e(\frac{(\upsilon_k + 1)(b-a)}{2} + a)) \\
&\quad \cdot f_p(\frac{(\upsilon_k + 1)(b-a)}{2} + a|x) \frac{b-a}{2},
\end{aligned} \tag{4.15}$$

where υ_k, for $k = 1, \cdots, K$, are abscissae and ω_k are the weights. They are fixed. If we can get the values of expectation function on a finite number of points, $x_k^p = \frac{(\upsilon_k+1)(b-a)}{2} + a$, for $k = 1, \cdots, K$, we can approximate the value of expectation function at any point in the support of X_i^p. To be specific, when we take $x = x_k^p$ for $k = 1, \cdots, K$, we will get nK equations for nK unknowns, $\left\{\xi_i^e(x_k^p)\right\}_{i,k}$'s. Using the homotopy method, we get a finite number of solutions, $\left\{\xi_i^{e,d}(x_k^p)\right\}_{i,k}$, for some $d = 1, \cdots, D$. For each d, plugging $\left\{\xi_i^{e,d}(x_k^p)\right\}_{i,k}$ back into (4.15), we derive an approximation of an equilibrium conditional expectation function. If X_i^p has a full support, we need to change variables so that integral is over a bounded interval. This quadrature method is used by Yang and Lee (2017) and Yang, Qu and Lee (2016) to derive approximations to equilibrium conditional expectation functions when there are private information in exogenous characteristics. The difference lies in the way to solve the system of equations (4.15). They search for the unique solution under condition (2.12). Instead, in this paper, the homotopy method is used to derive all the solutions, $\left\{\xi_i^{e,d}(x_k^p)\right\}_{i,k}$.

When X_i^p's are of multiple dimensions, the tensor product may be used. Details are introduced by Judd (1998). However, when the dimension of X_i^p's are high, computation can be intensive. Therefore, for high-dimension privately known characteristics, the stochastic integration may be used for approximation instead. To be specific, let $g(\cdot)$ be a density with its support containing the support of X_i^p such that $f_p(x^p|x)/g(x^p)$ is well defined. Then we can generate K random draws, say, x_k^p, from density $h(\cdot)$. The stochastic approximation will be

$$\xi_i^e(x) \approx \frac{1}{K}\sum_{k=1}^{K} H_i(u(x^g, x_i^c, x_k^p) + \lambda \sum_{j\neq i} W_{n,ij}\xi_j^e(x_k^p))\frac{f_p(x_k^p|x)}{g(x_k^p)}. \tag{4.16}$$

The fourth method is a combination of the quadrature method and basis function approximation. That is, an equilibrium function is approximated by a linear combination of a finite number of bases. In our model, $\xi = (\xi_1, \cdots, \xi_n)$ is an element of $(\Xi(W_n, \mathcal{J}), \|\cdot\|)$ if and only if each of its coordinate function, ξ_i is an element of the Lebesgue space, $L^1(\mathfrak{X}^p, \mathfrak{B}_p, \mu_p; \Re^1)$. We can use basis of this Banach space to approximate an equilibrium. For example, the dimension of X_i^p's is $k_p = 1$ and $\mathfrak{X}^p = [a, b]$. We first transform ξ into a function on $[0, 1]$ by changing variables.

1. When $-\infty < a < b < +\infty$, by setting $x = a + (b - a)\widetilde{x}$,
   $\int_a^b \xi_i(x) f_p(x) dx = \int_0^1 b\xi(a + (b - a)\widetilde{x}) f_p(a + (b - a)\widetilde{x}) d\widetilde{t}$.
   Define $\widetilde{\xi}_i(\widetilde{x}) = b\xi(a + (b - a)\widetilde{x}) f_p(a + (b - a)\widetilde{x})$.

2. When $a = -\infty$ and $b < +\infty$, by setting $x = \log(b\widetilde{x})$,
   $\int_a^b \xi_i(x) f_p(x) dx = \int_0^1 \xi(\log(b\widetilde{x})) f_p(\log(b\widetilde{x})) \frac{1}{\widetilde{x}} d\widetilde{x}$.
   Define $\widetilde{\xi}_i(\widetilde{x}) = \xi(\log(b\widetilde{x})) f_p(\log(b\widetilde{x})) \frac{1}{\widetilde{x}}$.

3. When $a > -\infty$ and $b = +\infty$, by setting $x = \log(a/(1 - \widetilde{x}))$,
   $\int_a^b \xi_i(x) f_p(x) dx = \int_0^1 \xi(\log(a/(1 - \widetilde{x})) f_p(\log(a/(1 - \widetilde{x}))) \frac{1}{1-\widetilde{x}} d\widetilde{x}$.
   Define $\widetilde{\xi}_i(\widetilde{x}) = \xi(\log(a/(1 - \widetilde{x})) f_p(\log(a/(1 - \widetilde{x}))) \frac{1}{1-\widetilde{x}}$.

4. When $a = -\infty$ and $b = +\infty$, by setting $t = \log(\widetilde{x}/(1 - \widetilde{x}))$,
   $\int_a^b \xi_i(x) f_p(x) dx = \int_0^1 \xi(\log(\widetilde{x}/(1 - \widetilde{x}))) f_p(\log(\widetilde{x}/(1 - \widetilde{x}))) \frac{1}{\widetilde{x}(1-\widetilde{x})} d\widetilde{x}$.
   Define $\widetilde{\xi}_i(\widetilde{x}) = \xi(\log(\widetilde{x}/(1 - \widetilde{x}))) f_p(\log(\widetilde{x}/(1 - \widetilde{x}))) \frac{1}{\widetilde{x}(1-\widetilde{x})}$.

We can see that ξ_i is an element of $L^1(\mathfrak{X}^p, \mathfrak{B}_p, \mu_p; \Re^1)$ if and only if $\widetilde{\xi}_i$ is in $L^1([0, 1], \mathfrak{B}_{[0,1]}, \mathfrak{m}; \Re^1)$, the space of Lebesgue integrable functions defined on the unit interval $[0, 1]$. Let

$$\tau_0(\widetilde{x}) = 0, \tag{4.17}$$

for all $\widetilde{x} \in [0, 1]$; and

$$\tau_{k,j} = 2^{k-1}(I((2j - 2)2^{-k} \leq \widetilde{x} < (2j - 1)2^{-k}) - I((2j - 1)2^{-k} \leq \widetilde{x} < 2j \cdot 2^{-k})), \tag{4.18}$$

for k and $1 \leq j < 2^{k-1}$ and $\widetilde{x} \in [0,1]$. Sort them in the order, τ_0, $\tau_{1,1}$, $\tau_{2,1}$, $\tau_{2,2}$, $\cdots$, and re-label those functions as $\widetilde{\tau}_0, \widetilde{\tau}_1, \widetilde{\tau}_2, \cdots$. Then $\{\widetilde{\tau}_k\}$ is called the Haar bases for $L^1([0,1], \mathfrak{B}_{[0,1]}, \mathfrak{m}; \Re^1)$.[16] See Figure 1 for a graphic illustration. Choose an integer L, we approximate $\widetilde{\xi}_i$ by a linear combination of a finite number of those basis functions, i.e., $\widetilde{\xi}_i \approx \sum_{l=0}^{L} \kappa_{i,L,l}\widetilde{\tau}_l$. From it, we get an approximation of ξ. Take the first case listed above as an example, $\xi_i(x) \approx \sum_{l=0}^{L} \Big(\kappa_{i,L,l}\widetilde{\tau}_l((x-a)/(b-a))\Big)/\Big(bf_p(x)\Big)$. Plugging this approximation back into the consistency condition, we get that

$$\begin{aligned} &\frac{\sum_{l=0}^{L} \kappa_{i,L,l}\widetilde{\tau}_l((x-a)/(b-a))}{bf_p(x)} \\ =&\int_a^b H_i\Big(u(X^g, X_i^c, \widetilde{y}) + \lambda \sum_{j\neq i} W_{n,ij} \sum_{l'=1}^{L} \frac{\kappa_{j,L,l'}\widetilde{\tau}_l((\widetilde{y}-a)/(b-a))}{bf_p(\widetilde{y})}\Big) f_p(\widetilde{y}|\widetilde{x})d\widetilde{y}, \end{aligned} \tag{4.19}$$

As for the integration, we can choose the quadrature points as we did before, i.e.,

$$\begin{aligned} &\sum_{l=0}^{L} \kappa_{i,L,l} \frac{\widetilde{\tau}_l((x-a)/(b-a))}{bf_p(x)} \\ =&\sum_{k=1}^{K} \omega_k H_i\Big(u(X^g, X_i^c, \frac{(\upsilon_k+1)(b-a)}{2} + a) + \lambda \sum_{j\neq i} W_{n,ij} \sum_{l'=1}^{L} \kappa_{j,L,l'} \frac{\widetilde{\tau}_l((\upsilon_k+1)/2)}{bf_p(\frac{(\upsilon_k+1)(b-a)}{2} + a)}\Big) \\ &\cdot f_p(\frac{(\upsilon_k+1)(b-a)}{2} + a|\widetilde{x})d\widetilde{y}\frac{b-a}{2}, \end{aligned} \tag{4.20}$$

for $i = 1, \cdots, n$, $l = 1, \cdots, L$, and $k = 1, \cdots, K$. The Gauss-Legendre quadrature abscrissae, υ_k, and weights, ω_k, are fixed. The Haar bases, $\widetilde{\tau}_l$'s, are known functions. From the data, it is possible to get the distribution density of X_i^p's conditional on public information. Consequently, once we plug in $x = \frac{(\upsilon_k+1)(b-a)}{2}$, the value of $\frac{\widetilde{\tau}_l((x-a)/(b-a))}{bf_p(x)}$ can be calculated. In this way, we derive a system of nK nonlinear equations for nL coefficients, $\kappa_{i,L,l}$'s. Choosing $K = L$ and applying the homotopy method when analytic extension is possible, we may get multiple solutions, $\kappa_{i,L,l}^d$, for $d = 1, \cdots, D$. They correspond to multiple equilibria. Approximations when $[a, b]$ is unbounded can be computed in a similar way by changing variables.

Comparing the above approaches, we can see that using the first method, simple function approximation, we first fix the class of functions which are used to make approximation and then pin down the unknown coefficients of equilibrium functions by the equilibrium condition. However, there is not a general guidance on the choice of partitioning cutoffs. For the second and third method, we do not fix the form of equilibrium expectation functions. Instead, we

[16]Since $\|\widetilde{\tau}_k\|_1 = 1$, the Haar basis is a basis with unit norm.

first solve the values of an equilibrium expectation function at a fixed finite set of points and then use those values to approximate the equilibrium expectation function at any point in its domain. These two methods also specify how those points are chosen. Using the last method, we construct a flexible form of functions using basis functions and then employ the quadrature method to approximate integration. Using the second method, approximation precision depends on the number of quadrature abscissae. For the fourth method, instead, the approximation performance hinges on the number of function basis chosen.

4.2.3 Equilibrium Selection and Parameter Identification

Given a precision $\eta > 0$, consider a finite approximation to the set of equilibria, $\mathfrak{E}(X, W_n)$, $\mathfrak{E}_0(X, W_n) = \{\xi^{e,1}(X, W_n), \cdots, \xi^{e,D}(X, W_n)\}$,
we can derive an approximation to the likelihood function (2.14):

$$\widetilde{L}(Y; X, W_n) = \sum_{d=1}^{D} \rho(\xi^{e,d}) \prod_{i=1}^{n} f(y_i|\xi^e). \tag{4.21}$$

For example, we can set

$$\widetilde{\rho}(\xi^{e,d}; \mathfrak{E}_0(X, W_n), \alpha) = \frac{\exp(\alpha' \gamma(\xi^{e,d}; X, W_n))}{\sum_{d'=1}^{D} \exp(\alpha' \gamma(\xi^{e,d'}; X, W_n))}. \tag{4.22}$$

Then we derived the (approximated) sample likelihood function:

$$\widetilde{L}(y; X, W_n) = \sum_{d=1}^{D} \widetilde{\rho}(\xi^{d,k}; \mathfrak{E}(X, W_n), \alpha) \prod_{i=1}^{n} f(y_i|\xi^{e,k}). \tag{4.23}$$

When there are G independent groups, the corresponding approximation to the log likelihood of the whole sample is

$$\begin{aligned} &\log \widetilde{L}(Y_1, \cdots, Y_G|\beta, \lambda, \sigma, \alpha) \\ =& \sum_{g=1}^{G} \log\Big(\sum_{d=1}^{D_g} \frac{\exp(\alpha' \gamma(\xi^{e,g,d}; X_g, W_g))}{\sum_{d'=1}^{D_g} \exp(\alpha' \gamma(\widetilde{\xi}^{e,g,d'}; X_g, W_g))} \prod_{i=1}^{n_g} f(y_{i,g}|\xi^{e,g,d})\Big). \end{aligned} \tag{4.24}$$

As for identification, β, λ, and σ can still be identified using the strategy of "identification at infinity". That is because those parameters can be separately from α using this technique. However, it is more difficult to identify α. The reason is that (4.23) is not the exact likelihood but an approximation. However, as the set of equilibria is determined by β, λ, and σ, it does not depend on α. If we assume that all independent groups choose the same approximation, $D_g = D$, and use the same probability mass, (4.22), we can identify α from variation across groups.

Parameters can be estimated either by directly maximizing the approximated likelihood function or simulated moment conditions, similar to the approach used when all exogenous characteristics are public information. The performance of estimates relies on the approximation. In this paper, actually, there are two types of approximations. First, approximate the set of equilibria by a finite subset. Second, approximate each of those functions.

Regarding the equilibrium conditional expectation on i's behavior based on a structure of private information, $J_{i,m}$, as a function of the realization of $X^p_{i,m}$, the model under a general form of incomplete information about X^p_i's can be estimated in a similar way.

5 Peer Effects

Consider the case that an agent's behaviors are affected by the performances of her peers. That is, $W_{n,ij} = 1$ for all $i \neq j$. Since every agent makes predictions on anyone else, we denote the set of all possible private information about X^p_i's in the group as $\widehat{J} = \left\{ \widetilde{J} : \widetilde{J} = J_i \text{ } for \text{ } some \text{ } i = 1, \cdots, n \right\}$. Denote the number of elements in this set as M_0. We can denote the set as $\widehat{J} = \left\{ \widetilde{J}_1, \cdots, \widetilde{J}_{M_0} \right\}$. For each i, there is a unique $m(i)$ with $1 \leq m(i) \leq M_0$ such that $J_i = \widetilde{J}_{m(i)}$. We use $x^p_{J,m}$ to represent one realization of the random vector $X^p_{J,m}$ corresponding to the private information, $\widetilde{J}_m$. The corresponding support is denoted as $\mathfrak{X}^p_{J,m}$. Note that $X^p_{J,m(i)} = X^p_{J_i}$, representing the private information known to i. The conditional expectation function, $\xi^e = (\xi^e_{i,m})$, takes the following special form:

$$\xi^e(x^p_{J,1}, \cdots, x^p_{J,M_0})_{i,m} = \xi^e_{i,m}(x^p_{J,m}) = E[H_i(u(X_i) + \lambda \sum_{j \neq i} \xi^e_{j,m}(X^p_{m(i)}))|X^p_{J,m} = x^p_{J,m}, Z], \quad (5.1)$$

for all $i = 1, \cdots, n$, $m = 1, \cdots, M_0$, and $x^p_{J,m} \in \mathfrak{X}^p_{J,m}$.[17] When all exogenous covariates are public information, ξ^e reduces to an $n \times 1$ vector, satisfying $\xi^e_i = E[H_i(u(X_i) + \lambda \sum_{j \neq i} \xi^e_j)|Z = z]$. Since only the total expected behaviors of peers is taken into account when an agent is making decisions, we wonder whether it is possible to represent an equilibrium by a vector-valued function, the dimension of whose range is less than n. That means the number of coordinate functions, $\xi^e_{i,m}$'s, is reduced. Lee, Li and Lin(2014) prove that this is possible in the binary choice

[17] In the general case, we define the expectation about i conditional on private information for agents other than i. For convenience, when discussing peer effects, we also define conditional expectation about i based on private information the same as i's. Since $W_{n,ii} = 0$ for all i, model equilibria and implications do not change with this small alteration.

model when all exogenous characteristics are public information and only the idiosyncratic shocks are privately known. Their results can be extended to the general setting in this paper.

Let $\overline{\xi} = (\overline{\xi}_1, \cdots, \overline{\xi}_{M_0})$ be a vector-valued function such that

$$\overline{\xi}(x^p_{J,1}, \cdots, x^p_{J,M_0})_m = \overline{\xi}_m(x^p_{J,m}), \tag{5.2}$$

for any $m = 1, \cdots, M_0$. For any i, there is a unique $m(i) = 1, \cdots, M_0$, such that $J_i = \widetilde{J}_{m(i)}$. Consider a function equation system, $G_i(\overline{\xi}_{m(i)}, \xi_i; X, W_n) = 0$, where $\xi_i = (\xi_{i,1}, \cdots, \xi_{i,M_0})$ is a vector-valued function which satisfies (5.2) and that for any m,

$$\begin{aligned}
&\big(G_i(\overline{\xi}_{m(i)}, \xi_i; X, W_n)(x^p_{J,1}, \cdots, x^p_{J,M_0})\big)_m \\
=&\big(G_i(\overline{\xi}_{m(i)}, \xi_i; X, W_n)_m\big)(x^p_{J,m}) \\
=&E[H_i(u(X_i) + \lambda\overline{\xi}_{m(i)}(X^p_{J,m(i)}) - \lambda\xi_{i,m(i)}(X^p_{J,m(i)}))|X^p_{J,m} = x^p_{J,m}, Z = z] - \xi_{i,m}(x^p_{J,m}) \\
=&0,
\end{aligned} \tag{5.3}$$

for all $x^p_{J,m} \in \mathfrak{X}^p_{J,m}$. Applying the Brouwer fixed point theorem, for any $\overline{\xi}_{m(i)}$, there is a function ξ_i satisfying the above system of equations. For any $\chi = (\chi_1, \cdots, \chi_{M_0})$ with $\chi_m \in L^1(\mathfrak{X}^p_{J,m}, \mathfrak{B}_{J,m}, \mu_p; \Re^1)$, consider the linear operator $\Delta_i(\overline{\xi}_{m(i)}, \xi_i; X, W_n)$ such that

$$\begin{aligned}
&\big(\Delta_i(\overline{\xi}_{m(i)}, \xi_i; X, W_n)(\chi)\big)_m(x^p_{J,m}) \\
=&-\lambda E[DH_i\big(u(X_i) + \lambda\overline{\xi}(X^p_{J,m(i)}) - \lambda\xi_{i,m(i)}(X^p_{J,m(i)})\big)\cdot \\
&\chi_m(X^p_{J,m(i)})|X^p_{J,m} = x^p_{J,m}, Z = z] - \chi_m(x^p_{J,m});
\end{aligned}$$

for $m = m(i)$; and

$$\big(\Delta_i(\overline{\xi}_{m(i)}, \xi_i; X, W_n)(\chi)\big)_m(x^p_{J,m}) = -\chi_m(x^p_{J,m}),$$

for $m = 1, \cdots, M_0$ and $m \neq m(i)$, where $DH_i(a)$ denotes the derivative of $H_i(\cdot)$ at point a. $\Delta_i(\overline{\xi}_{m(i)}, \xi_i; X, W_n)$ is the Fréchet derivative of G_i with respect to ξ_i at $(\overline{\xi}_{m(i)}, \xi_i; X, W_n)$. If $\Delta(\overline{\xi}_{m(i)}, \xi_i; X, W_n)$ is an isomorphism, by the Implicit Function Theorem in Banach spaces, for a neighborhood $\overline{\xi}^e_{m(i)}$, there is only one ξ^e_i such that the functional equation, $G_i(\overline{\xi}_{m(i)}, \xi_i; X, W_n) = 0$ is satisfied. In this way, we can derive an operator Λ_i, which defines each function ξ_i as the image of $\overline{\xi}_{m_i}$ such that $G_i(\overline{\xi}_{m(i)}, \Lambda_i(\overline{\xi}_{m(i)}); X, W_n) = 0$. If $\Delta_i(\overline{\xi}_{m(i)}, \xi_i; X, W_n)$ is an isomorphism for all i, we call the group to be regular.

Proposition 5.1 *Suppose that the group is regular. If there is a function* $\overline{\xi}^e = (\overline{\xi}^e_1, \cdots, \overline{\xi}^e_{M_0})$

such that it satisfies (5.2), *and*

$$\begin{aligned}\overline{\xi}^e_m(x^p_{J,m}) =& \sum_{i=1}^n E[H_i(u(X^g, X^c_i, X^p_i) + \lambda\overline{\xi}^e_{m(i)}(X^p_{J,m(i)}) \\ & - \lambda(\Lambda_i(\overline{\xi}^e_{m(i)}))_{m(i)}(X^p_{J,m(i)}))|X^p_{J,m} = x^p_{J,m}, z],\end{aligned} \tag{5.4}$$

where $\Lambda_i(\cdot)$ *is defined above, then there is a vector of functions,*

$$\xi^e = (\xi^e_{1,1}, \cdots, \xi^e_{1,M(0)}, \cdots, \xi^e_{n,1}, \cdots, \xi^e_{n,M_0}),$$

such that (5.1) *holds for all* i, m, *and* $x^p_{J,m} \in \mathfrak{X}^p_{J,m}$. *On the contrary, if there is a vector of functions,* $\xi^e = (\xi^e_{1,1}, \cdots, \xi^e_{1,M_0}, \cdots, \xi^e_{n,1}, \cdots, \xi^e_{n,M_0})$, *satisfying* (5.1), *there is a function* $\overline{\xi}^e = (\overline{\xi}^e_1, \cdots, \overline{\xi}^e_{M_0})$ *such that* (5.2) *and* (5.4) *hold.*

Proof. See Appendix F. ∎

Particularly, when all exogenous characteristics are public information, the functional equation system (5.3) reduces to

$$G_i(\overline{\xi}^e, \xi^e_i; X, W_n) = H_i(u(X_i) + \lambda\overline{\xi}^e - \lambda\xi^e_i) - \xi^e_i = 0, \tag{5.5}$$

which is just a nonlinear equation. The regularity condition then reduces to

$$-\lambda DH_i(u(X_i) + \lambda\overline{\xi}^e - \lambda\xi^e_i) - 1 \neq 0.$$

We can see that if $DH_i(a) > 0$ for all i and a and $\lambda \geq 0$, the regularity condition is satisfied. Then a BNE is equivalent to the expected group total outcomes, $\overline{\xi}^e$, which is a scalar. The system analyzed by Lee, Li and Lin (2014) corresponds to the special case that $H_i(a) = F_\epsilon(a)$. When each X^p_i is known to i only and the joint distribution of X^p_i's conditional on the public information $Z = z$ is exchangeable with a pdf $f_p(\cdot)$, (5.3) takes the following form:

$$\begin{aligned}& G_i(\overline{\xi}^e, \xi^e_i; X, W_n)(x) \\ =& \int_{\mathfrak{X}^p} H_i(u(X^g, X^c_i, \widetilde{x}) + \lambda\overline{\xi}^e(\widetilde{x}) - \lambda\xi^e_i(\widetilde{x}))f_p(\widetilde{x}|x)d\widetilde{x} - \xi^e_i(x) \\ =& 0,\end{aligned} \tag{5.6}$$

for all i and $x \in \mathfrak{X}^p$. In this case, a BNE is equivalent to the expectation of group total outcomes conditional on the realization of an individual's self-known characteristics. $\overline{\xi}^e(\cdot)$ is a function mapping the realization of that privately known characteristics to a real number. Therefore, focusing on peer effects, the dimension of an equilibrium can be reduced, simplifying estimation.

6 Binary Choice Models: Analysis and Experiments

This section discusses in detail the equilibrium sets of two forms of binary choice models, (2.5) and (2.6). Additionally, by Monte Carlo experiments, there is a comparison about the small sample performances between the maximum likelihood estimation with complete likelihood for multiple equilibria in this paper and the nested fixed point maximum likelihood estimation assuming unique equilibrium used by Yang and Lee(2017).

6.1 Binary Choice Model I

Consider the model, (2.1) and (2.2), where $h_i(z) = I(z > 0)$ for all i, such that (2.5) holds. According to Yang and Lee(2017), this corresponds to a game where agents in a group simultaneously choose between 0 and 1. In this case, the expected utility following choice "0" is normalized to be zero. Instead, if i chooses "1", her expected utility is affected by her expectations about others' actions, $u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} E[y_j | X^p_{J_i}, Z] - \epsilon_i$. Therefore, $y_i = I(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} E[y_j | X^p_{J_i}, Z] - \epsilon_i > 0)$. The entry game for a group of firms in the same industry is a case in point. Discussions in this section focuses on the case that all X_i's are public information and ϵ_i's are i.i.d. normal with mean 0 and variance 1. It follows from the previous analysis that an equilibrium is an $n \times 1$ vector in $\xi \in [0, 1]^n$, such that

$$\xi_i = \Phi(u_i + \lambda \sum_{j \neq i} W_{n,ij} \xi_j), \tag{6.1}$$

for $i = 1, \cdots, n$, where $\Phi(\cdot)$ is the cdf for standard normal distribution and u_i is a simplified notation for $u(X_i)$.

The investigation of the equilibrium set begins with the special case that every two group members are associated with each other. That is, $W_{n,ij} = 1$ for all $i \neq j$. As it is shown in Section H, in this case, an equilibrium conditional expectation is a scalar and can be represented as a zero of a nonlinear function. The characteristics of the equilibrium set is summarized by the following proposition,

Proposition 6.1 *In Binary Choice Model I* (2.5)*, consider a group of n agents such that any two of them are associated with each other, i.e., $W_{n,ij} = 1$ for all $i \neq j$. Suppose that $\lambda \neq 0$.*

- *If $-\sqrt{2\pi} < \lambda < 0$, there is a unique equilibrium;*

- *When $\lambda > 0$, there is a unique equilibrium if*

$$\min_{1\leq i\leq n} u_i > \frac{\lambda}{2}, \tag{6.2}$$

or

$$\max_{1\leq i\leq n} u_i < -\frac{\lambda}{2}, \tag{6.3}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are respectively the cdf and pdf of the standard normal distribution.

Proof. See Appendix G. ■

From Proposition 6.1, it is more likely to have multiple equilibria when $\lambda > 0$ than it is when $\lambda < 0$. Thus, further investigation of the equilibrium set focuses on the case that $\lambda > 0$. Figures **??** to **??** illustrates the characteristics of the equilibrium set as the group population and interaction intensity varies. Figures 2 and 3 show that there is a unique equilibrium when (6.2) is satisfied. Figure 2 corresponds to the case that agents are symmetric with $u_i = u = 2$ for all i. Figure 3 is for the case that u_i's are heterogeneous and vary from 1 to 2. When neither (6.2) nor (6.3) is satisfied, as it is shown by Figures 4 and 5, there can be multiple equilibria. In Figure 4, symmetric agents have $u_i = u = -2$ for all i. However, u_i's change from -3 to -2 in Figure 5.

In the above graphs, there are at most three equilibria. Actually, the number of equilibria is no more than three under a reasonable condition.

Lemma 6.1 *Suppose that $0 < \lambda < \frac{2\sqrt{2\pi}}{3}$. For the function $c(a;\lambda) = \lambda\phi(a)+1+a^2(2\lambda\phi(a)-1)$, there is $a_+(\lambda) > 0$, such that $c(a;\lambda) > 0$ for $-a_+(\lambda) < a < a_+(\lambda)$; $c(a_+(\lambda);\lambda) = c(-a_+(\lambda);\lambda) = 0$, and $c(a;\lambda) < 0$, for $a < -a_+(\lambda)$ or $a > a_+(\lambda)$. In addition, $a_+(\lambda)$ increases with λ.*

Proof. See Appendix G. ■

Proposition 6.2 *In Binary Choice Model I (2.5), consider a group of n agents such that any two of them are associated with each other, i.e., $W_{n,ij} = 1$ for all $i \neq j$. Suppose that $0 < \lambda < \frac{2\sqrt{2\pi}}{3}$. There are at most three equilibria if*

$$\min_{1\leq i\leq n} u_i > \lambda\Phi(a_+(\lambda)) + a_+(\lambda), \tag{6.4}$$

or

$$\max_{1\leq i\leq n} u_i < \lambda\Phi(-a_+(\lambda)) - a_+(\lambda) - \lambda n. \tag{6.5}$$

Proof. See Appendix G. ■

For general social relations, an equilibrium is represented by an $n \times 1$ vector, $\xi = (E[y_1], \cdots, E[y_n])'$ and is a zero of a system of nonlinear equations. Applying Proposition 3.2, there is a unique equilibrium if the sign of the determinant of the following matrix does not change as ξ varies in $[0,1]^n$:

$$\lambda \begin{pmatrix} \phi(u_1 + \lambda \sum_{j \neq 1} W_{n,1j}\xi_j) & 0 & \cdots & 0 \\ 0 & \phi(u_2 + \lambda \sum_{j \neq 2} W_{n,2j}\xi_j) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi(u_n + \lambda \sum_{j \neq n} W_{n,nj}\xi_j) \end{pmatrix} W_n - I_n. \quad (6.6)$$

For a graphical illustration, suppose that $u(X_i)$ is a linear function:

$$u(X_i) = \beta_0 + X_{i,1}^c \beta_1 + X_{i,2}^c \beta_2, \quad (6.7)$$

where $X_{i,1}^c$ and $X_{i,2}^c$ are two commonly known exogenous characteristics. Take $\beta_0^* = 0$, $\beta_1^* = \beta_2^* = 1$. Simulate one sample with $G = 100$ independent groups, each of which has $n = 5$ members. In each group, the number of social relations an agent can build is randomly determined, ranging from 0 to $n-1$. Based on the randomly generated total link number for agent i, $F_i = \sum_{j \neq i} W_{n,ij}$, for each $j \neq i$, generate a random number, $rn_{n,ij}$. Then $W_{n,ij} = 1$, if $rn_{n,ij}$ is among the F_i largest ones. The social relation matrix, W_n, is not row-normalized. Characteristics of the equilibrium set in this case are illustrated by Figures 6 and 7. It can be seen that for Binary Choice Model I, although the sufficient condition for equilibrium uniqueness in Yang and Lee(2017) is violated for a big proportion of groups when λ is a little bit bigger than 0.6, for groups in this sample, when λ is within 0 and 1, there is only a unique equilibrium. As λ increases, the average equilibrium outcomes and the proportion of agents who choose action 1 increase.

In the Monte Carlo experiments, the social relation matrix is constructed in the same way as it is for the illustration in Figures 6 and 7. There are $L = 400$ simulations. In each simulated sample, there are G independent groups with homogeneous population n. n is fixed at 5 in the experiments. There are two cases about the number of independent groups, $G = 100$ and $G = 200$. The true value of the interaction intensity is $\lambda^* = 0.8$. Table 1 summarizes the estimation results for three regression methods. Regression I is the conventional Probit estimation without social interactions. Regressions II and III take into account the interactions among socially related agents. Regression II assumes condition (2.12) is satisfied and uses the contraction mapping iteration method to solve for the (assumed) unique equilibrium. On the contrary, Regression III does not make restrictive assumptions on the intensity of social interactions, λ, or the number of equilibrium in estimation. It allows for multiple equilibria, uses the homotopy continuation method to compute the equilibrium set, and chooses the equilibria with maximal expected probabilities for choice "1" to complete the model. From the table, it can be seen that Regression III outperforms Regression I and II in terms of parameter estimation biases and the value of estimated average log likelihoods. That is because of the distortions imposed by (2.12). From Table 1, condition (2.12) is violated by 67.44% (67.35%)of the groups in the sample on average when the number of groups is $G = 100$ ($G = 200$) under the true parameter values. Additionally, the average upper bound on the interaction intensity for a sample imposed by (2.12), 0.6267, is very close to the estimates of λ in Regression II (The estimate for λ is 0.6263 when $G = 100$ and 0.6266 when $G = 200$.) Therefore, imposing (2.12) can be restrictive when it is violated by a large proportion of the groups in a sample.

6.2 Binary Choice Model II

In the basic framework, (2.1) and (2.2), take $h_i(z) = 2I(z > 0) - 1$ for all i. Then (2.6) holds. Similar to Brock and Durlauf(2001), this model describes the equilibrium outcomes of a simultaneous move game with discrete choices where the utility an agent gets depends on the difference between her own action and those of her friends. To be specific, suppose that in a group of n agents, an agent i can choose two actions, -1 and 1. Her utilities depend on her own choice and those of the agents who she is associated with. If she chooses 1, with others choosing y_{-i}, her utility is $\widetilde{u}(X_i, 1) + \widetilde{\lambda} \sum_{j \neq i} W_{n,ij} y_j + \widetilde{\epsilon}_i^1$. If she chooses -1, her utility is $\widetilde{u}(X_i, -1) - \widetilde{\lambda} \sum_{j \neq i} W_{n,ij} y_j + \widetilde{\epsilon}_i^{-1}$. When $\widetilde{\lambda} > 0$, an agent benefits from taking the same action as her friends and/or peers do. In contrast, when $\widetilde{\lambda} < 0$, an agent gets rewarded by distinguishing herself from her friends and/or peers. Suppose that all exogenous characteristics are public information but the idiosyncratic shocks are private information. As i does not know her friends' actions when her decisions are made, she has to maximize her expected utility, which is $\widetilde{u}(X_i,1)+\widetilde{\lambda}\sum_{j\neq i}W_{n,ij}E[y_j]+\widetilde{\epsilon}_i^1$ for action 1 and $\widetilde{u}(X_i,-1)-\widetilde{\lambda}\sum_{j\neq i}W_{n,ij}y_j+\widetilde{\epsilon}_i^{-1}$ for action -1. Thus, she will choose 1 if

$$\widetilde{u}(X_i,1)-\widetilde{u}(X_i,-1)+2\widetilde{\lambda}\sum_{j\neq i}W_{n,ij}E[y_j]-(\widetilde{\epsilon}_i^{-1}-\widetilde{\epsilon}_i^1)>0.$$

Define $u(X_i)=\widetilde{u}(X_i,1)-\widetilde{u}(X_i,-1)$, $\lambda=2\widetilde{\lambda}$, and $\epsilon_i=\widetilde{\epsilon}_i^{-1}-\widetilde{\epsilon}_i^1$. Plug them into (2.1) and (2.2). Choose $h_i(z)=2I(z>0)-1$. Then the Type II model for binary choices is derived. Suppose that $(\widetilde{\epsilon}_i^1,\widetilde{\epsilon}_i^{-1})$'s are i.i.d. normal with zero mean and variance $\frac{1}{2}$. An equilibrium conditional expectation, $\xi=(\xi_1,\cdots,\xi_n)'=(E[y_1],\cdots,E[y_n])'$, satisfies

$$\begin{aligned}\xi_i&=2\Phi(u(X_i)+\lambda\sum_{j\neq i}W_{n,ij}\xi_j)-1\\&=2\Phi(\widetilde{u}(X_i,1)-\widetilde{u}(X_i,-1)+2\widetilde{\lambda}\sum_{j\neq i}W_{n,ij}\xi_j)-1.\end{aligned}\tag{6.8}$$

If any two agents are associated with each other, i.e., $W_{n,ij}=1$ for any $i\neq j$, λ represents the intensity of influences from any other group member. Based on the previous discussions, in this case, the equilibrium can be described by the group total expected outcome, $\bar{\xi}=\sum_{i=1}^n\xi_i=\sum_{i=1}^n E[y_i]$. Similar to Binary Choice Model I, $\bar{\xi}$ can be described as a zero of a nonlinear function. It is possible to characterize the equilibrium set by analyzing this function.

Proposition 6.3 *In Binary Choice Model II* (2.5), *consider a group of n agents such that any two of them are associated with each other, i.e., $W_{n,ij}=1$ for all $i\neq j$. Suppose that $\lambda\neq 0$.*

- *If $-\frac{\sqrt{2\pi}}{2}<\lambda<0$, there is a unique equilibrium;*
- *When $\lambda>0$, there is a unique equilibrium if*

$$\min_{1\leq i\leq n}u_i>\lambda n,\tag{6.9}$$

or

$$\max_{1\leq i\leq n}u_i<-\lambda n,\tag{6.10}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are respectively the cdf and pdf of the standard normal distribution.

Proof. See Appendix G. ■

Compare Proposition 6.1 and Proposition 6.3, it is easy to see that for both types of the Binary choice models, it is easier to ensure uniqueness when $\lambda<0$ than it is when $\lambda>0$.

Additionally, the sufficient conditions for a unique equilibrium is more stringent for the Type II model of binary choices than it is for the Type I model. Equilibrium multiplicity for the Type II Binary Choice model is illustrated by Figures **??** to **??**. Figures 8 and 9 show that there is a unique equilibrium when (6.9) is satisfied. Figure **??** is for the case that in a group of population n, $u_i = n+1$ for all i. Figure 9 depicts the case for heterogeneous agents such that $\max_{1\leq i\leq n} u_i = n+3$ and $\min_{1\leq i\leq n} u_i = n+1$. Figures 10 and 11 show that there can be multiple equilibria when both (6.9) and (6.10) are violated. Figure 10 corresponds to the case of homogeneous agents with $u_i = u = 1$ for all i. The case for heterogeneous agents with u_i's randomly change from 1 to 2 is shown in Figure 11. It is interesting to see that there are multiple equilibria in this case for the Type II model of Binary choices (as it is shown in Figure 11) and there is a unique equilibrium for the Type I model (as it is shown in Figure 3). These graphical illustrations confirm that it is more likely to have multiple equilibria in the Type II model of binary choices.

Similar to the discussions for the Type I binary choice model, under a certain condition, there are at most three equilibria in the Type II model.

Lemma 6.2 *Suppose that $0 < \lambda < \frac{\sqrt{2\pi}}{3}$. Fix λ, define a function $\widetilde{c}(a;\lambda) = 2\lambda\phi(a) + 1 + a^2(4\lambda\phi(a) - 1)$. There is $\widetilde{a}_+(\lambda) > 0$, such that $\widetilde{c}(a;\lambda) > 0$, for $-\widetilde{a}_+(\lambda) < a < \widetilde{a}_+(\lambda)$; $\widetilde{c}(\widetilde{a}_+(\lambda);\lambda) = \widetilde{c}(-\widetilde{a}_+(\lambda);\lambda) = 0$; and $\widetilde{c}(a;\lambda) < 0$, for $a < -\widetilde{a}_+(\lambda)$ or $a > \widetilde{a}_+(\lambda)$. In addition, $\widetilde{a}_+(\lambda)$ increases with λ.*

Proof. See Appendix G. ■

Proposition 6.4 *In Binary Choice Model II (2.6), consider a group of n agents such that any two of them are associated with each other, i.e., $W_{n,ij} = 1$ for all $i \neq j$. Suppose that $0 < \lambda < \frac{\sqrt{2\pi}}{3}$. There are at most three equilibria if*

$$\min_{1\leq i\leq n} u_i > \lambda(2\Phi(\widetilde{a}_+(\lambda)) - 1 + n) + \widetilde{a}_+(\lambda), \tag{6.11}$$

or

$$\max_{1\leq i\leq n} u_i < \lambda(2\Phi(-\widetilde{a}_+(\lambda)) - 1 - n) - \widetilde{a}_+(\lambda). \tag{6.12}$$

Proof. See Appendix G. ■

For general social relation matrix, W_n, an equilibrium is an $n \times 1$ vector satisfying a system of nonlinear equations. Applying Proposition 3.2, there is a unique equilibrium if the sign of the determinant of the following matrix does not change as ξ varies in $[-1,1]^n$:

$$2\lambda \begin{pmatrix} \phi(u_1 + \lambda \sum_{j\neq 1} W_{n,1j}\xi_j) & 0 & \cdots & 0 \\ 0 & \phi(u_2 + \lambda \sum_{j\neq 2} W_{n,2j}\xi_j) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi(u_n + \lambda \sum_{j\neq n} W_{n,nj}\xi_j) \end{pmatrix} W_n - I_n. \tag{6.13}$$

When there are multiple equilibria, the selection rule is applied to complete the model. Based on previous discussions, there are a finite number of equilibria. Denote them by $\{\xi^{e,1}, \cdots, \xi^{e,L}\}$. Suppose that equilibria are selected according to the total (ex ante) expected utilities. That is, $\gamma(\xi^{e,l}, X, W) = \sum_{i=1}^n U_i(\xi^{e,l})$, where

$$\begin{aligned}
&U_i(\xi^{e,l}) \\
=&\int I(\widetilde{u}(X_i,1) + \widetilde{\lambda}\sum_{j\neq i} W_{n,ij}\xi_j^{e,l} + \widetilde{\epsilon}_i^1 > \widetilde{u}(X_i,-1) - \widetilde{\lambda}\sum_{j\neq i} W_{n,ij}\xi_j^{e,l} + \widetilde{\epsilon}_i^{-1}) \\
&\cdot (\widetilde{u}(X_i,1) + \widetilde{\lambda}\sum_{j\neq i} W_{n,ij}\xi_j^{e,l} + \widetilde{\epsilon}_i^1)\frac{1}{\sigma^2}\phi(\frac{\widetilde{\epsilon}_i^1}{\sigma})\phi(\frac{\widetilde{\epsilon}_i^{-1}}{\sigma})d\widetilde{\epsilon}_i^1 d\widetilde{\epsilon}_i^{-1} \\
&+\int I(\widetilde{u}(X_i,1) + \widetilde{\lambda}\sum_{j\neq i} W_{n,ij}\xi_j^{e,l} + \widetilde{\epsilon}_i^1 \leq \widetilde{u}(X_i,-1) - \widetilde{\lambda}\sum_{j\neq i} W_{n,ij}\xi_j^{e,l} + \widetilde{\epsilon}_i^{-1}) \\
&\cdot (\widetilde{u}(X_i,-1) - \widetilde{\lambda}\sum_{j\neq i} W_{n,ij}\xi_j^{e,l} + \widetilde{\epsilon}_i^{-1})\frac{1}{\sigma^2}\phi(\frac{\widetilde{\epsilon}_i^1}{\sigma})\phi(\frac{\widetilde{\epsilon}_i^{-1}}{\sigma})d\widetilde{\epsilon}_i^1 d\widetilde{\epsilon}_i^{-1} \\
=&(\widetilde{u}(X_i,1) - \widetilde{u}(X_i,-1))\frac{\xi_i^{e,l}+1}{2} + \widetilde{\lambda}\sum_{j\neq i} W_{n,ij}\xi_j^{e,l}\xi_i + \widetilde{u}(X_i,-1) \qquad (6.14)\\
&+\int \widetilde{\epsilon}_i^1 \Phi(\frac{\widetilde{u}(X_i,1) - \widetilde{u}(X_i,-1) + 2\widetilde{\lambda}\sum_{j\neq i} W_{n,ij}\xi_j^{e,l} + \widetilde{\epsilon}_i^1}{\sigma})\frac{1}{\sigma}\phi(\frac{\widetilde{\epsilon}_i^1}{\sigma})d\widetilde{\epsilon}_i^1 \\
&+\int \widetilde{\epsilon}_i^{-1} \Phi(\frac{\widetilde{u}(X_i,-1) - \widetilde{u}(X_i,1) - 2\widetilde{\lambda}\sum_{j\neq i} W_{n,ij}\xi_j^{e,l} + \widetilde{\epsilon}_i^{-1}}{\sigma})\frac{1}{\sigma}\phi(\frac{\widetilde{\epsilon}_i^{-1}}{\sigma})d\widetilde{\epsilon}_i^{-1} \\
=&(u(X_i) + \lambda\sum_{j\neq i} W_{n,ij}\xi_j^{e,l})\frac{\xi_i^{e,l}}{2} + u(X_i)\frac{1}{2} + \widetilde{u}(X_i,-1) \\
&+\int \widetilde{\epsilon}_i^1 \Phi(\frac{u(X_i) + \lambda\sum_{j\neq i} W_{n,ij}\xi_j^{e,l} + \widetilde{\epsilon}_i^1}{\sigma})\frac{1}{\sigma}\phi(\frac{\widetilde{\epsilon}_i^1}{\sigma})d\widetilde{\epsilon}_i^1 \\
&+\int \widetilde{\epsilon}_i^{-1} \Phi(\frac{-u(X_i) - \lambda\sum_{j\neq i} W_{n,ij}\xi_j^{e,l} + \widetilde{\epsilon}_i^{-1}}{\sigma})\frac{1}{\sigma}\phi(\frac{\widetilde{\epsilon}_i^{-1}}{\sigma})d\widetilde{\epsilon}_i^{-1}
\end{aligned}$$

In this model, $\sigma = \sqrt{\frac{1}{2}}$. $\widetilde{u}(X_i,-1)$ is normalized to be equal to 0. Thus, individual expected utilities can be computed given the model primitives, u_i's, and an equilibrium set. Although there are no explicit analytical forms for the last two integrals in (6.14), they can be computed numerically by the Gaussian quadrature.

For further investigation of the equilibrium set for general social relations, suppose that $u(X_i) = \widetilde{u}(X_i, 1) - \widetilde{u}(X_i, -1)$ takes the linear function form (6.7). Like the discussions in the Type I Binary Choice model, take $\beta_0^* = 0$ and $\beta_1^* = \beta_2^* = 1$. Consider a sample of $G = 100$ independent groups. Each of them has $n = 5$ members. In a group, the number of social links an agent has is randomly determined. For each i, generate a random number, f_i. If $f_i \geq 0.5$, the number of social links for i, $F_i = \sum_{j \neq i} W_{n,ij}$, is $n - 1$; otherwise, $F_i = n - 2$. Given F_i, generate random numbers, rn_{ij} for each $j \neq i$. Then $W_{n,ij} = 1$ if rn_{ij} is among the F_i largest. Use the homotopy continuation to compute the equilibrium set for this sample. Figures 12 and 13 illustrate the characteristics of the equilibrium set and the (selected) equilibrium outcomes. According to Yang and Lee(2017), when $|\lambda| < 0.3133$, there is a unique equilibrium. From Figure 12, there is still a unique equilibrium when λ is a little bit bigger than 0.3133. However, unlike the Type I model for binary choices, as λ continues to increase, some groups have more than one equilibria. Additionally, there is an increasing tendency for the sample average number of equilibria and the proportion of groups with multiple equilibria. Figure 13 shows the characteristics of the selected outcomes corresponding to the selection criterion (6.14). As the interaction intensity, λ, increases, there is not an obvious trend for the sample average expected equilibrium outcomes, actual outcomes, the proportion of agents who choose 1, and the proportion of agents who choose -1. This is different from the characteristics of the equilibrium outcomes in the Type I model of binary choices, where as λ increases more agents choose action 1 and the average equilibrium outcome also increases (see Figure 7).

Samples are generated in the same way for the Monte Carlo experiments. There are $L = 400$ simulations. Consider two values that λ takes, 0.2 and 0.8. When $\lambda = 0.2$, the sufficient condition for equilibrium uniqueness in Yang and Lee (2017) is satisfied. However, when $\lambda = 0.8$, as it is shown by Figure **??**, it is possible to have multiple equilibria. In experiments, four estimation method are compared: (1) conventional maximum likelihood estimation without social interactions, i.e., assuming $\lambda = 0$; (2) nested fixed point maximum likelihood estimation which assumes a unique equilibrium, restricts λ, and uses contraction mapping iterations to compute the equilibrium; (3) nested fixed point maximum likelihood estimation which assumes a unique equilibrium and computes the equilibrium by solving nonlinear equations without restricting λ; and (4) maximum likelihood estimation for complete likelihood with equilibrium selection which uses the homotopy continuation method to compute the set of equilibria. Results are summarized in the Tables 2 and 3.

It is obvious that the conventional regression which ignores the interactive effects among socially associated agents can bring in more biases than regressions which take into account social interactions even when the intensity of social interactions is moderate. When $\lambda^* = 0.2$, there is a unique equilibrium according to Yang and Lee(2017). From Table 2, the last three regression methods have similar performances. However, with large interaction intensity, $\lambda^* = 0.8$, from the tabulated results in Table 3, Regression IV outperforms the other three regressions. As the upper bound on interaction intensity which ensures equilibrium uniqueness is 0.3133 on average, which is far smaller than the true parameter value, $\lambda^* = 0.8$, when $|\lambda|$ is restricted within this bound, estimations are biased, which is shown by the performances of Regression II. Since the average estimates for λ by Regression I is 0.3109, which is very close to 0.3133, this upper bound is nearly binding. A potential improvement may be achieved by relaxing the restrictions on $|\lambda|$ and computing the equilibrium by solving a system of nonlinear equations, using numerical algorithms such as the Newton's method. This method is used in Regression III. Although this method performs well for moderate interactions when there is a unique equilibrium (as Table 2 shows), when $\lambda^* = 0.8$, it brings in biased estimation. That is because this method still assumes that there is a unique equilibrium. However, the average number of equilibria in the simulated sample is 2.2918 and 78.11% of groups have more than one equilibria. In Regression IV, equilibrium multiplicity is considered and equilibria are selected according their expected total utilities. It can be seen that Regression IV has much smaller biases and much larger estimated sample log likelihoods than those of other regression methods. That is, the computation intensity of Regression IV is rewarded with good estimation performances when there are multiple equilibria.

7 Conclusion

In a general framework of social interactions under incomplete information with multiple equilibria, this paper investigates the approach to complete the model, along with identification, computation, and estimation issues. The proposed solution to multiple equilibria extends the random equilibrium selection method used by Bajari et al(2010b) and (2010c) to a setting, which is general in the types of behaviors and information structures. Although this all-solution method can be computationally intensive, it does not impose strong assumptions on the data generating process and can be applied to a broad range of empirical studies. Although the model have specific structures on the interdependence among socially associated agents, it incorporates discrete and continuous choices, bounded and unbounded outcomes, unbounded idiosyncratic shocks, as well as different information structures. Therefore, the characterization of equilibria in this paper complements the existence theorems for the Bayesian Nash Equilibrium in the recent theory literature.

Group unobservables are important in empirical studies. In this paper, there is a brief discussion for the case when those unobservables only affect individual choices but not the equilibrium selection rule. It will be an interesting extension if that assumption is relaxed.

Using the stochastic selection rule, we do not need to worry about the case that two equilibria score the same according to that criterion, for only the distribution of equilibrium selection matters. If the deterministic rule is used, however, the model will still be incomplete when there is a tie between two equilibria according to the deterministic objective function. However, if we can apply some mathematical tools, such as optimal control, optimization over the equilibrium set may be less computationally intensive than the computations of all the equilibria. Consequently, further investigations of the deterministic rule are of significance.

Appendices

A Games and Equilibria

A.1 Game Explanations

In the model, (2.1) and (2.2), an agent's s behavior is interacted with those of others when she is uncertain about some of their attributes. Hence, we can view outcomes of the model as the outcome of an equilibrium for a simultaneous move game with incomplete information. According to Harsanyi (1967a; 1967b), assuming that agents' payoffs are related to a randomly determined "state", an agent's uncertainty comes from the fact that her signal does not completely recover the true state. Then predicting others' unknown characteristics is equivalent to making inference about the realized state from her own signal. We follow the setup in Osborne and Rubinstein (1994).

For n group members, let $\mathfrak{X}_i^p$ represent the support of X_i^p and $\mathcal{E}$ the common support of ϵ_i's. The set of states, $\prod_i^n \mathfrak{X}_i \times \mathcal{E}^n$, is the set of all possible X_i^p's and ϵ_i's for all players. In this case, player i's "type" is her private information, $X_{J_i}^p$, and idiosyncratic shocks, ϵ_i. Her set of types is then $\mathcal{T}_i = \prod_{k:J_i(k)=1} \mathfrak{X}_k \times \mathcal{E}$. The signal function is a mapping from the states to her type, $\tau_i : \prod_i^n \mathfrak{X}_i \times \mathcal{E}^n \rightarrow \prod_{k:J_i(k)=1} \mathfrak{X}_k \times \mathcal{E}$. Her prior belief on the set of states is the joint distribution of X_i^p's and the distribution of the i.i.d. shocks, $F_\epsilon(\cdot)$, which is the same for all agents. The set of actions for agent i is denoted by $\mathcal{Y}_i$. Her strategy is a contingent plan specifying the action to take for each realized type, $s_i : \prod_{k:J_i(k)=1} \mathfrak{X}_k \times \mathcal{E} \rightarrow \mathcal{Y}_i$. The payoff received by an agent depends on actions taken by all group members, $y = (y_1, \cdots, y_n) \in \prod_{i=1}^n \mathcal{Y}_i$, as well as the uncertain state. $s^e = (s_1^e(\cdot), \cdots, s_n^e(\cdot))$ is a Bayesian Nash Equilibrium (BNE) in this model if

$$s_i^e(X_{J_i}^p, \epsilon_i) = h_i(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} E[s_j^e(X_{J_j}^p, \epsilon_j) | X_{J_i}^p, Z, \epsilon_i] - \epsilon_i). \tag{A.1}$$

With specific $h_i(\cdot)$ functions, it is possible to build a structural model. See Yang and Lee(2017) for continuous and binary choices and Yang, Qu and Lee (2016) for the Tobit model when agents' actions are subject to the non-negative constraint.

A.2 Equilibrium and Expectations

Under the assumption that observed outcomes are realizations of a BNE (A.1), we relate equilibrium strategies, $s^e = (s_1^e(\cdot), \cdots, s_n^e(\cdot))$, to conditional expected outcomes in an equilibrium, $\left\{E[y_j | X_{J_i}^p, Z = z]\right\}$. Pick any i and k such that $i \neq k$. By consistency, we get that

$$E[y_i | X_{J_k}^p, Z] = E[h_i(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij} E[y_j | X_{J_i}^p, Z = z] - \epsilon_i) | X_{J_k}^p, Z]. \tag{A.2}$$

Therefore, given public information $Z = z$, for any i, conditional expectations on i's behavior depends on the private information. Two agents k and k', where $k \neq i$, $k' \neq i$, and $k \neq k'$, have the same expectations about i's behavior if they have the same private information, i.e., $X_{J_k}^p = X_{J_{k'}}^p$.

Similar to Yang and Lee (2017), conditional expectations can be modeled as functions and embedded into a function space. In this paper, however, conditional expectation functions are defined in a different way in order to utilize properties of classical function spaces.[18] Given social relations W_n and information structure $\mathfrak{J} = (J_1, \cdots, J_n)$, for each i, we collect all possible types of private information which are used by others to predict her actions as,

$$\widehat{J}_i = \left\{ \widetilde{J} \in \{0,1\}^n : \widetilde{J} = J_j \textit{ for some } j \textit{ s.t. } W_{n,ji} \neq 0 \right\}. \tag{A.3}$$

Denote the number elements in this set by M_i.[19] Considering that two agents may have the same type of private information, $M_i \leq \sum_{j \neq i} M_{n,ji}$. Labeling elements in $\widehat{J}_i$ by $m = 1, \cdots, M_i$, we get

$$\widehat{J}_i = \left\{ \widetilde{J}_{i,1}, \cdots, \widetilde{J}_{i,M_i} \right\}. \tag{A.4}$$

For any j with $W_{n,ji} \neq 0$, there is exactly one vector in $\widehat{J}_i$ representing j's private information, J_j. That is, there is a unique $m_i(j) \in \{1, \cdots, M_i\}$, such that $J_j = \widetilde{J}_{i,m_i(j)}$. The mapping, $m_i(\cdot)$, defined in this way is onto. Let $\mathfrak{X}_i^p$ be the support of X_i^p (conditional on public information $Z = z$). It is a subset of $\Re^{k_p}$.(Recall that k_p is the dimension of X_i^p's.) For any (i, m), denote by $\mathfrak{X}_{i,m}^p$ the support of privately known characteristics contained in $\widetilde{J}_{i,m}$. Then it is a subset of the Euclidean space with dimension $k_p(\sum_{j=1}^n \widetilde{J}_{i,m}(j))$. We denote its elements simply by $x_{i,m}^p$. For example, if $\widetilde{J}_{i,m}(j) = 1$, for $j = 2, 3$; and $\widetilde{J}_{i,m}(j) = 0$, otherwise. $\mathfrak{X}_{i,m}^p$ is the support for (X_2^p, X_3^p). Its elements are realizations, $(x_2^{p'}, x_3^{p'})'$. Define $\xi_{i,m}^e : \mathfrak{X}_{i,m}^p \to \Re^1$ as

$$\xi_{i,m}^e(x_{i,m}^p) = E[y_i | X_{\widetilde{J}_{i,m}}^p = x_{\widetilde{J}_{i,m}}^p, Z = z]. \tag{A.5}$$

Then $\xi_{i,m}^e$ is a mapping from a subset of an Euclidean space with dimension $k_p \sum_{j=1}^n \widetilde{J}_{i,m}(j)$ to

[18]Treating the privately known characteristics used to make predictions as a random vector, Yang and Lee (2017) define the conditional expectation about i's behaviors as a function of all possible random vectors used to make predictions. That function maps a random vector in its domain to a random variable, specifying the value of conditional expectations for each realization of that random vector. Consider a group of 3 people as an example. Agent 1 is linked to agents 2 and 3, i.e., $W_{3,21} \neq 0$ and $W_{3,31} \neq 0$. $X_{J_2}^p$ and $X_{J_3}^p$ are the private information for 2 and 3 respectively. The conditional expectation about y_1, ψ_1, is then defined on the set of random vectors, $\mathfrak{A}_1 = \left\{X_{J_2}^p, X_{J_3}^p\right\}$. $\psi_1(X_{J_2}^p)$ is a random vector, such that for each ω in the sample space of X_i^p's, $\psi_1(X_{J_2}^p)(\omega) = E[y_1 | X_{J_2}^p = X_{J_2}^p(\omega), Z = z]$. In our model, however, we define exclusively a function for every possible type of private information that is used to predict i's behaviors. For the aforementioned example, if $J_2 \neq J_3$, we define functions ξ_{1,J_2} and ξ_{1,J_3}, on the support of those random vectors. That is, $\xi_{1,J_2}(x_{J_2}^p) = E[y_1 | X_{J_2}^p = x_{J_2}^p, Z = z]$ and $\xi_{1,J_3}(x_{J_3}^p) = E[y_1 | X_{J_3}^p = x_{J_3}^p, Z = z]$. The conditional expectation functions defined in this way are mappings from a subset of an Euclidean space to $\Re^1$, which makes it convenient to apply the properties of the classical L^p spaces.

[19]If $M_i = 0$, i's actions does not influence others' choices. Then expectations on her behaviors do not influence the distribution of outcomes. It is redundant in the system. In computation, we can just exclude the redundancy.

$\Re^1$.[20] Collecting all those functions, we derive a vector-valued function,

$$\xi^e = (\xi^e_{1,1}, \cdots, \xi^e_{1,M_1}, \cdots, \xi^e_{n,1}, \cdots, \xi^e_{n,M_n}),$$

whose domain is $\prod_{i=1}^n \prod_{m=1}^{M_i} \mathfrak{X}^p_{i,m}$ and range is $\Re^M$, where $M = \sum_{i=1}^n M_i$. ξ^e has two properties:

$$\xi^e(x^p_{1,1}, \cdots, x^p_{1,M_1}, \cdots, x^p_{n,1}, \cdots, x^p_{n,M_n})_{i,m} = \xi^e_{i,m}(x^p_{i,m}); \tag{A.6}$$

and the equilibrium condition,

$$\xi^e_{i,m}(x^p_{i,m}) = E[h_i(u(X^g, X^c_i, X^p_i) + \lambda \sum_{j \neq i} W_{n,ij} \xi^e_{j,m_j(i)}(X^p_{j,m_j(i)}) - \epsilon_i) | X^p_{i,m} = x^p_{i,m}, Z = z], \tag{A.7}$$

for all $i = 1, \cdots, n$, $m = 1, \cdots, M_i$, and $x^p_{i,m} \in \mathfrak{X}_{i,m}$.[21] In particular, when all exogenous characteristics are public information, conditional expectations only depend on public information. In that case, $E[y_i|Z = z]$ is a scalar for any i and ξ^e reduces to an $n \times 1$ vector, $\xi^e = (\xi^e_1, \cdots, \xi^e_n)'$.

B Expectations, Equilibria, and Functions

In this appendix, we embed conditional expectation functions into a function space. For a group with size n, social relations W_n and information structure J, we define a function space, $\Xi(W_n, J)$, such that each $\xi \in \Xi(W_n, \mathcal{J})$ is a mapping from an $\prod_{i=1}^n \prod_{m=1}^{M_i} \mathfrak{X}^p_{i,m}$ to $\Re^M$, satisfying

1. For all $i = 1, \cdots, n$, $m = 1, \cdots, M_i$, and $x^p_{i,m} \in \mathfrak{X}^p_{i,m}$

$$\xi(x^p_{1,1}, \cdots, x^p_{1,M_1}, \cdots, x^p_{n,1}, \cdots, x^p_{n,M_n})_{i,m} = \xi_{i,m}(x^p_{i,m}). \tag{B.1}$$

2.

$$\max_{1 \le i \le n} \max_{1 \le m \le M_i} \int |\xi_{i,m}(x^p_{i,m})| d\mu_p < \infty, \tag{B.2}$$

where μ_p represents the L-S measure implied by the conditional distribution of X^p_i's given public information $Z = z$.

[20] In principle, $\xi^e_{i,m}$ can be defined on the whole Euclidean space. However, considering the support of X^p_i's may not be full, it will be convenient to work with bounded subset of the Euclidean space sometimes. Therefore, we just use an abstract subset at this stage.

[21] By our definition, given i, for each $j \neq i$, via the mapping $m_j(\cdot)$, we find exactly the private information $\widetilde{J}_{i,m_j(i)} = J_i$. Thus, in (A.7), all the $X^p_{j,m_j(i)}$'s are the same. They are just $X^p_{J_i}$, the random vector of exogenous characteristics which are known by i.

Define summation and scalar product in $\Xi(W_n, \mathcal{J})$ in a conventional way. According to (B.2), define the norm on $\Xi(W_n, J)$ as

$$\|\xi\| = \max_{1\le i\le n} \max_{1\le m\le M_i} \int |\xi_{i,m}(x^p_{i,m})| d\mu_p. \tag{B.3}$$

Lemma B.1 *The norm, $\|\cdot\|$, defined in* (B.3) *is well-defined.*

Proof. It is obviously that $\|\xi\| \ge 0$ for any $\xi \in (\Xi(W_n, \mathcal{J}), \|\cdot\|)$ and $\|\xi\| = 0$ if and only if $\xi = 0$ a.e. according to μ_p. For any real scalar α and $\xi \in (\Xi(W_n, \mathcal{J}), \|\cdot\|)$,

$$\|\alpha\xi\| = \max_{1\le i\le n} \max_{1\le m\le M_i} \int |\alpha\xi_{i,m}(x^p_{i,m})| d\mu_p = |\alpha| \max_{1\le i\le n} \max_{1\le m\le M_i} \int |\xi_{i,m}(x^p_{i,m})| d\mu_p = |\alpha| \|\xi\|.$$

For any two elements, $\xi, \xi' \in (\Xi(W_n, J), \|\cdot\|)$,

$$\begin{aligned}
\|\xi + \xi'\| &= \max_{1\le i\le n} \max_{1\le m\le M_i} \int |\xi_{i,m}(x^p_{i,m}) + \xi'_{i,m}(x^p_{i,m})| d\mu_p \\
&\le \max_{1\le i\le n} \max_{1\le m\le M_i} \int |\xi_{i,m}(x^p_{i,m})| d\mu_p + \max_{1\le i\le n} \max_{1\le m\le M_i} \int |\xi'_{i,m}(x^p_{i,m})| d\mu_p \\
&= \|\xi\| + \|\xi'\|.
\end{aligned}$$

■

$L^1(\mathfrak{X}^p_{i,m}, \mathfrak{B}_{i,m}, \mu_p; \Re^1)$ is the space of all real-valued functions on $\mathfrak{X}^p_{i,m}$ which are measurable under μ_p such that $\|\chi\|_1 := \int_{\mathfrak{X}^p_{i,m}} \int |\chi| d\mu_p < \infty$ for all $\chi \in L^1(\mathfrak{X}^p_{i,m}, \mathfrak{B}_{i,m}, \mu_p; \Re^1)$. This space belongs to the class of Lebesgue spaces. According to Dunford and Schwartz (1958), it is a Banach space. Since ξ is a vector-valued function composing of a finite number of coordinate functions, with the norm defined as the maximal of the absolute integrable of its coordinate functions, $\xi \in (\Xi(W_n, \mathcal{J}), \|\cdot\|)$ if and only if each of its coordinates, $\xi_{i,e} \in L^1(\mathfrak{X}^p_{i,m}, \mathfrak{B}_{i,m}, \mu_p; \Re^1)$. Based on this finding, we derive the following result.

Proposition B.1 $(\Xi(W_n, \mathcal{J}), \|\cdot\|)$ *is complete. So it is a Banach space.*

Proof. Take any Cauchy sequence, $\{\xi^k\}$ in $(\Xi(W_n, J), \|\cdot\|)$. That is, $\|\xi^k - \xi^l\| = \max_{1\le i\le n} \max_{1\le m\le M_i} \int |\xi^k_{i,m}(x^p_{i,m}) - \xi^l_{i,m}(x^p_{i,m})| d\mu_p \to 0$, as $k, l \to \infty$. Then $\int |\xi^k_{i,m}(x^p_{i,m}) - \xi^l_{i,m}(x^p_{i,m})| d\mu_p \to 0$ as $k, l \to \infty$ uniformly for all $i = 1, \cdots, n$ and $m = 1, \cdots, M_i$ in the L^1 norm. For any $\eta > 0$, for each (i, m), the completeness of
$L^1(\mathfrak{X}^p_{i,m}, \mathfrak{B}_{i,m}, \mu_p; \Re^1)$, implied that there is $\xi_{i,m} \in L^1(\mathfrak{X}^p_{i,m}, \mathfrak{B}_{i,m}, \mu_p; \Re^1)$ such that there is $K_{i,m}(\eta) >$, whenever $k > K_{i,m}(\eta)$, $\int_{\mathfrak{X}^p_{i,m}} |\xi^k_{i,m}(x^p_{i,m}) - \xi_{i,m}(x^p_{i,m})| d\mu_p < \eta$. Since the total number

of coordinate function is finite, define $\xi = (\xi_{i,m})_{1\le i\le n, 1\le m\le M_i}$. For $K = \max_{i,m} K_{i,m}(\eta)$, whenever $k > K$, we have that $\|\xi^k - \xi\| = \max_{1\le i\le n}\max_{1\le m\le M_i}\int_{\mathfrak{X}^p_{i,m}} |\xi^k_{i,m}(x^p_{i,m}) - \xi_{i,m}(x^p_{i,m})| d\mu_p < \eta$. Moreover, since $\xi_{i,m} \in L^1(\mathfrak{X}^p_{i,m}, \mathfrak{B}_{i,m}, \mu_p; \Re^1)$ for all (i, m),
$\max_{1\le i\le n}\max_{1\le m\le M_i}\int_{\mathfrak{X}^p_{i,m}} |\xi_{i,m}| < \infty$. That is, $\xi \in (\Xi(W_n, \mathcal{J}), \|\cdot\|)$. ■

Proposition B.2 claims the equivalence between the BNEs and the equilibrium conditional expectations. This proposition is proved below.

Proposition B.2 *Conditional on public information $Z = z$, if $s^e = (s^e_1, \cdots, s^e_n) : \prod_{i=1}^n \mathcal{T}_i \times \mathcal{E}^n \to \prod_{i=1}^n \mathcal{Y}_i$ is a BNE of this model, then there is an equilibrium conditional expectation function ξ^e satisfying conditions* (A.6) *and* (A.7). *On the contrary, if there is an equilibrium conditional expectation function, there is a BNE for this model.*

Proof of Proposition B.2. If s^e is a BNE, define ξ^e by

$$\xi^e(x^p_{1,1}, \cdots, x^p_{1,M_1}, \cdots, x^p_{n,1}, \cdots, x^p_{n,M_n})_{i,m} = \xi^e_{i,m}(x^p_{i,m}) = E[s^e(X^p_{J_i}, \epsilon_i)|X^p_{\widetilde{J}_{i,m}} = x^p_{i,m}, z],$$

for any i, m, and $x^p_{i,m} \in \mathfrak{X}^p_{i,m}$. It follows from (A.1) that

$$\begin{aligned}
\xi^e_{i,m}(x^p_{i,m}) =& E[h_i(u(X_i) + \lambda\sum_{j\neq i} W_{n,ij} E[s^e_j(X^p_{J_j}, \epsilon_j)|X^p_{J_i}, z, \epsilon_i] - \epsilon_i)|X^p_{\widetilde{J}_{i,m}} = x^p_{i,m}, z]\\
=& E[h_i(u(X_i) + \lambda\sum_{j\neq i} W_{n,ij} E[s^e_j(X^p_{J_j}, \epsilon_j)|X^p_{J_i}, z] - \epsilon_i)|X^p_{\widetilde{J}_{i,m}} = x^p_{i,m}, z]\\
=& E[h_i(u(X_i) + \lambda\sum_{j\neq i} W_{n,ij} \xi^e_{j,m_j(i)}(X^p_{j,m_j(i)}) - \epsilon_i)|X^p_{\widetilde{J}_{i,m}} = x^p_{i,m}, z].
\end{aligned}$$

In the above equation, the second equality comes from the independence among ϵ_i's, and the independence between the idiosyncratic shocks and exogenous covariates and social relations. The third equality follows directly from the way in which ξ^e is defined. Note that $X^p_{j,m_j(i)} = X^p_{J_i}$ for any j with $W_{n,ij} \neq 0$. On the contrary, assume that ξ^e is an equilibrium conditional expectation function, i.e., (A.6) and (A.7) are satisfied. Define s^e by $s^e_i(X^p_{J_i}, \epsilon_i) = h_i(u(X_i) + \lambda\sum_{j\neq i} W_{n,ij}\xi^e_{j,m_j(i)}(X^p_{j,m_j(i)}) - \epsilon_i)$. Then we have that

$$\begin{aligned}
E[S^e_j(X^p_{J_j}, \epsilon_j)|X^p_{J_i}, z, \epsilon_i] =& E[h_j(u(X_j) + \lambda\sum_{k\neq j} W_{n,jk}\xi^e_{k,m_k(j)}(X^p_{k,m_k(j)}) - \epsilon_j)|X^p_{J_i}, z, \epsilon_i]\\
=& E[h_j(u(X_j) + \lambda\sum_{k\neq j} W_{n,jk}\xi^e_{k,m_k(j)}(X^p_{k,m_j(k)}) - \epsilon_j)|X^p_{J_i}, z]\\
=& \xi^e_{j,m_j(i)}(X^p_{j,m_j(i)}),
\end{aligned}$$

where $X^p_{j,m_j(i)}$ corresponds to $X^p_{J_i}$. The second equality follows from the independence among all ϵ_i's and the independence between those shocks and exogenous characteristics. The third equality is derived by applying (A.7). Therefore, by the above definition of s^e,

$$s_i^e(X^p_{J_i}, \epsilon_i) = h_i(u(X_i) + \lambda \sum_{j \neq i} W_{i,j} E[S_j^e(X^p_{J_j}, \epsilon_j)|X^p_{J_i}, z, \epsilon_i] - \epsilon_i).$$

■

C Proofs for Equilibrium Characterizations with Public Characteristics

In this section, we discuss in detail the structure of the set of equilibria when all exogenous covariates are public information. We first prove that there is no loss of generality by focusing on regular groups. We impose Assumption C.1 in order to apply theorems in differential topology.

Assumption C.1 *The functions, $u(\cdot)$ and $H_i(\cdot)$, for $i = 1, \cdots, n$, are smooth. That is, they have continuous partial derivatives of all orders.*

In most models used in empirical and theoretical studies, $u(\cdot)$ is linear function of exogenous covariates. Although $h_i(\cdot)$ can be discrete, $H_i(\cdot)$ defined in Assumption 2.2 is usually smooth as our examples show. In the subsequent discussions, we focus on the following case.

Definition C.1 *For a group (X, W_n), an equilibrium ξ^e is regular if the derivative of $S(\cdot; X, W_n)$ at ξ^e, $DS(\xi^e; X, W_n)$ is non-singular. A group (X, W_n) is regular if each of its equilibrium is regular. That is, $DS(\xi^e; X, W_n)$ is non-singular for any ξ^e such that $S(\xi^e; X, W_n) = 0$.*

When a group if regular, applying the Inverse Function Theorem, in a neighborhood of an equilibrium, ξ^e, there is no other equilibria. That is to say, all equilibria are isolated from each other. That property is important for the following discussion. We show in the proposition below that there is no loss of generality by focusing on regular groups. Denote the support of X by $\mathfrak{X}$. For any social matrix W_n, define function $\widetilde{S} : \Re^n \times \mathfrak{X} \to \Re^n$ by $\widetilde{S}(\xi, X; W_n) = S(\xi; X, W_n)$. That is, given W_n, $S(\cdot; X, W_n)$ can be viewed as a family of smooth maps, indexed by the exogenous covariates, X. By Assumption C.1, $\widetilde{S}$ is smooth. Proposition C.1 follows from the transversality theorem in differential topology.

Proposition C.1 *Given social matrix W_n, if $D\widetilde{S}(\xi, X; W_n)$ has full row rank for all (ξ, X) with $\widetilde{S}(\xi^e; X, W_n) = 0$, then for almost every $X \in \mathfrak{X}$, $DS(\xi^e; X, W_n)$ is non-singular. That is, almost all groups are regular.*

Proof. The result follows from the transversality theorem in the context book by Guillemin and Pollack (1974). If $D\widetilde{S}(\xi^e, X; W_n)$ has full row-rank, 0 is a regular value for $\widetilde{S}(\cdot, X; W_n)$. Since $\{0\}$ is a singleton in $\Re^n$, $\widetilde{S}(\cdot, X; W_n)$ is transversal to $\{0\}$. From the transversality theorem, $S(\cdot; X, W_n)$ is transveral to $\{0\}$ for almost every X. ■

The following corollary shows that we can apply Proposition C.1 for a large class of models.

Corollary C.1 *Suppose that $u(\cdot)$ is linear in exogenous covariates, i.e., $u(X_i) = \beta_{0,0} + X^{g'}\beta_{0,1} + X_i^{c'}\beta_1$. If $\beta_1 \neq 0$ and $dH_i(a)/da \neq 0$ for any i and $x \in \Re$, $D\widetilde{S}(\xi, X; W_n)$ has full row rank for all W_n. Then almost all groups are regular.*

Proof of Corollary C.1. Without loss of generality, suppose that $\beta_{1,1} \neq 0$, we have that $D\widetilde{S}(\xi, X; W_n) = \begin{pmatrix} \lambda DH - I_n & \beta_{1,1}DH & * \end{pmatrix}$, where $DH = diag(dH_1(\widetilde{x}_1)/dx, \cdots, dH_n(\widetilde{x}_n)/dx)$ is a diagonal matrix whose diagonal is composed of the derivatives of H_i's evaluated at $\widetilde{x}_i = u(X_i) + \lambda \sum_{j \neq i} W_{i,j}\psi_j$ for all i. By the assumption, none of $dH_i(\widetilde{x}_i)/dx$ is zero. Therefore, the rows for $D\widetilde{S}(\xi, X; W_n)$ are linearly independent. So it has full row rank. ■

Although the function $S(\cdot; X, W_n)$ is defined on the whole space, $\Re^n$, we usually begin searching for an equilibrium in a region. According to Guillemin and Pollack (1974), some properties of the set of solutions for $S(\xi; X, W_n) = 0$ inside a region can be derived by analyzing the properties of a function on the boundary of that region. Denote the closed ball in $\Re^n$ with radius $r > 0$ centered at the origin by $B[0, r] = \{x \in \Re^n : \|x\|_E \leq r\}$, where $\|\cdot\|_E$ is the Euclidean norm. Its interior is the open ball, $B(0, r) = \{x \in \Re^n : \|x\|_E < r\}$. Its boundary, $\partial B[0, r] = \{x \in \Re^n : \|x\|_E = r\}$, is a sphere, centering at the origin with radium $r > 0$. In particular, we call the ball a unit sphere if its radius is 1. It is standard to denote a unit sphere in $\Re^n$ as S^{n-1}. If $S(\xi; X, W_n) \neq 0$ for any $\xi \in \partial B[0, r]$, we can define a function, $\widehat{S}(\cdot; X, W_n)$, on $\partial B[0, r]$, as

$$\widehat{S}(\xi; X, W_n)_i = \frac{S(\xi; X, W_n)_i}{\|S(\xi; X, W_n)\|_E} = \frac{H_i(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij}\xi_j) - \xi_i}{\sqrt{\sum_{i=1}^n (H_i(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij}\xi_j) - \xi_i)^2}}, \quad \text{(C.1)}$$

for $i = 1, \cdots, n$. We can see that $\widehat{S}(\cdot; X, W_n)$ maps points on $\partial B[0,r]$ to a point in $S^{n-1} \subseteq \Re^n$. We now associate $\widehat{S}(\cdot; X, W_n)$ to a class of functions via homotopy. We analyze properties of this function through deformation.

Definition C.2 *$\mathcal{C}$ and $\mathcal{D}$ are smooth manifolds in Euclidean spaces.*[22] *Smooth maps, $R_0 : \mathcal{C} \to \mathcal{D}$ and $R_1 : \mathcal{C} \to \mathcal{D}$ are homotopic, if there exists a smooth map, $\widetilde{R} : \mathcal{C} \times [0,1] \to \mathcal{D}$, such that $\widetilde{R}(c,0) = R_0(c)$ and $\widetilde{R}(c,1) = R_1(c)$, for all $c \in \mathcal{C}$, $\widetilde{R}$ is called a homotopy between R_0 and R_1.*

For any given $t \in [0,1]$, if $tH_i(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij}\psi_j) - \psi_i \neq 0$ for all $i = 1, \cdots, n$ and $\xi \in \partial B[0,r]$, by setting

$$R_t(\xi; X, W_n)_i = \frac{tH_i(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij}\xi_j) - \xi_i}{\sqrt{\sum_{i=1}^n (tH_i(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij}\xi_j) - \xi_i)^2}}, \tag{C.2}$$

we can derive a mapping from $\partial B[0,r]$ to S^{n-1}. If that is possible for all $t \in [0,1]$, we get a homotopy, $\widetilde{R}(\cdot,\cdot; X, W_n) : \partial B[0,r] \times [0,1] \to S^1$ such that $\widetilde{R}(\cdot,\cdot; X, W_n) = R_t(\xi; X, W_n)$. In that case, $\widehat{S}(\cdot; X, W_n)$ is homotopic to the function $R_0(\cdot; X, W_n) : \partial B[0,r] \to S^{n-1}$:

$$R_0(\xi; X, W_n)_i = \frac{-\xi_i}{\sqrt{\sum_{i=1}^n (-\xi_i)^2}} = \frac{-\xi_i}{r}, \tag{C.3}$$

for all $i = 1, \cdots, n$. The simplicity of that function makes it convenient to derive properties which are invariant to smooth changes in a homotopy and can be applied to $\widehat{S}(\cdot; X, W_n)$. To make sure that this homotopy is well-defined, we impose Assumption 3.1 about the function, T, as in (3.4). Figure 14 is a graphic illustration of the homotopy constructed for the $H(\cdot)$ function:

$$H(\xi)_i = \frac{\exp(\beta(h + J\xi_j)) - \exp(-(\beta(h + J\xi_j)))}{\exp(\beta(h + J\xi_j)) + \exp(-(\beta(h + J\xi_j)))}, \tag{C.4}$$

for $i, j = 1, 2$ and $i \neq j$, which corresponds to the Binary choice model analyzed by Brock and Durlauf(2007) without imposing rational expectations, $\xi_1 = \xi_2$. We trace out the images of $R_t(\xi; X, W_n)$ for a point in the circle centerer at the original with a radius of 3, when t takes various values and depict them with the same color and marks. We do that for four points, $(r,0)$, $(0,r)$, $(-r,0)$, and $(-r,-r)$. For each of those points, their images in such a homotopy form an interval in the unit circle in a smooth way. We also depict the image of the original function, $S(\cdot; X, W_n)$, at $a(1,0)$, $a(0,1)$, $a(-1,0)$, and $a(-1,-1)$, when a runs from 0 to 1.

[22] Intuitively speaking, a manifold is a subset of an Euclidean space which looks like an open subset of an Euclidean space locally. Any open subsets of an Euclidean space is, of course a manifold. See Guillemin and Pollack (1974) for a rigorous definition.

Utlitizing the homotopy (C.2), we derive some properties of the set of equilibria, which are summarized in Proposition 3.1. It is proved in detail below.

Proof of Proposition 3.1. The proof is composed of two parts. We first show the existence of an equilibrium in the interior, $B(0, r)$ and then prove finiteness. (1) Under Assumption 3.1, when the positive number r is sufficiently large, for any ξ with $\|\xi\|_E = r$, we have that

$$\|S(\xi; X, W_n)\|_E = \|T(\xi; X, W_n) - \xi\|_E \geq |\|T(\xi; X, W_n)\|_E - \|\xi\|_E| > 0.$$

Thus, $S(\xi; X, W_n) \neq 0$ on $\partial B[0, r]$ and $\widehat{S}(\cdot; X, W_n)$ in (C.1) is well-defined on $\partial B[0, r]$. Similarly, for any $t \in [0, 1]$, $\|tT(\xi; X, W_n) - \xi\|_E \geq |\|tT(\xi; X, W_n)\|_E - \|\xi\|_E| > 0$. Thus, $tH_i(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij}\psi_j) - \psi_i \neq 0$, for all $t \in [0, 1]$, $i = 1, \cdots, n$, and all $\xi \in \partial B[0, r]$. Therefore, we can define a homotopy, $\widetilde{R}(\cdot, \cdot; X, W_n) : \partial B[0, r] \times [0, 1] \to S^1$ such that $\widetilde{R}(\cdot, \cdot; X, W_n) = R_t(\xi; X, W_n)$, which is defined in (C.2). $R_0(\xi; X, W_n)$ is a linear transformation from $\partial B[0, r]$ to S^{n-1} with degree $(-1)^n \neq 0$. Therefore, the degree of $\widehat{S}(\cdot; X, W_n) : \partial B[0, r] \to S^{n-1}$ is equal to $(-1)^n \neq 0$. (2) If there is no solution to $S(\xi; X, W_n) = 0$ in the interior, $B(0, r)$, $\widehat{S}(\cdot; X, W_n)$ can be extended to the whole closed ball, $B[0, r]$. According to Guillemin and Pollack (1974), the degree of $\widehat{S}(\cdot; X, W_n)$ on $\partial B[0, r]$ is equal to zero. That is a contradiction. Therefore, there must be at least one point $\xi^e \in IntB[0, r]$ such that $S(\xi^e; X, W_n) = 0$. (3) For any group (X, W_n), the set of equilibria, $\mathfrak{E}(X, W_n)$ is the zeros for the continuous function, $S(\cdot; X, W_n)$. Therefore, it is closed. As a closed subset of the closed ball $B[0, r]$, $\mathfrak{E}(X, W_n)$ is compact. Given regularity, for any $\xi^e \in B[0, r]$, $det(DS(\xi; X, W_n) \neq 0$. By the Inverse Function theorem, $S(\cdot; X, W_n)$ is a local diffeomorphism around ξ. Thus, there is an open neighborhood, $O(\xi^e) \in B[0, r]$ for ξ^e such that $O(\xi^e) \cap \mathfrak{E}(X, W_n) = \{\xi^e\}$. Then the relative open sets, $\{O(\xi^e) \cap \mathfrak{E}(X, W_n)\}$ is an open cover for $\mathfrak{E}(X, W_n)$. By compactness, there is a finite subcover. That is, there is an integer, K, such that $\mathfrak{E}(X, W_n) \subseteq \cup_{k=1}^K O(\xi_k^e) \cap \mathfrak{E}(X, W_n)$. Because each $O(\xi_k^e) \cap \mathfrak{E}(X, W_n) = \{\xi_k^e\}$ is a singleton, $\mathfrak{E}(X, W_n)$ contains just a finite number of points, $\xi_1^e, \cdots, \xi_K^e$. ■

Proof of Proposition 3.2. Pick $B[0, r]$ according to Proposition 3.1, define a function $\widehat{R}(\cdot, \cdot; X, W_n) : B(0, r) \times [0, 1] \to \Re^n$, such that for all $t \in [0, 1]$, $i = 1, \cdots, n$, and $\xi \in \Re^n$,

$$\widehat{R}(\xi, t; X, W_n)_i = tH_i(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij}\xi_j) - \xi_i.$$

We do not lose any zeros of $S(\cdot; X, W_n)$, for there is no zeros for this homotopy on the boundary $\partial B[0, r]$. We can see $\widehat{R}_1(\xi; X, W_n) = \widehat{R}(\xi, 1; X, W_n)$ is the restriction of $S(\cdot; X, W_n)$ in $B(0, r)$.

$\widehat{R}_0(\xi; X, W_n) = \widehat{R}(\xi, 0; X, W_n)$ is just a linear transformation with $\widehat{R}_0(\xi; X, W_n)_i = -\psi_i$ for all $i = 1, \cdots, n$. The restriction, $S(\cdot; X, W_n) : B(0, r) \to \Re^n$ is homotopic to $\widehat{R}_0(\cdot; X, W_n)$. By the Sard's theorem, pick a point $b \in \Re^n$, such that b is a regular value of $\widehat{R}$. Since $\widehat{R}^{-1}(b)$ is a closed set in $B(0, r) \subseteq B[0, r]$, it is compact. Due to transversality, $\widehat{F}^{-1}(b)$ is a one-dimension compact submanifold in $B(0, r)$. Therefore, the sum of the orientation numbers at points in $\partial\widehat{R}^{-1}(b)$ is zero. Since the boundary, $\partial(B(0, r) \times [0, 1]) = (B(0, r) \times \{0\}) \cup (B(0, r) \times \{1\}$, $\partial\widehat{R}$ is equal to $\widehat{R}_0$ on $B(0, r) \times \{0\}$ and $\widehat{R}_1$ on $B(0, r) \times \{1\}$. Therefore, the intersection numbers of those three functions at $\{b\}$ satisfy the following relation:

$$I(\partial\widehat{R}, \{b\}) = I(\widehat{R}_1, \{b\}) - I(\widehat{R}_0, \{b\}).$$

That is, the two homotopic maps, $\widehat{R}_0(\cdot; X, W_n)$ and $\widehat{R}_1(\cdot; X, W_n)$, have the same intersection numbers at $\{b\}$. According to Guillemin and Pollack (1974), since $\Re^n$ is connected and has the same dimension with $B(0, r) \in \Re^n$, the intersection number is invariant with the point picked and is defined as the degree of a function. Therefore, choose point $\{0\}$, we get that

$$deg(\widehat{R}_1) = I(\widehat{R}_1, \{0\}) = I(\widehat{R}_0, \{0\}) = deg(\widehat{R}_0).$$

Since $\widehat{R}_0^{-1}(\{0\}) = \{0\}$ and $det(D\widehat{R}_0) = (-1)^n$, we get that the degree of $\widehat{R}_0$ is also $(-1)^n$, which is equal to $+1$ when n is even and is equal to -1 when n is odd. Since 0 is a regular value for $\widehat{R}_1$ which is the restriction of $S(\cdot; X, W_n)$ on $B(0, r)$, $deg(\widehat{R}_1) = I(\widehat{R}_1, \{0\})$ is actually the sum of orientation numbers for points in $S^{-1}(\cdot; X, W_n)$, which are model equilibrium expectations. Because the orientation numbers of those points are by definition either $+1$ or -1, if their sum is either $+1$ or -1, there must be an odd number of such points. In that case, the equilibria with orientation number $+1$ (-1) outnumbers the equilibria with orientation number -1 by exactly 1. In addition, if the sign of $det(DS(\cdot; X, W_n))$ does not change in $B(0, r)$, all the equilibria will have the same orientation numbers, either $+1$ or -1. If there are more than one equilibria, the absolute value of their sum will be bigger than 1, which contradicts with $deg(\widehat{R}_1(\cdot; X, W_n) = (-1)^n$. Therefore, in that case, there is a unique equilibrium. ■

Proof of Lemma 3.1. By calculation,

$$DS(\xi;X,W_n)=\lambda\Delta_H W-I_n=\lambda\begin{pmatrix}\frac{dH_1(a_1)}{da} & 0 & \cdots & 0\\ 0 & \frac{dH_2(a_2)}{da} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \frac{dH_n(a_n)}{da}\end{pmatrix}W-I_n,$$

where $a_i=u(X_i)+\lambda\sum_{j\neq i}W_{n,ij}\xi_j^e$ for $i=1,\cdots,n$, and I_n is the n-dimension identity matrix. $DS(\xi;X,W_n)$ is an $n\times n$ matrix with all diagonal elements equal to -1, i.e., $DS(\xi;X,W_n)_{i,i}=-1$. All of its off-diagonal elements are equal to their counterparts in $\lambda\Delta_H W$, i.e, $DS(\xi;X,W_n)_{i,j}=(\lambda\Delta_H W)_{i,j}$. By the Gershgorin circle theorem, every eigenvalue of $DS(\xi;X,W_n)$ lies within one of the discs, $B[-1,|\lambda||dH_i(a_i)/da|\sum_{j\neq i}W_{i,j}]$, for $i=1,\cdots,n$.[23] Those circles are all centered at -1. By (2.12), all of their radii are strictly less than 1. Therefore, every real eigenvalues of $DS(\xi;X,W_n)$ is strictly negative. Since the trace of $DS(\xi;X,W_n)$ is real, if τ is one of $DS(\xi;X,W_n)$'s eigenvalues, so be its conjugate. But their product is $\tau\overline{\tau}>0$. Let $2k$ denote the number of complex eigenvalues. The sign of the product of all eigenvalues, which is equal to the sign of the determinant, is equal to $(-1)^{n-2k}=(-1)^n$. Therefore, $sgn(det(DS(\xi;X,W_n)))=(-1)^n$ for all $\xi\in\Re^n$. ∎

D Proofs for Identification with Public Characteristics

Proof of Proposition 3.3. With public information on all exogenous covariates, simply denote $(X^g,X^c)=X$. By calculation, $E[Y|X]=E[(I_n-\lambda\overline{W}_n)^{-1}X\beta|X]$. Therefore, if (β,λ,σ) and $(\widetilde{\beta},\widetilde{\lambda},\widetilde{\sigma})$ are observationally equivalent,

$$E[(I_n-\lambda\overline{W}_n)^{-1}X\beta-(I_n-\widetilde{\lambda}\overline{W}_n)^{-1}X\widetilde{\beta}|X]=0.$$

for any X in its support. Multiply both sides by the non-random matrix, $(I_n-\lambda\overline{W}_n)(I_n-\widetilde{\lambda}\overline{W}_n)$. Notice that $(I_n-\lambda\overline{W}_n)(I_n-\widetilde{\lambda}\overline{W}_n)=(I_n-\widetilde{\lambda}\overline{W}_n)(I_n-\lambda\overline{W}_n)$, we have that

$$E[(I_n-\widetilde{\lambda}\overline{W}_n)X\beta-(I_n-\lambda\overline{W}_n)X\widetilde{\beta}|X]=0.$$

[23] Let $A=(a_{ij})$ be an $n\times n$ matrix. Denote by $R_i=\sum_{j\neq i}|a_{i,j}|$ the sum of absolute values of off-diagonal elements in the i-th row. Denote the closed disc centered at a_{ii} with a radius R_i by $B[a_{ii},R_i]$. By the Gershgorin circle theorem, every eigenvalue of A lies within one of those closed discs, $B[a_{ii},R_i]$, for $i=1,\cdots,n$. A brief explanation and proof for this theorem can be found at http://en.wikipedia.org.

Denote by l_n the $n \times 1$ vector of 1's,

$$E\Big[\Big(l_n \quad \overline{W}_n l_n \quad X^c \quad \overline{W}_n X^c\Big)\Big(\beta_{0,0} - \widetilde{\beta}_{0,0} \quad \lambda\widetilde{\beta}_{0,0} - \widetilde{\lambda}\beta_{0,0} \quad \beta_1' - \widetilde{\beta}_1' \quad \lambda\widetilde{\beta}_1' - \widetilde{\lambda}\beta_1'\Big)' |X\Big] = 0,$$

which is equivalent to

$$E\Big[\Big(l_n \quad \overline{W}_n l_n \quad X^c \quad \overline{W}_n X^c\Big)' \Big(l_n \quad \overline{W}_n l_n \quad X^c \quad \overline{W}_n X^c\Big) |X\Big]$$
$$\Big(\beta_{0,0} - \widetilde{\beta}_{0,0} \quad \lambda\widetilde{\beta}_{0,0} - \widetilde{\lambda}\beta_{0,0} \quad \beta_1' - \widetilde{\beta}_1' \quad \lambda\widetilde{\beta}_1' - \widetilde{\lambda}\beta_1'\Big)' = 0.$$

Taking expectations over X, we get that

$$E\Big[\Big(l_n \quad \overline{W}_n l_n \quad X^c \quad \overline{W}_n X^c\Big)' \Big(l_n \quad \overline{W}_n l_n \quad X^c \quad \overline{W}_n X^c\Big)\Big]$$
$$\Big(\beta_{0,0} - \widetilde{\beta}_{0,0} \quad \lambda\widetilde{\beta}_{0,0} - \widetilde{\lambda}\beta_{0,0} \quad \beta_1' - \widetilde{\beta}_1' \quad \lambda\widetilde{\beta}_1' - \widetilde{\lambda}\beta_1'\Big)' = 0.$$

Under assumption (3.12), $E\Big[\Big(l_n \quad \overline{W}_n l_n \quad X^c \quad \overline{W}_n X^c\Big)' \Big(l_n \quad \overline{W}_n l_n \quad X^c \quad \overline{W}_n X^c\Big)\Big]$ is positive definite. Then $\Big(\beta_{0,0} - \widetilde{\beta}_{0,0} \quad \lambda\widetilde{\beta}_{0,0} - \widetilde{\lambda}\beta_{0,0} \quad \beta_1' - \widetilde{\beta}_1' \quad \lambda\widetilde{\beta}_1' - \widetilde{\lambda}\beta_1'\Big)'$ implies that $\beta = \widetilde{\beta}$ and $\lambda = \widetilde{\lambda}$. If $\overline{W}_n$ is row-normalized, $\overline{W}_n l_n = l_n$. Then observationally equivalence implies that

$$E\Big[\Big(l_n \quad X^c \quad \overline{W}_n X^c\Big)' \Big(l_n \quad X^c \quad \overline{W}_n X^c\Big)\Big]$$
$$\Big(\beta_{0,0} - \widetilde{\beta}_{0,0} + \lambda\widetilde{\beta}_{0,0} - \widetilde{\lambda}\beta_{0,0} \quad \beta_1' - \widetilde{\beta}_1' \quad \lambda\widetilde{\beta}_1' - \widetilde{\lambda}\beta_1'\Big)' = 0.$$

If (3.12′) holds, we will also have $\beta = \widetilde{\beta}$ and $\lambda = \widetilde{\lambda}$. If $\overline{W}_n$ is row-normalized, $\overline{W}_n l_n = l_n$. Because $E[(Y - E[Y|X])(Y - E[Y|X])'|X] = \sigma^2 I_n$. We can identify σ through conditional variance of y_i's given X. ■

Proof of Lemma 3.2. Without loss of generality, suppose that $\beta_{1,1} > 0$. For $\omega_{-i} \in \{0,1\}^{n-1}$, choose $\mathcal{X}(\omega)^c$ as

$$\mathcal{X}(\omega)^c = \left\{X^c \in \Re^{nL} : X^c_{j,1}(2\omega_j - 1) \geq 0, j \neq i\right\}.$$

We can see that for any $j \neq i$, as $|X^c_{j,1}| \to \infty$ in $\mathcal{X}^c(\omega) \subseteq \Re^{nL}$, with the restriction, $X^c_{j,1}(2\omega_j - 1) \geq 0$, $u(X_j)$ goes to $+\infty$ when ω_j is 1; and $u(X_j)$ goes to $-\infty$ when ω_j is 0. Therefore, $\lim_{|X^c_{j,1}| \to \infty, j \neq i, X^c \in \mathcal{X}^c(\omega)} P(y_{-i} = \omega|X^c) = 1$. Since all X^c_i's have full support, $P(X^c \in \mathcal{X}^c(\omega)) > 0$. ■

Proof of Proposition 3.4. By Lemma 3.2,

$$\lim_{|X^c_{j,1}|\to\infty, j\neq i, X^c\in\mathcal{X}^c(\omega_0)_{-i}} P(\beta_{0,0} + X_i^{c'}\beta_1 + \lambda\sum_{j\neq i}\overline{W}_{n,ij}\psi_j - \epsilon_i > 0|X^c)$$
$$= \lim_{|X^c_{j,1}|\to\infty, j\neq i, X^c\in\mathcal{X}^c(\omega_0)_{-i}} P(\beta_{0,0} + X_i^{c'}\beta_1 + \lambda\sum_{j\neq i}\overline{W}_{n,ij}\omega_{0,j} - \epsilon_i > 0|X^c).$$

Therefore,

$$\lim_{|X^c_{j,1}|\to\infty, j\neq i, X^c\in\mathcal{X}^c(\omega_0)} \log P(y_i, \omega_0|X^c)$$
$$= \lim_{|X^c_{j,1}|\to\infty, j\neq i, X^c\in\mathcal{X}^c(\omega_0)} y_i \log F_\epsilon(\beta_{0,0} + X_i^{c'}\beta_1 + \lambda\sum_{j\neq i}\overline{W}_{n,ij}\omega_{0,j})$$
$$+ (1-y_i)\log(1 - F_\epsilon(\beta_{0,0} + X_i^{c'}\beta_1 + \lambda\sum_{j\neq i}\overline{W}_{n,ij}\omega_{0,j})).$$

Under condition (3.13), when $X^c \in \mathcal{X}^c(\omega_0)$ and $|X^c_{j,1}| \geq D$, for all $i \neq j$,
$E[(\partial\log P(y_i, \omega_0|X^c)/\partial\theta)(\partial\log P(y_i, \omega_0|X^c)/\partial\theta)^{'}|X^c \in \mathcal{X}^c(\omega_0), |X^c_{j,1}| \geq D, i \neq j]$
is positive definite. From Rothenberg(1971), $\theta = (\beta^{'}, \lambda)^{'}$ can be identified. ■

Proof of Lemma 3.3. Similar to the proof of Lemma3.2, choose

$$\mathcal{X}^c_1 = \left\{X^c \in \Re^{nL} : X^c_{j,1} \geq 0, 1 \leq j \leq n\right\}.$$

In this set, as $\|X^c_{i,1}\|_E \to \infty$, $X^c_{i,1}\beta_{1,1} \to +\infty$. As $u(X_i) = \beta_{0,0} + X_i^{c'}\beta_1 \to +\infty$, in the limit, none outcomes are censored. ■

Proof of Proposition 3.5. It follows from Lemma3.3 that in $\mathcal{X}^c_1$, as $|X^c_{i,1}| \to \infty$ in this set, no choices are censored. Therefore, for the distribution of observed outcomes we have that

$$\lim_{|X^c_{j,1}|\to\infty, 1\leq j\leq n, X^c\in\mathcal{X}^c_1} f(Y|X^c) = \lim_{|X^c_{j,1}|\to\infty, 1\leq j\leq n, X^c\in\mathcal{X}^c_1} f(Y^*|X^c).$$

That is, in the limit, the observed outcomes are the latent variables which are associated with each other just as continuous choices in linear models. Since $E[|y_i||X^c] = H_i(u(X_i) + \lambda\sum_{j\neq i}\overline{W}_{n,ij}\xi^e_j) < \infty$ and $E[|y_i^*||X^c] = E[y_i|X^c]$, by the Lebesgue Control convergence theorem,

$$\lim_{|X^c_{j,1}|\to\infty, 1\leq j\leq n, X^c\in\mathcal{X}^c_1} E[Y|X^c] = \lim_{|X^c_{j,1}|\to\infty, 1\leq j\leq n, X^c\in\mathcal{X}^c_1} E[Y^*|X^c]$$
$$= \lim_{|X^c_{j,1}|\to\infty, 1\leq j\leq n, X^c\in\mathcal{X}^c_1} (I_n - \lambda\overline{W}_n)^{-1}(\beta_{0,0}l_n + X^c\beta_1).$$

Thus, if (β, λ, σ) and $(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\sigma})$ are observationally equivalent,

$$\lim_{|X^c_{j,1}|\to\infty, 1\le j\le n, X^c\in\mathcal{X}^c_1} E\Big[\begin{pmatrix} l_n & \overline{W}_n l_n & X^c & \overline{W}_n X^c \end{pmatrix} \begin{pmatrix} \beta_{0,0}-\widetilde{\beta}_{0,0} & \lambda\widetilde{\beta}_{0,0}-\widetilde{\lambda}\beta_{0,0} & \beta_1'-\widetilde{\beta}_1' & \lambda\widetilde{\beta}_1'-\widetilde{\lambda}\beta_1' \end{pmatrix}' |X^c\Big] = 0,$$

for all $X^c \in \mathcal{X}^c_1$. Similar to the proof of Proposition 3.3, we derive that

$$\lim_{|X^c_{j,1}|\to\infty, 1\le j\le n, X^c\in\mathcal{X}^c_1} E\Big[\begin{pmatrix} l_n & \overline{W}_n l_n & X^c & \overline{W}_n X^c \end{pmatrix}' \begin{pmatrix} l_n & \overline{W}_n l_n & X^c & \overline{W}_n X^c \end{pmatrix}\Big] \begin{pmatrix} \beta_{0,0}-\widetilde{\beta}_{0,0} & \lambda\widetilde{\beta}_{0,0}-\widetilde{\lambda}\beta_{0,0} & \beta_1'-\widetilde{\beta}_1' & \lambda\widetilde{\beta}_1'-\widetilde{\lambda}\beta_1' \end{pmatrix}' = 0.$$

Under (3.15), there is a positive-measure subset of covariates such that
$E\Big[\begin{pmatrix} l_n & \overline{W}_n l_n & X^c & \overline{W}_n X^c \end{pmatrix}' \begin{pmatrix} l_n & \overline{W}_n l_n & X^c & \overline{W}_n X^c \end{pmatrix}\Big]$
is positive definite. Therefore, $(\beta, \lambda) = (\widetilde{\beta}, \widetilde{\lambda})$. Identification of σ follows from Yang, Qu, and Lee (2016). ■

Proof of Proposition 3.6. For a group $(X, \overline{W}_n)$, the sample log likelihood can be written as $\log \sum_{\xi^e} \rho(\alpha, \xi^e) f(Y|\xi^e)$.

By calculation, we get that

$$\begin{aligned} & E[\frac{\partial \log L(Y|X^c, \overline{W}_n)}{\partial \alpha} \frac{\partial \log L(Y|X^c, \overline{W}_n)}{\partial \alpha'}|X^c] \\ =& E[(\frac{1}{\sum_{\xi^e} \rho(\xi^e; \mathfrak{E}(X, W_n), \alpha) f(Y|\xi^e)})^4 (\frac{1}{\sum_{\xi^e} \exp(\alpha'\gamma(\xi^e; X, \overline{W}_n))})^2 \\ & (\frac{1}{\sum_{\xi^e} \exp(\alpha'\gamma(\xi^e; X, \overline{W}_n) f(Y|\xi^e))})^2 \Gamma'(X^c, \overline{W}_n; \beta, \sigma, \lambda)(D(X, \overline{W}_n))^2 \Gamma(X^c, \overline{W}_n; \beta, \sigma, \lambda)|X^c]. \end{aligned}$$

When there are M equilibria, $D(X, \overline{W}_n)$ is a $M \times M$ diagonal matrix. Its (m, m) element is $\frac{\exp(\alpha'\gamma(\xi^e; X, \overline{W}_n)) f(y|\xi^e)}{\sum_{\widetilde{\xi}^e} \exp(\alpha'\gamma(\widetilde{\xi}^e; X, \overline{W}_n)) f(y|\widetilde{\xi}^e)} - \frac{\exp(\alpha'\gamma(\xi^e; X, \overline{W}_n))}{\sum_{\widetilde{\xi}^e} \exp(\alpha'\gamma(\widetilde{\xi}^e; X, \overline{W}_n))}$. We can see that as long as there are multiple equilibria, $D(X, \overline{W}_n)$ is positive definite. If $E[\Gamma'(X^c, \overline{W}_n; \beta, \sigma, \lambda)\Gamma(X^c, \overline{W}_n; \beta, \sigma, \lambda)|X^c]$ has full column rank, so does $E[\frac{\partial \log L(Y|X^c, \overline{W}_n)}{\partial \alpha} \frac{\partial \log L(Y|X^c, \overline{W}_n)}{\partial \alpha'}|X^c]$. From Rothenberg(1971), α can be identified. ■

E Equilibrium with Privately Known Characteristics

In this section, we discuss the existence and property of equilibria when some exogenous characteristics are private information. The case in Section 4 is one special case. In the Banach space, $(\Xi(W_n, \mathcal{J}), \|\cdot\|)$. define an operator, $T : \Xi(W_n, \mathcal{J}) \to \Xi(W_n, \mathcal{J})$, such that for all $i = 1, \cdots, n$

and $m = 1, \cdots, M_i$,

$$T(\xi)(x^p_{\widetilde{J}_{1,1}}, \cdots, x^p_{\widetilde{J}_{1,M_1}}, \cdots, x^p_{\widetilde{J}_{n,1}}, \cdots, x^p_{\widetilde{J}_{n,M_n}})_{i,m} = E[H_i(u(X_i) + \lambda \sum_{j \neq i} W_{n,ij}\xi_{j,J_i}(X^p_{J_i}))|x^p_{\widetilde{J}_{i,m}}, z]. \quad \text{(E.1)}$$

An equilibrium conditional expectation function, ξ^e, corresponds to one of T's fixed points. To apply the Schauder fixed point theorem (Proposition E.4), we need to capture a compact set in the function space, $(\Xi(W_n, \mathcal{J}), \|\cdot\|)$. Analogous to conventional discussions on Banach spaces, we focus on a weaker condition, relative compactness.

Definition E.1 *A set in a metric space is relatively compact if its closure is compact.*

Simon (1987) introduces a property about relative compact sets, which is used for our proof. it is cited as the lemma below.

Lemma E.1 *A set V is a normed space U is relatively compact if and only if for any $\eta > 0$, there are a finite subset $\{v_1, \cdots, v_L\} \subseteq V$ such that for any $v \in V$, there is v_i for some $i = 1, \cdots, n$ with $\|v - v_i\|_U < \eta$.*

Thus, as long as we find a set which is relatively compact, all of its own points and its limit points form a new set which is compact. Because $\xi \in (\Xi(W_n, \mathcal{J}), \|\cdot\|)$ if and only if each of its coordinate functions, $\xi_{i,m}$ is a function in the Lebesgue space, $L^1(\mathfrak{X}^p_{i,m}, \mathfrak{B}_{i,m}, \mu_p; \Re^1)$, we may utilize the properties of a relatively compact set in $L^1(\mathfrak{X}^p_{i,m}, \mathfrak{B}_{i,m}, \mu_p; \Re^1)$. That is possible due to the following lemma.

Lemma E.2 Γ_0 *is a relatively compact subset of* $(\Xi(W_n, \mathcal{J}), \|\cdot\|)$ *if and only if*
$\Gamma_{0,i,m} = \left\{\xi_{i,m} : \mathfrak{X}^p_{i,m} \to \Re^1 : (\xi_{i,m}, \xi_{-im}) \in \Gamma_0 \text{ for some } \xi_{-im} : \prod_{(i',m') \neq (i,m)} \mathfrak{X}^p_{i',m'} \to \Re^{M-1}\right\}$
is relatively compact in $L^1(\mathfrak{X}_{i,m}, \mathfrak{B}_{i,m}, \mu_p; \Re^1)$ for all $i = 1, \cdots, n$, $m = 1, \cdots, M_i$.

Proof. Suppose that $\Gamma_{0,i,m}$ is relatively compact in $L^1(\mathfrak{X}^p_{i,m}, \mathfrak{B}_{i,m}, \mu_p; \Re^1)$ for each (i,m). By Lemma E.1, for an arbitrarily chosen $\eta > 0$, for any (i,m), there is $(\xi_{i,m,1}, \cdots, \xi_{i,m,L_{i,m}})$ in $\Gamma_{0,i,m}$, such that for any $\xi_{i,m} \in \Gamma_{0,i,m}$, there is $\xi_{i,m,l}$ for some $1 \leq l \leq L_{i,m}$ with $\|\xi_i - \xi_{i,l}\|_1 < \eta$. Construct a finite subset of Γ_0 as
$\Gamma_{0b} = \{(\xi_{1,1,l_1}, \cdots, \xi_{n,m_n,l_n}) : 1 \leq l_i \leq L_{i,m} \text{ for all } i = 1, \cdots, n, m = 1, \cdots, M_i\}$.
Then for any $\xi = (\xi_{1,1}, \cdots, \xi_{n,M_n}) \in \Gamma_0$, for each (i,m), pick $l_{i,m}$ such that $\|\xi_{i,m} - \xi_{i,m,l_{i,m}}\|_1 = \int |\xi_{i,m} - \xi_{i,m,l_{i,m}}| d\mu_p < \eta$. Denoting $\xi^b = (\xi_{1,1,l_{1,1}}, \cdots, \xi_{n,M_n,l_{n,M_n}}) \in \Gamma_{0b}$, we have that $\|\xi -$

$\xi^b\| = \max_{1\le i\le n}\max_{1\le m\le M_i}\|\xi_{i,m} - \xi_{i,m,l_{i,m}}\|_1 < \eta$. Therefore, Γ_0 is relatively compact in $(\Xi(W_n,\mathcal{J}),\|\cdot\|)$. On the contrary, suppose that Γ_0 is relatively compact in $(\Xi(W_n,\mathcal{J}),\|\cdot\|)$. If for some (i_0,m_0), Γ_{0,i_0,m_0} is not relatively compact in $L^1(\mathfrak{X}^p_{i_0,m_0},\mathfrak{B}_{i_0,m_0},\mu_p;\Re^1)$, there is $\eta_0>0$, such that for any finite subset of Γ_{0,i_0,m_0}, $\left\{\xi_{i_0,m_0,1},\cdots,\xi_{i_0,m_0,L_{i_0,m_0}}\right\}$, there is $\xi^*_{i_0,m_0}\in\Gamma_{0,i_0,m_0}$ with $\|\xi^*_{i_0,m_0}-\xi_{i_0,m_0,l}\|_1>\eta$ for all $1\le l\le L_{i_0,m_0}$. For any finite subset of Γ_0, $\{\xi^1,\cdots,\xi^L\}$, $\left\{\xi^1_{i_0,m_0},\cdots,\xi^L_{i_0,m_0}\right\}$ is a finite subset of Γ_{0,i_0,m_0}. Take $\xi^*=(\xi^*_{i_0,m_0},\xi^{-i_0m_0})\in\Gamma_0$. We have that $\|\xi^*-\xi^l\|\ge\|\xi^*_{i_0,m_0}-\xi^l_{i_0,m_0}\|_1>\eta_0$, contradicting with Γ_0 being relatively compact. Therefore, each $\Gamma_{0,i,m}$ is relatively compact in $L^1(\mathfrak{X}^p_{i,m},\mathfrak{B}_{i,m},\mu_p;\Re^1)$ for all (i,m). ■

Owing to Lemma E.2, to characterize relatively compact sets in $(\Xi(W_n,\mathcal{J}),\|\cdot\|)$, we need to capture the relatively compact sets in the Lebesgue space $L^1(\mathfrak{X}_{\widetilde{J}_{i,m}},\Sigma_{i,m},\mu_p;\Re^1)$. There are some characterizations for Lebesgue spaces of functions whose domains are general measurable spaces and ranges are Banach spaces, such as the results by Brooks and Dinculeanu(1979) and more recently, the Diaz-Mayoral Theorem (See van Neerven(2014) for an elementary proof). Here, we apply the classical results by Dunford and Schwartz(1958).

If $\Omega=\Re^1$, $\mathfrak{B}_R$ is the Borel σ-algebra, μ is the Lebesgue measure, $\mathfrak{m}$, and $\mathfrak{Y}$ is a Banach space with norm $\|\cdot\|_Y$. Dunford and Schwartz (1958) have a characterization for a relatively compact set in the Lebesgue space, $L^q(\Re^1,\mathfrak{B}_R,\mathfrak{m};\mathfrak{Y})$, the space consisting of mappings from $\Re^1$ to $\mathfrak{Y}$, integrable under $\mathfrak{m}$, with the L^q norm.

Proposition E.1 *For $1\le q<\infty$, $\Upsilon_0\subseteq L^q(\Re^1,\mathfrak{B}_R,\mathfrak{m};\mathfrak{Y})$ is relatively compact if and only if:*

1. *It is bounded, i.e., $\sup_{\chi\in\Upsilon_0}(\int_{-\infty}^{+\infty}\|\chi(t)\|^q_Y dt)^{1/q}<\overline{U}$ for some $\overline{U}>0$;*
2. *$\int_{-\infty}^{+\infty}\|\chi(t+s)-\chi(t)\|^q_Y dt\to 0$ as $s\to 0$ uniformly for all $\chi\in\Upsilon_0$; and*
3. *$(\int_r^{+\infty}+\int_{-\infty}^{-r})\|\chi(t)\|^q_Y dt\to 0$ as $r\to\infty$ uniformly for all $\chi\in\Upsilon_0$.*

For $\Omega=[a,b]$, (1) and (2) and necessary and sufficient conditions for relative compactness.

Proof. See Dunford and Schwartz(1958) Theorem IV.8.20 (pp.298). ■

The results above can be extended to the n-dimension Euclidean space.

Proposition E.2 *For $1\le q<\infty$, $\Upsilon_0\subseteq L^q(\Re^n,\mathfrak{B},\mathfrak{m};\mathfrak{Y})$ is relatively compact if and only if:*

1. *It is bounded, i.e., $\sup_{\chi\in\Upsilon_0}(\int_{\Re^n}\|\chi(t)\|^q_Y dt)^{1/q}<\overline{U}$ for some $\overline{U}>0$;*

2. $\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \|\chi(t_1+s_1,\cdots,t_n+s_n) - \chi(t_1,\cdots,t_n)|_Y^q dt_1 \cdots dt_n \to 0$, *as* $s = (s_1, \cdots, s_n) \to 0$ *uniformly for all* $\chi \in \Upsilon_0$*; and*

3. $\int_{\Re^n - C_r} \|\chi(t)\|_Y^q dt \to 0$ *as* $r \to \infty$ *uniformly for all* $\chi \in \Upsilon_0$*, where the cube* $C_r = \{t = (t_1, \cdots, t_n) \in \Re^n : -r \le t_i \le r \ \forall i = 1, \cdots, n\}$.

If $\Omega = \prod_{i=1}^n [a_i, b_i]$*, (1) and (2) are necessary and sufficient for relative compactness.*

Proof. See Dunford and Schwartz(1958) Theorem IV.8.21 (pp.301). ■

Although those results are very general, they are about Lebesgue measure. In our model, particularly, each coordinate function ξ_i for $\xi \in \Xi$ is defined on a subset of the Euclidean space with the probability measure, μ_p, which is induced by the distribution of X_i^p's conditional on the public information $Z = z$. We show that when there is a pdf for this conditional distribution, we can apply Proposition E.1 and Proposition E.2.

Lemma E.3 *Let* Ω *be a subset of* $\Re^n$*. When* $\mu << \mathfrak{m}$*,* $d\mu/d\mathfrak{m} = f$ *is a strictly positive on* Ω*,* $\Upsilon_0 \in L^1(\Omega, \mathfrak{B}_\Omega, \mu; \mathfrak{Y})$ *is relatively compact if and only if* $f\Upsilon_0 = \{f\chi : \chi \in \Upsilon_0\}$ *is relatively compact in* $L^1(\Omega, \Sigma_B, \mathfrak{m}; \mathfrak{B})$.

Proof. On one hand, if $f\Upsilon_0$ is relatively compact in $L^1(\Omega, \Sigma_B, \mathfrak{m}; \mathfrak{B})$, by Lemma E.1, for any $\eta > 0$, there are a finite subset, $\{\chi^1, \cdots, \chi^K\}$ in $L^1(\Omega, \mathfrak{B}_\Omega, \mathfrak{m}; \mathfrak{Y})$, such that for any $\chi \in L^1(\Omega, \mathfrak{B}_\Omega, \mathfrak{m}; \mathfrak{Y})$, there is χ^k in that finite set with $\int_\Omega \|\chi - \chi^k\|_Y d\mathfrak{m} < \eta$. $\{\chi^1/f, \cdots, \chi^K/f\}$ is a finite subset of $L^1(\Omega, \mathfrak{B}_\Omega, \mu; \mathfrak{Y})$. For any $\widetilde{\chi} \in \Upsilon_0$, $f\widetilde{\chi} \in f\Upsilon_0$. Therefore, $\int_\Omega \|f\widetilde{\chi} - \chi^l\|_Y d\mathfrak{m} = \int_\Omega \|\widetilde{\chi} - \chi^l/f\|_Y f d\mathfrak{m} = \int_\Omega \|\widetilde{\chi} - \chi^l/f\|_Y d\mu < \eta$. Therefore, Υ_0 is relatively compact in $L^1(\Omega, \mathfrak{B}_\Omega, \mu; \mathfrak{Y})$. On the other hand, if Υ_0 is relatively compact, for any $\eta > 0$, there is a finite subset $\{\widetilde{\chi}^1, \cdots, \widetilde{\chi}^K\}$ such that for any $\widetilde{\chi} \in \Upsilon_0$, there is a function $\widetilde{\chi}^k$ in that finite subset such that $\int_\Omega \|\widetilde{\chi} - \widetilde{\chi}^k\|_Y d\mu < \eta$. Note that $\{f\widetilde{\chi}^1, \cdots, f\widetilde{\chi}^K\}$ is a finite set in $L^1(\Omega, \mathfrak{B}_\Omega, \mathfrak{m}; \mathfrak{Y})$. Take $\chi \in L^1(\Omega, \mathfrak{B}_\Omega, \mathfrak{m}; \mathfrak{Y})$, $\chi/f \in L^1(\Omega, \mathfrak{B}_\Omega, \mu; \mathfrak{Y})$. Hence, we have that $\int_\Omega \|\chi - f\widetilde{\chi}^k\|_Y d\mathfrak{m} = \int_\Omega \|\chi/f - \widetilde{\chi}^k\|_Y f d\mathfrak{m} = \int_\Omega \|\chi/f - \widetilde{\chi}^k\|_Y d\mu < \eta$. Therefore, $f\Upsilon_0$ is relatively compact in $L^1(\Omega, \mathfrak{B}_\Omega, \mathfrak{m}; \mathfrak{Y})$. ■

Corollary E.1 *Suppose that* $\Omega = \prod_{i=1}^n [a_i, b_i]$ *for* $-\infty \le a_i < b_i \le +\infty$*.* $\mu << \mathfrak{m}$*.* $d\mu/d\mathfrak{m} = f$ *is strictly positive on* Ω*. When* $|a_i| < \infty$ *and* $|b_i| < \infty$ *for all* $i = 1, \cdots, n$*,* $\Upsilon_0 \in L^1(\Omega, \mathfrak{B}_\Omega, \mu; \mathfrak{Y})$ *is relatively compact if and only if:*

1. $\sup_{\chi \in \Upsilon_0} \int_\Omega \|\chi(t)\|_B dt < \overline{U}$ *for some* $\overline{U} > 0$*;*

2.

$$\int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} |\chi(t_1+s_1,\cdots,t_n+s_n)f(t_1+s_1,\cdots,t_n+s_n) - \chi(t_1,\cdots,t_n)f(t_1,\cdots,t_n)|dt_1\cdots dt_n \to 0,$$

as $s = (s_1, \cdots, s_n) \to 0$*, uniformly for all* $\chi \in \Upsilon_0$*.*

When $|a_i| = \infty$ *for some* i *or* $|b_j| = \infty$ *for some* j*, the necessary and sufficient conditions for* $\Upsilon_0 \in L^1(\Omega, \mathfrak{B}_\Omega, \mu; \mathfrak{Y})$ *to be relatively compact include both (1) and (2), as well as (3):* $\int_{\omega - C_r} \|\chi(t_1,\cdots,t_n)\|_Y f(t_1,\cdots,t_n)dt_1\cdots dt_n \to 0$*, as* $r \to \infty$ *uniformly for all* $\chi \in \Upsilon_0$*, where the cube* $C_r = \{t = (t_1,\cdots,t_n) \in \Omega : -r \le t_i \le r\ \forall i = 1,\cdots,n\}$*.*

Combining Lemma E.2 and Corollary E.1, we derive a characterization of a relatively compact subset of $(\Xi(W_n,\mathcal{J}), \|\cdot\|)$ using properties of functions.

Proposition E.3 *Suppose that conditional on public information* $Z = z$*, the support of* X_i^p*,* $\mathfrak{X}_i^p$*, is a cube in* $\Re^{k_p}$ *(It can be bounded or unbounded), and the joint distribution of* X_i^p*'s has a pdf* $f_p(\cdot)$*.*

1. *When all the* $\mathfrak{X}_i^p$*'s are bounded,* Γ_0 *is relatively compact in* $(\Xi(W_n,\mathcal{J}), \|\cdot\|)$*, if and only if,*
   - *(a) (uniformly bounded) there is a real number* $\overline{B} > 0$ *such that* $\sup_{\xi\in\Gamma_0}\|\xi\| = \sup_{\xi\in\Gamma_0}\max_{1\le i\le n}\max_{1\le m\le M_i}\int_{\mathfrak{X}_{i,m}^p}|\xi_{i,m}(x)|f_p(x)dx \le \overline{B}$*;*
   - *(b)* $\max_{1\le i\le n}\max_{1\le m\le M_i}\int_{\mathfrak{X}_{i,m}^p}|\xi_{i,m}(x+\widetilde{x})f_p(x+\widetilde{x}) - \xi_{i,m}(x)f_p(x)|dx \to 0$ *as* $\widetilde{x} \to 0$ *uniformly for any* $\xi \in \Gamma_0$*.*

2. *When some of the support is unbounded, the necessary and sufficient conditions for* Γ_0 *to be relatively compact in* $(\Xi(W_n,\mathcal{J}), \|\cdot\|)$ *include (1) and (2) as well as (3):* $\max_{1\le i\le n}\max_{1\le m\le M_i}\int_{\mathfrak{X}_{i,m}^p - C_{r,i,m}}|\xi_{i,m}(x)|dx \to 0$ *as* $r \to \infty$ *uniformly for all* $\xi \in \Upsilon_0$*, where the cube* $C_{r,i,m} = \left\{x \in \mathfrak{X}_{i,m}^p : |x_j| \le r\ \forall j \text{ with } J_{i,m}(j) = 1\right\}$*.*

Proof. (a) Suppose that Γ_0 is relatively compact in $(\Xi(W_n,\mathcal{J}), \|\cdot\|)$. Then for each i and m, the set, $\Gamma_{0,i,m} = \{\xi_{i,m} : (\xi_{i,m}, \xi_{-im}) \in \Gamma_0 \text{ for some } \xi_{-im}\}$

is relatively compact in $L^1(\mathfrak{X}^p_{i,m}, \mathfrak{B}_{i,m}, \mu_p; \Re^1)$ according to LemmaE.2. It follows from Corollary E.1 that for any i and m, there is $\overline{U}_{i,m}$, such that $\sup_{\xi_{i,m}\Gamma_{0,i,m}} \int_{\mathfrak{X}^p_{i,m}} |\xi_i(x)|f_p(x)dx < \overline{U}_{i,m}$. Therefore,

$$\begin{aligned}\sup_{\xi\in\Gamma_0} \|\xi\| =& \sup_{\xi\in\Gamma_0} \max_{1\leq i\leq n} \max_{1\leq m\leq M_i} \int_{\mathfrak{X}^p_{i,m}} |\xi_{i,m}(x)|f_p(x)dx \\ \leq& \max_{1\leq i\leq n} \max_{1\leq m\leq M_i} \overline{U}_{i,m} \\ =&\overline{U},\end{aligned}$$

where $\overline{U} = \max_{1\leq i\leq n} \max_{1\leq m\leq M_i} \overline{U}_{i,m} < \infty$. Additionally, for any $\eta > 0$, there is $\delta_{i,m} > 0$, such that $\int_{\mathfrak{X}^p_{i,m}} |\xi(x + \widetilde{x}^{i,m})f(x + \widetilde{x}^{i,m} - \xi(x)f(x)|dx < \eta$ when $\|\widetilde{x}^{i,m}\|_E < \delta_{i,m}$ for all $\xi \in \Gamma_0$. Take $\delta = \min 1 \leq i \leq n \max_{1\leq m\leq M_i} \delta_{i,m} > 0$. For a vector
$\widetilde{x} = (\widetilde{x}^{1,1'}, \cdots, \widetilde{x}^{1,M_1'}, \cdots, \widetilde{x}^{n,1'}, \cdots, \widetilde{x}^{n,M_n'})'$, when $\|\widetilde{x}\|_E < \delta$, $\|\widetilde{x}^{i,m}\|_E < \delta_{i,m}$ for all (i,m). Then $\max_{1\leq i\leq n} \max_{1\leq m\leq M_i} \int_{\mathfrak{X}^p_{i,m}} |\xi_{i,m}(x + \widetilde{x}^{i,m})f(x + \widetilde{x}^{i,m}) - \xi_{i,m}(x)f(x)|dx < \eta$ for all $\xi \in \Gamma_0$. When some of the X_i^p's have an unbounded support, for any $\eta > 0$, there is $R_{i,m} > 0$, such that for all (i,m), $\int_{\mathfrak{X}^p_{i,m}-C_{r,i,m}} |\chi(x)|dx < \eta$ for all $r > R$ and all $\xi_{i,m}$. Take $R = \max_{1\leq i\leq n} \max_{1\leq m\leq M_i} R_{i,m} < +\infty$. Then when $r > R$,
$\max_{1\leq i\leq n} \max_{1\leq m\leq M_i} \int_{\mathfrak{X}^p_{i,m}-C_{r,i,m}} |\xi_{i,m}(x)|dx < \eta$ for all $\xi \in \Upsilon_0$.
(b) One the contrary, if (1) and (2) hold, suppose that there is (i_0, m_0), such that for any $\overline{U} > 0$, there is $\xi_{i_0,m_0} \in \Gamma_{0,i_0,m_0}$, such that $\|\xi_{i_0,m_0}\|_1 = \int_{\mathfrak{X}^p_{i_0,m_0}} |\xi_{i_0,m_0}(x)|f_p(x)dx > \overline{U}$. Then pick any $\xi_{-i_0m_0} \in \Gamma_{0,-i_0m_0}$ such that $\xi = (\xi_{i_0,m_0}, \xi_{-i_0m_0}) \in \Gamma_0$. Then
$\|\xi\| = \max_{1\leq i\leq n} \max_{1\leq m\leq M_i} \int_{\mathfrak{X}^p_{i,m}} |\xi_{i,m}(x)|f_p(x)dx \geq \int_{\mathfrak{X}^p_{i_0,m_0}} |\xi_{i_0,m_0}(x)|f_p(x)dx > \overline{U}$,
which is a contradiction. Therefore, every $\Gamma_{0,i,m}$ is uniformly bounded under the $\|\cdot\|_1$ norm. Similarly, suppose that there is some (i_0, m_0), with some $\eta_{i_0,m_0} > 0$, for any $\delta_{i_0,m_0} > 0$, there is $\widetilde{x}^{i_0,m_0}$ with $\|\widetilde{x}^{i_0}\|_E < \delta_{i_0,m_0}$ and $\xi_{i_0,m_0} \in \Gamma_{0,i_0,m_0}$, such that $\int_a^b |\xi_{i_0,m_0}(x+\widetilde{x}^{i_0,m_0})f_p(x+\widetilde{x}^{i_0,m_0}) - \xi_{i_0,m_0}(x)f(x)|dx > \eta_{i_0,m_0}$. Pick any $\xi_{-i_0m_0} \in \Gamma_{0,-i_0m_0}$ such that $\xi = (\xi_{i_0,m_0}, \xi_{-i_0m_0}) \in \Gamma_0$. Then $\max_{1\leq i\leq n} \max_{1\leq m\leq M_i} \int_{\mathfrak{X}^p_{i,m}} |\xi_{i,m}(x + \widetilde{x}^{i,m})f_p(x + \widetilde{x}^{i,m}) - \xi_{i,m}(x)f(x)|dx \geq \int_{\mathfrak{X}^p_{i_0,m_0}} |\xi_{i_0,m_0}(x + \widetilde{x}^{i_0,m_0})f_p(x + \widetilde{x}^{i_0,m_0}) - \xi_{i_0,m_0}(x)f(x)|dx > \eta$, which contradicts that Γ_0 satisfies (2). Similarly, suppose that there is some (i_0, m_0), with some $\eta_{i_0,m_0} > 0$, for any $R > 0$, there is $r > R$, such that $\int_{\mathfrak{X}^p_{i_0,m_0}} |\xi_{i_0,m_0}(x)|f_p(x)dx > \eta_{i_0,m_0}$. Then $\max_{1\leq i\leq n} \max_{1\leq m\leq M_i} \int_{\mathfrak{X}^p_{i,m}-C_{r,i,m}} |\xi_{i,m}(x)|dx \geq \int_{\mathfrak{X}^p_{i_0,m_0}} |\xi_{i_0,m_0}(x)|f_p(x)dx > \eta$, which is also a contradiction. By Corollary E.1, each $\Gamma_{0,i,m}$ is relatively compact in $L^1(\mathfrak{X}^p_{\widetilde{J}_{i,m}}, \mathfrak{B}_{i,m}, \mu_p; \Re^1)$. It then follows from LemmaE.2 that Γ_0 is

relatively compact in $(\Xi(W_n, \mathfrak{J}), \|\cdot\|)$. ■

The existence of an equilibrium is established by Schauder fixed point theorem, which is cited below.

Proposition E.4 *[Schauder fixed point theorem] Let $\mathcal{K}$ be a nonempty, closed, and convex subset of a normed space. Let T be a continuous mapping from $\mathcal{K}$ into a compact subset of $\mathcal{K}$. Then T has a fixed point in $\mathcal{K}$.*

In order to apply this theorem to operator T defined in (E.1), we impose two assumptions, Assumptions 4.2 and 4.3.

Lemma E.4 *Under Assumption 4.2, T is continuous. Actually, it is a Lipschitz function.*

Proof. For any $\xi, \xi^{'} \in (\Xi(W_n, \mathcal{J}), \|\cdot\|)$,

$$\begin{aligned}
&\|T(\xi) - T(\xi^{'})\| \\
&= \max_{1\le i\le n} \max_{1\le m\le M_i} \int E\Big[H_i(u(X_i) + \lambda \sum_{j\neq i} W_{n,i}\xi_{j,J_i}(x^p_{J_i})) - H_i(u(X_i) + \lambda \sum_{j\neq i} W_{n,i}\xi^{'}_{j,J_i}(x^p_{J_i}))|x^p_{i,m}, z\Big] dF_p \\
&\le \max_{1\le i\le n} \max_{1\le m\le M_i} \sup_c |dH_i(c)/dc||\lambda| \sum_{j\neq i} W_{n,ij} \int E\Big[|\xi_{i,J_i}(x^p_{J_i}) - \xi^{'}_{i,J_i}(x^p_{J_i})|\Big|x^p_{i,\tilde{J}_{i,m}}, z\Big] dF_p \\
&= \max_{1\le i\le n} \max_{1\le m\le M_i} \sup_c |dH_i(c)/dc||\lambda| \sum_{j\neq i} W_{n,ij} \int |\xi_{i,J_i}(x^p_{J_i}) - \xi^{'}_{i,J_i}(x^p_{J_i})| dF_p \\
&\le \max_{1\le i\le n} \sup_c |dH_i(c)/dc||\lambda|\|W_n\|_\infty \|\xi - \xi^{'}\|.
\end{aligned}$$

That is to say, T is a Lipschitz function. Thus, it is continuous in $(\Xi(W_n, \mathcal{J}), \|\cdot\|)$. ■

Lemma E.5 *Under Assumption 4.3, there is $r_0 > 0$, such that there is no equilibria out of the closed ball, $B[0, r_0] = \{\xi \in (\Xi(W_n, \mathcal{J}), \|\cdot\|) : \|\xi\| \le r_0\}$. In addition, for any $\xi \in B[0, r_0]$, $T(\xi) \in B[0, r_0]$.*

Proof. Because $\|T(\xi) - \xi\| \ge \big|\|T(\xi)\| - \|\xi\|\big|$, under Assumption 4.3, there is $r_1 > 0$, such that $\|T(\xi) - \xi\| > 0$ for all ξ with $\|\xi\| > r_1$. Now we show that there is $r_2 > 0$, such that for any $r \ge r_2$, $T(B[0, r]) \subseteq B[0, r]$. If this statement does not hold, for any positive r, there is $r^* \ge r$ and ξ^{r*} with $\|\xi^{r*}\| \le r^*$, $\|T(\xi^{r*})\| > r^* \ge \|\xi^{r*}\|$. Then $\|T(\xi^{r*})\|/\|\xi^{r*}\| > 1$, which contradicts Assumption 4.3. Choose $r_0 = \max\{r_1, r_2\}$, we get the results. ■

Proposition E.5 *Under Assumptions 4.2 and 4.3, if in addition,*

$$\max_{1\le i\le n}\max_{1\le m\le M_i}\int_{\mathfrak{X}^p}|T(\xi)_{i,m}(x+\widetilde{x})f_p(x+\widetilde{x})-T(\xi)_{i,m}(x)f_p(x)|dx\to 0, \tag{E.2}$$

as $\widetilde{x}\to 0$, uniformly for any $\xi\in B[0,r_0]$; and

$$\max_{1\le i\le n}\max_{1\le m\le M_i}\int_{\mathfrak{X}^p-C_r}|T(\xi)_{i,m}(x)|f_p(x)dx\to 0, \tag{E.3}$$

as $r\to\infty$, uniformly for all $\xi\in B[0,r_0]$, the set of equilibria, $\mathfrak{E}(X,W_n)$ is a nonempty and compact subset of $(\Xi(W_n,\mathcal{J}),\|\cdot\|)$ and is contiained in the closed ball $B[0,r_0]$. In particular, (E.2) *and* (E.3) *are satisfied, if*

1. *$H_i(\cdot)$'s are uniformly bounded, i.e., $\max_{1\le i\le n}\sup_{a\in\Re^1}|H_i(a)|\le\overline{B}^{'}$, for some $\overline{B}^{'}$;*

2. *$E[X_i^p|Z=z]<\infty$, for all i; and*

3. *For some $\delta_0>0$, for each (i,m), there is an function $g_{i,m}(x,\widehat{x})$ such that $\int_{\mathfrak{X}^p_{i,m}}\int_{\mathfrak{X}^p_{J_i}}|g_{i,m}(x,\widehat{x})|dxd\widehat{x}<\infty$, $|f_{p,i,m}(x+\widetilde{x},\widehat{x})|\le g_{i,m}(x,\widehat{x})$, a.e., for any $\widetilde{x}$ in the cube C_{δ_0}, where $f_{p,i,m}(\cdot,\cdot)$ is the joint density of $X^p_{i,m}=X^p_{J_{i,m}}$ and $X^p_{J_i}$ conditional on public information $Z=z$.*[24]

Proof. Choose the closed ball $B[0,r_0]$ satisfying the properties stated in Lemma E.5. It is nonempty, closed, and convex in space $(\Xi(W_n,\mathcal{J}),\|\cdot\|)$. For any $\xi^{'}\in T(B[0,r_0])$, $\xi^{'}=T(\xi)$, for some $\xi\in B[0,r_0]$. Therefore,

$$\begin{aligned}&\max_{1\le i\le n}\max_{1\le m\le M_i}\int_{\mathfrak{X}^p}|\xi^{'}_{i,m}(x+\widetilde{x})f_p(x+\widetilde{x})-\xi^{'}_{i,m}(x)f_p(x)|dx\\ &=\max_{1\le i\le n}\max_{1\le m\le M_i}\int_{\mathfrak{X}^p}|T(\xi)_{i,m}(x+\widetilde{x})f_p(x+\widetilde{x})-T(\xi)_{i,m}(x)f_p(x)|dx;\end{aligned}$$

$$\max_{1\le i\le n}\max_{1\le m\le M_i}\int_{\mathfrak{X}^p-C_r}|\xi^{'}_{i,m}(x)|f_p(x)dx=\max_{1\le i\le n}\max_{1\le m\le M_i}\int_{\mathfrak{X}^p-C_r}|T(\xi)_{i,m}(x)|f_p(x)dx.$$

Therefore, if (E.2) and (E.3) are satisfied, according to Proposition E.3, $T(B[0,r_0])$ is relatively compact in the normed space $(\Xi(W_n,\mathcal{J}),\|\cdot\|)$. Its closure, $\overline{T}(B[0,r_0])$, is compact. Moreover, $\overline{T}(B[0,r_0])\subseteq B[0,r_0]$, for $B[0,r_0]$ is closed. By the Schauder fixed point, T has a fixed point in $B[0,r_0]$. Thus, the set of equilibria, $\mathfrak{E}(X,W_n)$, is nonempty. Since it is the set of fixed points for the continuous operator T, $\mathfrak{E}(X,W_n)$ is closed. As a closed subset of the compact set $\overline{T}(B[0,r_0])$, $\mathfrak{E}(X,W_n)$ is compact.

[24] It is possible that they have overlaps between $X^p_{i,m}$ and $X^p_{J_i}$. For example, both agent 1 and agent 2 know X^p_3. We write $f_{p,i,m}(x,y)$ in order to simplify notations.

In particular, if $H_i(\cdot)$'s are uniformly bounded,

$$
\begin{aligned}
&\max_{1\leq i\leq n}\max_{1\leq m\leq M_i}\int_{\mathfrak{X}^p_{i,m}}\big|T(\xi)_{i,m}(x+\widetilde{x})f_p(x+\widetilde{x})-T(\xi)_{i,m}(x)f_p(x)\big|dx\\
&=\max_{1\leq i\leq n}\max_{1\leq m\leq M_i}\int_{\mathfrak{X}^p_{i,m}}\big|E[H_i(u(X^g,X_i^c,\widehat{x})+\lambda\sum_{j\neq i}W_{n,ij}\xi_j(\widehat{x}))|x+\widetilde{x},z]f_p(x+\widetilde{x})\\
&\quad-E[H_i(u(X^g,X_i^c,\widehat{x})+\lambda\sum_{j\neq i}W_{n,ij}\xi_j(\widehat{x}))|x,z]f_p(x)\big|dx\\
&=\max_{1\leq i\leq n}\max_{1\leq m\leq M_i}\int_{\mathfrak{X}^p_{i,m}}|\int_{\mathfrak{X}^p_{J_i}}H_i(u(X^g,X_i^c,\widehat{x})+\lambda\sum_{j\neq i}W_{n,ij}\xi_j(\widehat{x}))(f_p(x+\widetilde{x},\widehat{x})-f_p(x,\widehat{x}))d\widehat{x}|dx\\
&\leq\max_{1\leq i\leq n}\max_{1\leq m\leq M_i}\sup_{a\in\Re^1}|H_i(a)|\int_{\mathfrak{X}^p_{i,m}}\int_{\mathfrak{X}^p_{J_i}}|f_p(x+\widetilde{x},\widehat{x})-f_p(x,\widehat{x}))|d\widehat{x}dx,
\end{aligned}
$$

which, following from the Lebesgue dominated convergence theorem, goes to 0 uniformly for all ξ's as $\widetilde{x}\to 0$, under the above distribution assumption. Similarly, for any $r>0$,

$$
\begin{aligned}
&\max_{1\leq i\leq n}\max_{1\leq m\leq M_i}\int_{\mathfrak{X}^p-C_r}|T(\xi)_{i,m}(x)|f_p(x)dx\\
&=\max_{1\leq i\leq n}\max_{1\leq m\leq M_i}\int_{\mathfrak{X}^p-C_r}|\int_{\mathfrak{X}^p_{J_i}}H_i(u(X^g,X_i^c,\widehat{x})+\lambda\sum_{j\neq i}W_{n,ij}\xi_j(y))f_p(\widehat{x}|x)d\widehat{x}|f_p(x)dx\\
&\leq\max_{1\leq i\leq n}\max_{1\leq m\leq M_i}\sup_{a\in\Re^1}|H_i(a)|\int_{\mathfrak{X}^p-C_r}f_p(x)dx \qquad \text{(E.4)}\\
&\leq\max_{1\leq i\leq n}\max_{1\leq m\leq M_i}\sup_{a\in\Re^1}|H_i(a)|\sum_{j:J_{i,m}(j)=1}P(|X_j^p|>r)\\
&\leq\max_{1\leq i\leq n}\max_{1\leq m\leq M_i}\sup_{a\in\Re^1}|H_i(a)|\sum_{j:J_{i,m}(j)=1}E[|X_j^p||z]/r,
\end{aligned}
$$

where the last inequality follows from the Chebyshev's inequality. As $r\to\infty$, the above formula goes to zero uniformly for all ξ. ∎

It is obvious that if X_i^p's have a bounded support and their joint density conditional on public information is continuous, $|f_{p,i,m}(x+\widetilde{x},y)|$ can be dominated by the constant function on the support, which is Lebesgue integrable. Now we show that we can dominate the density for jointly normal random vectors.

Lemma E.6 *Suppose that $(X_1',X_2')'$ are jointly normal with mean $(\mu_1',\mu_2')'$ and variance-covariance matrix,*

$$
\Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}.
$$

Take $\delta_0 > 0$ arbitrarily. Then there is a function $g : \Re^{k_1} \times \Re^{k_2} \to \Re^1$, such that the joint density $f(x+\widetilde{x}, \widehat{x}) \leq g(x, \widehat{x})$ a.e. for any $\widetilde{x} \in C_{\delta_0}$.

Proof. Denote the inverse of Σ by

$$\Sigma^{-1} = \begin{pmatrix} \widetilde{\Sigma}_{11} & \widetilde{\Sigma}_{12} \\ \widetilde{\Sigma}_{21} & \widetilde{\Sigma}_{22} \end{pmatrix}.$$

The joint density takes the following form:

$$\begin{aligned} & f(x+\widetilde{x}, \widehat{x}) \\ =& (2\pi)^{-(k_1+k_2)/2}(det(\Sigma))^{-1/2} \exp\left\{-(1/2)(x' + \widetilde{x}' - \mu_1', \widehat{x}' - \mu_2')\Sigma^{-1}(x+\widehat{x}-\mu_1, \widehat{x}-\mu_2)\right\} \\ =& (2\pi)^{-(k_1+k_2)/2}(det(\Sigma))^{-1/2} \exp\left\{-(1/2)(\widehat{x}-\mu_2)'(\widetilde{\Sigma}_{22} - \widetilde{\Sigma}_{21}\widetilde{\Sigma}_{11}^{-1}\widetilde{\Sigma}_{12})(\widehat{x}-\mu_2)\right\} \\ & \cdot \exp\left\{-(1/2)(\widetilde{x} + (x-\mu_1) + \widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2))'\widetilde{\Sigma}_{11}(\widetilde{x} + (x-\mu_1) + \widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2))\right\}. \end{aligned}$$

For any $(x, \widehat{x})$, since the cube C_{δ_0} is compact in an Euclidean space, we can define $\widetilde{g}(x, \widehat{x}) = \min_{\widetilde{x} \in C_{\delta_0}}(\widetilde{x} + (x-\mu_1) + \widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2))'\widetilde{\Sigma}_{11}(\widetilde{x} + (x-\mu_1) + \widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2))$. This function is defined on the basis of an optimization problem about a quadratic form with respect to linear constraints. For any $c \geq 0$, define the lower contour set,

$$L(c, x, \widehat{x}) = \left\{\widetilde{x} : (\widetilde{x} + (x-\mu_1) + \widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2))'\widetilde{\Sigma}_{11}(\widetilde{x} + (x-\mu_1) + \widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2)) \leq c\right\}.$$

Each of those sets is an aera composed of an ellipse and its interior. It is convex. Fixing $(x, \widehat{x})$, we either have a solution inside C_{δ_0}, $-(x-\mu_1) - \widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2)$, or a corner solution at a boundary point of the cube C_{δ_0}. In the space for $\widetilde{x}$, fixing c, when $(x, \widehat{x})$ varies, the center of the ellipse moves; while the axes do not change. Therefore, we can divide the space for $(x, \widehat{x})$ into several regions, such that two points in the same region either both have interior solutions or corner solutions on the same edge of C_{δ_0}.[25] If $-(x-\mu_1) - \widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2) \in C_{\delta_0}$, $\widetilde{g}(x, \widehat{x}) = 0$. Otherwise, its value depends on the minimal boundary point. In this case, as the edges of C_{δ_0} are bounded, for any $\widetilde{x}$ on the boundary of C_{δ_0},

$$\begin{aligned} & \Big|(\widetilde{x} + (x-\mu_1) + \widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2))'\widetilde{\Sigma}_{11}(\widetilde{x} + (x-\mu_1) + \widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2)) \\ & - ((x-\mu_1) + \widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2))'\widetilde{\Sigma}_{11}((x-\mu_1) + \widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2))\Big| / \|(x, \widehat{x})\|_E^2 \to 0 \end{aligned}$$

[25] This can be seen clearly for the special case when $k_1 = k_2 = 1$. In that case, for any $(x, \widehat{x})$, if $-(x-\mu_1) - \widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2) < -\delta_0$, the minimizing solution is $-\delta_0$; if $-\delta_0 \leq -(x-\mu_1) - \widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2) \leq \delta_0$, the minimizing solution is interior; if $-(x-\mu_1) - \widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2) > \delta_0$, the minimizing solution is δ_0.

as $\|(x,\widehat{x})\|_E \to \infty$. Therefore, there is $R>0$, such that when $\|(x,\widehat{x})\|_E > R$,

$\widetilde{g}(x,\widehat{x}) \geq (1/4)((x-\mu_1)+\widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2))'\widetilde{\Sigma}_{11}((x-\mu_1)+\widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2))$.

Define $g(x,\widehat{x}) = (2\pi)^{-(k_1+k_2)/2}(det(\Sigma))^{-1/2}\exp\{-(1/2)\widetilde{g}(x,\widehat{x})\}$. Then $f(x+\widetilde{x},\widehat{x}) \leq g(x,\widehat{x})$, for any $(x,\widehat{x})$. By the Maximum Theorem, $\widetilde{g}(x,\widehat{x})$ is continuous, so does $g(x,\widehat{x})$. Moreover, if $-(x-\mu_1)-\widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2) \in C_{\delta_0}$, $g(x,\widehat{x}) = (2\pi)^{-(k_1+k_2)/2}(det(\Sigma))^{-1/2}$; otherwise, $g(x,\widehat{x})$ is continuous when $\|(x,\widehat{x})\|_E \leq R$, and when $\|(x,\widehat{x})\|_E > R$,

$$\begin{aligned} g(x,\widehat{x}) \leq &(2\pi)^{-(k_1+k_2)/2}(det(\Sigma))^{-1/2} \\ &\cdot \exp\left\{-(1/8)((x-\mu_1)+\widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2))'\widetilde{\Sigma}_{11}((x-\mu_1)+\widetilde{\Sigma}_{11}^{-1/2}\widetilde{\Sigma}_{12}(\widehat{x}-\mu_2))\right\}. \end{aligned}$$

Hence, $g(x,\widehat{x})$ is integrable. ■

F Equilibrium for Peer Effects

Proof of Proposition 5.1. On one hand, suppose that $\overline{\xi}^e$ satisfying (5.4). With a regular group, we can define $\xi^e_{j,m} = \left(\Lambda_j(\overline{\xi}^e_{m(j)})\right)_m$ for all $j=1,\cdots,n$ and $m=1,\cdots,M_0$. (5.4) implies that

$$\begin{aligned} &\overline{\xi}^e_{m(i)}(x^p_{J,m(i)}) \\ =&\sum_{j=1}^n E[H_j(u(X_j)+\lambda\overline{\xi}^e_{m(j)}(X^p_{J,m(j)})-\lambda(\Lambda_j(\overline{\xi}^e_{m(j)}))_{m(j)}(X^p_{J,m(j)}))|X^p_{J,m(i)}=x^p_{J,m(i)},z] \\ =&\sum_{j=1}^n \left(\Lambda_j(\overline{\xi}^e_{m(j)})\right)_{m(i)}(x^p_{J,m(i)}) \\ =&\sum_{j=1}^n \xi^e_{j,m(i)}(x^p_{J,m(i)}). \end{aligned}$$

for all i and $x^p_{J,m(i)} \in \mathfrak{X}^p_{J,m(i)}$. Therefore,

$$\begin{aligned} \xi^e_{i,m}(x^p_{J,m}) =&\left(\Lambda_i(\overline{\xi}^e_{m(i)})\right)_m(x^p_{J,m}) \\ =&E[H_i(u(X_i)+\lambda\overline{\xi}^e_{m(i)}(X^p_{J,m(i)})-\lambda\xi^e_{i,m(i)}(X^p_{J,m(i)}))|X^p_{J,m}=x^p_{J,m},z] \\ =&E[H_i(u(X_i)+\lambda\sum_{j\neq i}\xi^e_{j,m(i)}(X^p_{J,m(i)}))|X^p_{J,m}=x^p_{J,m},z]. \end{aligned}$$

On the other hand, given that $\xi^e = (\xi^e_{1,1},\cdots,\xi^e_{1,M_0},\cdots,\xi^e_{n,1},\cdots,\xi^e_{n,M_0})$ satisfies (5.1), define

$\overline{\xi}^e = (\overline{\xi}^e_1, \cdots, \overline{\xi}^e_{M_0})$ such that

$$\overline{\xi}^e(x^p_{J,1}, \cdots, x^p_{J,M_0})_m = \overline{\xi}^e_m(x^p_{J,m}) = \sum_{i=1}^n \xi^e_{i,m}(x^p_{J,m}),$$

for all m and $x^p_{J,m} \in \mathfrak{X}^p_{J,m}$. Then we get that

$$\begin{aligned}
\xi^e_{i,m}(x^p_{J,m}) =& E[H_i(u(X_i) + \lambda \sum_{j \neq i} \xi^e_{j,m(i)}(X^p_{J,m(i)}))|X^p_{J,m} = x^p_{J,m}, z] \\
=& E[H_i(u(X_i) + \lambda \overline{\xi}^e_{m(i)}(X^p_{J,m(i)}) - \lambda \xi^e_{i,m(i)}(X^p_{J,m(i)}))|X^p_{J,m} = x^p_{J,m}, z].
\end{aligned}$$

Applying the Implicit Function theorem in Banach spaces, $\xi^e_{i,m} = (\Lambda_i(\overline{\xi}^e_{m(i)}))_m$, for all $i = 1, \cdots, n$ and $m = 1, \cdots, M_0$. Therefore,

$$\begin{aligned}
\overline{\xi}^e_m(x^p_{J,m}) =& \sum_{i=1}^n \xi^e_{i,m}(x^p_{J,m}) \\
=& \sum_{i=1}^n (\Lambda_i(\overline{\xi}^e_{m(i)}))_m(x^p_{J,m}) \\
=& \sum_{i=1}^n E[H_i(u(X_i) + \lambda \overline{\xi}^e_{m(i)}(X^p_{J,m(i)}) - \lambda \xi^e_{i,m(i)}(X^p_{J,m(i)}))|X^p_{J,m} = x^p_{J,m}, z] \\
=& \sum_{i=1}^n E[H_i(u(X_i) + \lambda \overline{\xi}^e_{m(i)}(X^p_{J_i}) - \lambda(\Lambda_i(\overline{\xi}^e_{m(i)}))(X^p_{J_i}))|X^p_{J,m} = x^p_{J,m}, z].
\end{aligned}$$

■

G Proofs for Equilibrium Set Characterization in Binary Choice Models

Proof for Proposition 6.1. Let $\overline{\xi}$ denote the total expected outcome in the group. That is, $\overline{\xi} = \sum_{i=1}^n \xi_i = \sum_1^n E[y_i]$. Given $\overline{\xi}$, for agent i, $K(\xi_i, \overline{\xi}; u_i, \lambda) = \Phi(u_i + \lambda\overline{\xi} - \lambda\xi_i) - \xi_i = 0$. $K(0, \overline{\xi}; u_i, \lambda) > 0$. $K(1, \overline{\xi}; u_i, \lambda) < 0$. $\frac{\partial K}{\partial \xi_i} = -(\lambda\phi(u_i + \lambda\overline{\xi} - \lambda\xi_i) + 1)$. If $\lambda\phi(u_i + \lambda\overline{\xi} - \lambda\xi_i) + 1 > 0$, for each $\overline{\xi}$, there is a unique $\xi_i \in (0, 1)$ such that $K(\xi_i, \overline{\xi}; u_i, \lambda) = 0$. Thus, individual expected outcomes are determined by a function, $\xi_i = G(u_i, \overline{\xi})$, such that

$$\frac{\partial G(u_i, \overline{\xi})}{\partial \overline{\xi}} = \frac{\lambda\phi(u_i + \lambda\overline{\xi} - \lambda G(u_i, \overline{\xi}))}{\lambda\phi(u_i + \lambda\overline{\xi} - \lambda G(u_i, \overline{\xi})) + 1},$$

$$\frac{\partial^2 G(u_i,\overline{\xi})}{\partial \overline{\xi}^2} = -\frac{\lambda^2 \phi(u_i+\lambda\overline{\xi}-\lambda G(u_i,\overline{\xi}))(u_i+\lambda\overline{\xi}-\lambda G(u_i,\overline{\xi}))}{[\lambda\phi(u_i+\lambda\overline{\xi}-\lambda G(u_i,\overline{\xi}))+1]^3}.$$

$\overline{\xi}$ is determined by

$$S(u,\overline{\xi}) = \sum_{i=1}^{n} G(u_i,\overline{\xi}) - \overline{\xi}.$$

As a result, $\frac{\partial S}{\partial \overline{\xi}} = \sum_{i=1}^n \frac{\partial G(u_i,\overline{\xi})}{\partial \overline{\xi}} - 1$ and $\frac{\partial^2 S}{\partial \overline{\xi}^2} = \sum_{i=1}^n \frac{\partial^2 G(u_i,\overline{\xi})}{\partial \overline{\xi}^2}$. As $\xi_i \in [0,1]$ for $i = 1, \cdots, n$, the valid equilibrium $\overline{\xi}^e \in [0,n]$. Since $G(u_i,\overline{\xi}) \in (0,1)$, $S(u,0) > 0$ and $S(u,n) < 0$. Therefore, there must be an equilibrium between 0 and n.

- When $-\sqrt{2\pi} < \lambda < 0$, $\lambda\phi(u_i+\lambda\overline{\xi}-\lambda G(u_i,\overline{\xi}))+1 > 0$. Thus, $\frac{\partial G(u_i,\overline{\xi})}{\partial \overline{\xi}} < 0$ for all i and $\frac{\partial S}{\partial \overline{\xi}} < 0$. As a result, there is a unique equilibrium.
- When $\lambda > 0$, $\frac{\partial G(u_i,\overline{\xi})}{\partial \overline{\xi}} > 0$ for all i. If $\min_{1\le i\le n} u_i > \frac{\lambda}{2}$, $\Phi(u_i - \frac{\lambda}{2}) - \frac{1}{2} > 0$. That is, $K(\frac{1}{2}, 0; u_i, \lambda) > 0$. Then for all i, $G(u_i,\overline{\xi}) \ge G(u_i,0) > \frac{1}{2}$. Therefore, $u_i + \lambda\overline{\xi} - \lambda G(u_i,\overline{\xi}) = \Phi^{-1}(G(u_i,\overline{\xi})) > 0$. It follows that $\frac{\partial^2 G(u_i,\overline{\xi})}{\partial \overline{\xi}^2} < 0$ for any $\overline{\xi} \in [0,n]$. Similarly, if $\max_{1\le i\le n} u_i < \frac{\lambda}{2} - \lambda n$, for any i, $\Phi(u_i + \lambda n - \frac{\lambda}{2}) - \frac{1}{2} < 0$. That is, $K(\frac{1}{2}, n; u_i, \lambda) < \frac{1}{2}$. Then $G(u_i,\overline{\xi}) \le G(u_i,n) < \frac{1}{2}$. Then $u_i + \lambda\overline{\xi} - \lambda G(u_i,\overline{\xi}) = \Phi^{-1}(G(u_i,\overline{\xi}) < 0$. In this case, $\frac{\partial^2 G(u_i,\overline{\xi})}{\partial \overline{\xi}^2} > 0$ for all $\overline{\xi} \in [0,n]$. In both cases, $\frac{\partial^2 G(u_i,\overline{\xi})}{\partial \overline{\xi}^2}$ does not change sign as $\overline{\xi}$ runs in $[0,n]$. Hence, there is a unique equilibrium.

■

Proof of Lemma 6.1. It is easy to see that $c(a;\lambda) = c(-a;\lambda)$. $c(0;\lambda) = \lambda\phi(0) + 1 > 0$. $\lim_{a\to+\infty} c(a;\lambda) = \lim_{a\to-\infty} c(a;\lambda) = -\infty$. $c^{'}(a;\lambda) = a(3\lambda\phi(a) - 2 - 2\lambda a^2\phi(a))$. When $0 < \lambda < \frac{2\sqrt{2\pi}}{3}$, $3\lambda\phi(a) - 2 - 2\lambda a^2\phi(a) < 0$. Thus, $c^{'}(a;\lambda) < 0$ for $a > 0$ and $c^{'}(a;\lambda) > 0$ for $a < 0$. Therefore, there is $a_+ > 0$ with the claimed properties. Additionally, $\frac{da_+}{d\lambda} = -\frac{(2a^2+1)\phi(a)}{a(3\lambda\phi(a)-2-2\lambda a^2\phi(a))} > 0$. ■

Proof of Proposition 6.2. For any i,

$$\frac{\partial^3 G(u_i,\overline{\xi})}{\partial \overline{\xi}^3} = -\frac{\lambda^3 \phi(u_i+\lambda\overline{\xi}-\lambda G(u_i,\overline{\xi}))c(u_i+\lambda\overline{\xi}-\lambda G(u_i,\overline{\xi});\lambda)}{[\lambda\phi(u_i+\lambda\overline{\xi}-\lambda G(u_i,\overline{\xi}))+1]^5}.$$

$\frac{\partial^3 S}{\partial \overline{\xi}^3} = \sum_{i=1}^n \frac{\partial^3 G(u_i,\overline{\xi})}{\partial \overline{\xi}^3}$. As $\lambda > 0$, $\frac{\partial G(u_i,\overline{\xi})}{\partial \overline{\xi}} > 0$ for all i. If (6.4) holds, for any i, $\Phi(u_i - \lambda\Phi(a_+(\lambda))) - \Phi(a_+(\lambda)) > 0$. That is, $K(\Phi(a_+(\lambda)), 0; u_i, \lambda) > 0$. Therefore, $G(u_i,\overline{\xi}) \ge G(u_i,0) > \Phi(a_+(\lambda))$ for all $\overline{\xi} \in [0,n]$. It follows that $u_i + \lambda\overline{\xi} - \lambda G(u_i,\overline{\xi}) = \Phi^{-1}(G(u_i,\overline{\xi})) > a_+(\lambda)$. From Lemma 6.1, $c(u_i + \lambda\overline{\xi} - \lambda G(u_i,\overline{\xi});\lambda) < 0$, for all i and $\overline{\xi} \in [0,n]$. Analogously, under condition

(6.5), for any i, $\Phi(u_i + \lambda n - \lambda\Phi(-a_+(\lambda))) - \Phi(-a_+(\lambda)) < 0$. Then $K(\Phi(-a_+(\lambda)), n; u_i, \lambda) < 0$. Therefore, $G(u_i, \bar{\xi}) \le G(u_i, n) < \Phi(-a_+(\lambda))$. That is, for any i. $u_i + \lambda\bar{\xi} - \lambda G(u_i, \bar{\xi}) = \Phi^{-1}(G(u_i, \bar{\xi})) < -a_+(\lambda)$. By Lemma 6.1, $c(u_i + \lambda b - \lambda G(u_i, \bar{\xi}); \lambda) < 0$, for all i and $\bar{\xi} \in [0, n]$. Thus, in both cases, $\frac{\partial^3 S}{\partial \bar{\xi}^3} > 0$ when $\bar{\xi}$ runs from 0 to n. If there are more than three equilibria, $\frac{\partial^2 S}{\partial \bar{\xi}^2}$ must change its sign, which is impossible if $\frac{\partial^3 S}{\partial \bar{\xi}^3}$ keeps on being positive. Consequently, under condition (6.4) or (6.5), there are at most three equilibria. ■

Proof of Proposition 6.3. Fix λ, for any i, $\widetilde{K}(\xi_i, \bar{\xi}; u_i, \lambda) = 2\Phi(u_i + \lambda\bar{\xi} - \lambda\xi_i) - 1 - \xi_i = 0$. As $\widetilde{K}(-1, \bar{\xi}; u_i, \lambda) > 0$, $\widetilde{K}(1, \bar{\xi}; u_i, \lambda) < 0$, and $\frac{\partial \widetilde{K}}{\partial \xi_i} = -(2\lambda\phi(u_i + \lambda\bar{\xi} - \lambda\xi_i) + 1)$, when $2\lambda\phi(u_i + \lambda\bar{\xi} - \lambda\xi_i) + 1 > 0$, for each $\bar{\xi}$, there is a unique $\xi_i \in [-1, 1]$ such that $\widetilde{K}(\xi_i, \bar{\xi}; u_i, \lambda) = 0$. Thus, ξ_i is implicitly a function of $\bar{\xi}$ and u_i. Denote this function as $\widetilde{G}(u_i, \bar{\xi}; \lambda)$. By computation,

$$\frac{\partial \widetilde{G}(u_i, \bar{\xi})}{\partial \bar{\xi}} = \frac{2\lambda\phi(u_i + \lambda\bar{\xi} - \lambda\widetilde{G}(u_i, \bar{\xi}))}{2\lambda\phi(u_i + \lambda\bar{\xi} - \lambda\widetilde{G}(u_i, \bar{\xi})) + 1},$$

$$\frac{\partial^2 \widetilde{G}(u_i, \bar{\xi})}{\partial \bar{\xi}^2} = -\frac{2\lambda^2\phi(u_i + \lambda\bar{\xi} - \lambda\widetilde{G}(u_i, \bar{\xi}))(u_i + \lambda\bar{\xi} - \lambda\widetilde{G}(u_i, \bar{\xi}))}{[2\lambda\phi(u_i + \lambda\bar{\xi} - \lambda\widetilde{G}(u_i, \bar{\xi})) + 1]^3}.$$

$\bar{\xi}$ is determined by

$$\widetilde{S}(u, \bar{\xi}) = \sum_{i=1}^{n} \widetilde{G}(u_i, \bar{\xi}) - \bar{\xi}.$$

As a result, $\frac{\partial \widetilde{S}}{\partial \bar{\xi}} = \sum_{i=1}^{n} \frac{\partial \widetilde{G}(u_i, \bar{\xi})}{\partial \bar{\xi}} - 1$ and $\frac{\partial^2 \widetilde{S}}{\partial \bar{\xi}^2} = \sum_{i=1}^{n} \frac{\partial^2 \widetilde{G}(u_i, \bar{\xi})}{\partial \bar{\xi}^2}$. As $\xi_i \in [-1, 1]$ for $i = 1, \cdots, n$, the valid equilibrium $\bar{\xi}^e \in [-n, n]$. Since $\widetilde{G}(u_i, \bar{\xi}) \in (-1, 1)$, $\widetilde{S}(u, -n) > 0$ and $\widetilde{S}(u, n) < 0$. Therefore, there must be an equilibrium between $-n$ and n.

- When $-\frac{\sqrt{2\pi}}{2} < \lambda < 0$, $2\lambda\phi(u_i + \lambda\bar{\xi} - \lambda\widetilde{G}(u_i, \bar{\xi})) + 1 > 0$. Then $\frac{\partial \widetilde{G}(u_i, \bar{\xi})}{\partial \bar{\xi}} < 0$ for all i and $\frac{\partial \widetilde{S}}{\partial \bar{\xi}} < 0$. Thus, in this case, there is a unique equilibrium.

- When $\lambda > 0$, $\frac{\partial \widetilde{G}(u_i, \bar{\xi})}{\partial \bar{\xi}} > 0$ for all i. If $\min_{1 \le i \le n} u_i > \lambda n$, $2\Phi(u_i - \lambda n) - 1 > 0$. That is, $\widetilde{K}(0, -n; u_i, \lambda) > 0$. Then for all i, $\widetilde{G}(u_i, \bar{\xi}) \ge \widetilde{G}(u_i, -n) > 0$. Therefore, $u_i + \lambda\bar{\xi} - \lambda\widetilde{G}(u_i, \bar{\xi}) = \Phi^{-1}(\frac{\widetilde{G}(u_i, \bar{\xi}) + 1}{2}) > 0$. It follows that $\frac{\partial^2 \widetilde{G}(u_i, \bar{\xi})}{\partial \bar{\xi}^2} < 0$ for any $\bar{\xi} \in [-n, n]$. Similarly, if $\max_{1 \le i \le n} u_i < -\lambda n$, for any i, $2\Phi(u_i + \lambda n) - 1 < 0$. That is, $\widetilde{K}(0, n; u_i, \lambda) < 0$. Hence, $\widetilde{G}(u_i, \bar{\xi}) \le \widetilde{G}(u_i, n) < 0$. Then $u_i + \lambda\bar{\xi} - \lambda\widetilde{G}(u_i, \bar{\xi}) = \Phi^{-1}(\frac{\widetilde{G}(u_i, \bar{\xi}) + 1}{2}) < 0$. In this case, $\frac{\partial^2 \widetilde{G}(u_i, \bar{\xi})}{\partial \bar{\xi}^2} > 0$ for all $\bar{\xi} \in [-n, n]$. That is to say, in both cases, $\frac{\partial^2 \widetilde{G}(u_i, \bar{\xi})}{\partial \bar{\xi}^2}$ does not change its sign as $\bar{\xi}$ runs in $[-n, n]$. Hence, there is a unique equilibrium in these two cases.

■

Proof of Lemma 6.2. First, it is easy to see that $\widetilde{c}(a;\lambda)=\widetilde{c}(-a;\lambda)$. $\widetilde{c}(0;\lambda)=2\lambda\phi(0)+1>0$. $\lim_{a\to+\infty}\widetilde{c}(a;\lambda)=\lim_{a\to-\infty}\widetilde{c}(a;\lambda)=-\infty$. $\frac{d\widetilde{c}(a;\lambda)}{da}=2a(3\lambda\phi(a)-1-2\lambda\phi(a)a^2)$. If $0<\lambda<\frac{\sqrt{2\pi}}{3}$, $3\lambda\phi(a)-1-2\lambda\phi(a)a^2<0$. Then $\frac{d\widetilde{c}(a;\lambda)}{da}>0$ if $a<0$; and $\frac{d\widetilde{c}(a;\lambda)}{da}<0$ if $a>0$. Thus, $a=0$ is the unique peak for $\widetilde{c}(\cdot;\lambda)$. By symmetry, there is $\widetilde{a}_+>0$ such that $\widetilde{c}(a;\lambda)>0$ if $-\widetilde{a}_+<a<\widetilde{a}_+$; $\widetilde{c}(\widetilde{a}_+;\lambda)=\widetilde{c}(-\widetilde{a}_+;\lambda)=0$; and $\widetilde{c}(a;\lambda)<0$ for $a>\widetilde{a}_+$ or $a<-\widetilde{a}_+$. In addition, from $\widetilde{c}(\widetilde{a}_+;\lambda)=0$, $\frac{d\widetilde{a}_+}{d\lambda}=\frac{\phi(\widetilde{a}_+)(1+2\widetilde{a}_+^2)}{\widetilde{a}_+(2\lambda\phi(\widetilde{a}_+)\widetilde{a}_+^2+1-3\lambda\phi(\widetilde{a}_+))}>0$. ■

Proof of Proposition 6.4. $\frac{\partial^3\widetilde{s}}{\partial\overline{\xi}^3}=\sum_{i=1}^n\frac{\partial^3\widetilde{G}}{\partial\overline{\xi}^3}$, where

$$\frac{\partial^3\widetilde{G}}{\partial\overline{\xi}^3}=-\frac{2\lambda\phi(u_i+\lambda\overline{\xi}-\lambda\widetilde{G}(u_i;\lambda))\widetilde{c}(u_i+\lambda\overline{\xi}-\lambda\widetilde{G}(u_i;\lambda);\lambda)}{(2\lambda^3\phi(u_i+\lambda\overline{\xi}-\lambda\widetilde{G}(u_i;\lambda))+1)^5}.$$

When $\lambda>0$, the sign of $\frac{\partial^3\widetilde{G}}{\partial\overline{\xi}^3}$ is determined by the sign of $\widetilde{c}(u_i+\lambda\overline{\xi}-\lambda\widetilde{G}(u_i;\lambda);\lambda)$. If $\min_{1\le i\le n}u_i>\lambda(2\Phi(\widetilde{a}_+(\lambda))-1+n)+\widetilde{a}_+(\lambda)$, for any i, $2\Phi\big(u_i-\lambda n-\lambda(2\Phi(\widetilde{a}_+(\lambda))-1)\big)-1>2\Phi(\widetilde{a}_+(\lambda))-1$. That is, $\widetilde{K}(2\Phi(\widetilde{a}_+(\lambda))-1,-n;u_i,\lambda)>0$. Thus, for any $\overline{\xi}\in[-n,n]$, $G(u_i,\overline{\xi})\ge G(u_i,-n)>2\Phi(\widetilde{a}_+(\lambda))-1$. Therefore, $u_i+\lambda\overline{\xi}-\lambda G(u_i,\overline{\xi})=\Phi^{-1}(\frac{G(u_i,\overline{\xi})+1}{2})>\widetilde{a}_+$. It then follows from Lemma 6.2 that $\widetilde{c}(u_i+\lambda\overline{\xi}-\lambda\widetilde{G}(u_i;\lambda);\lambda)<0$ for all i and $\overline{\xi}\in[-n,n]$. Similarly, if $\max_{1\le i\le n}u_i<\lambda(2\Phi(-\widetilde{a}_+(\lambda))-1-n)-\widetilde{a}_+(\lambda)$, for all i, $2\Phi\big(u_i+\lambda n-\lambda(2\Phi(-\widetilde{a}_+(\lambda))-1)\big)-1<2\Phi(-\widetilde{a}_+(\lambda))-1$. That is, $\widetilde{K}(2\Phi(-\widetilde{a}_+(\lambda))-1,n;u_i,\lambda)<0$. Then $G(u_i,\overline{\xi})\le G(u_i,n)<2\Phi(-\widetilde{a}_+(\lambda))-1$. Therefore, $u_i+\lambda\overline{\xi}-\lambda G(u_i,\overline{\xi})=\Phi^{-1}(\frac{G(u_i,\overline{\xi})+1}{2})<-\widetilde{a}_+$. Applying Lemma 6.2, $\widetilde{c}(u_i+\lambda\overline{\xi}-\lambda\widetilde{G}(u_i;\lambda);\lambda)<0$ for all i and $\overline{\xi}\in[-n,n]$. That is to say, when (6.11) or (6.12) holds, $\frac{\partial^3\widetilde{G}}{\partial\overline{\xi}^3}>0$ for all i and $\frac{\partial^3\widetilde{s}}{\partial\overline{\xi}^3}>0$ for all $\overline{\xi}\in[-n,n]$. In this case, $\frac{\partial^2\widetilde{s}}{\partial\overline{\xi}^2}$ does not change its sign in $[-n,n]$ and there are at most three equilibria. ■

H Discussions and Extensions

H.1 Group Unobservables

In the previous discussions, all exogenous characteristics, X^g, X_i^c's, and X_i^p's are observable to econometricians. In reality, however, some variables are known to agents but unavailable from the data. For example, researchers studying students' class performances may not know how good the teachers are, which is known to the students. Similarly, in market sale data, it is possible that no information about the wealth of customers is available from the data. But the firms may have some relevant information. In this section, we take into account unobservable variables that are public information in a group and can affect the payoff of all group members. Modeling them as random effects, we derive the following framework:

$$y_{i,g} = h_{i,g}(y^*_{i,g}), \tag{2.1$'$}$$

and

$$y^*_{i,g} = u(X^g, X^c_i, X^p_i, \zeta^g) + \lambda \sum_{j \neq i} W_{g,ij} E[y_{j,g} | X^p_{J_{i,g}}, Z] - \epsilon_{i,g}. \tag{2.2$'$}$$

where g is the group index. By (2.2$'$), we implicitly assume that all group unobervables are additive and summarize them as a single variable, ζ^g. Assume that ζ^g's are i.i.d. independent of all the other exogenous variables, social relations, as well as the idiosyncratic shocks. Their distribution is denoted by the pdf, $f_\zeta(\cdot;\vartheta)$.

Because each ζ^g is publicly known to agents in g, for interactions among group members, it acts the same as X^g. For any group g, we can characterize the equilibrium set and use a parametric stochastic selection rule to complete the model. Suppose that the distribution of equilibrium selection is

$$\rho(\xi^e; \mathfrak{E}(X^g, X^c, X^p, \zeta^g, W_g), \alpha) = \rho(\alpha' \gamma(\xi^e; X^g, X^c, X^p, W_g); \mathfrak{E}(X^g, X^c, X^p, \zeta^g, W_g)), \tag{H.1}$$

with some known selection rule $\gamma(\xi^e; X^g, X^c, X^p, W_g)$ and unknown parameter α. The above form shows that the group unobservables only affect the set of equilibria, but not the selection rule. Therefore, given a set of equilibria(finite, or approximated by a finite number of equilibrium expectation functions), the realizations of unobserved group features does not influence the distribution of equilibrium outcomes. This assumption is reasonable if the selection rule is consistent with social welfare maximization or Pareto optimization. The selection rules in previous sections satisfy (H.1). Then we can complete the model and write down the sample log likelihood function as follows:

$$\begin{aligned} \log L(Y; X, W) = \sum_{g=1}^{G} \log \Big[\int \int_{\xi^e \in \mathfrak{E}(X^g, X^c, X^p, \widetilde{\zeta}^g, W_g)} \prod_{i=1}^{n_g} f(y_{i,g} | \xi^e_g) \cdot \\ \rho(\alpha' \gamma(\xi^e; X^g, X^c, X^p, W_g); \mathfrak{E}(X^g, X^c, X^p, \widetilde{\zeta}^g, W_g)) f_\zeta(\widetilde{\zeta}^g; \vartheta) d\widetilde{\zeta}^g \Big]. \end{aligned} \tag{H.2}$$

Because X^g's are observed from the data set and ζ^g's are not, identification and estimation methods will be different from those in previous sections.

Assuming that $u(\cdot)$ is a linear function of exogenous characteristics, i.e.,

$u(X^g, X^c_{i,g}, X^p_{i,g}, \zeta^g) = \beta_{0,0} + X^{g'}\beta_{0,1} + X^{c'}_{i,g}\beta_1 + X^{p'}_{i,g}\beta_2 + \zeta^g$,

for all i and g. We normalize the mean of ζ^g to be equal to 0, i.e., $E[\zeta^g] = 0$. For data about a single group, applying the technique of identification at infinity, we can identify β_1, β_2, and $\widetilde{\beta}_0 = \beta_{0,0} + X^{g'}\beta_{0,1} + \zeta^g$ (as a whole) under certain conditions, based on our previous discussions. If $H_i(\cdot)$ is strictly increasing, the group average behaviors will be increasing in $\widetilde{\beta}_0 = \beta_{0,0} + X^{g'}\beta_{0,1} + \zeta^g$, which helps us identify $\beta_{0,0}$, $\beta_{0,1}$, and the distribution of ζ^g's from variations across groups.

As for estimation, we can calculate integration over unobserved ζ^g's by stochastic simulations. That is to say, we randomly take S draws from the distribution $f_\zeta(\cdot;\vartheta)$ for each g, $\left\{\zeta^{g,1}, \cdots, \zeta^{g,S}\right\}$ and calculate the simulated log likelihood:

$$\begin{aligned}\log \widehat{L}(Y; X, W) = &\sum_{g=1}^{G} \log \Big[\frac{1}{S}\sum_{s=1}^{S} \int_{\xi^e \in \mathfrak{E}(X^g, X^c, X^p, \zeta^{g,s}, W_g)} \prod_{i=1}^{n_g} f(y_{i,g}|\xi^e_g)\cdot \\ &\rho(\alpha^{'}\gamma(\xi^e; X^g, X^c, X^p, W_g); \mathfrak{E}(X^g, X^c, X^p, \zeta^{g,s}, W_g))\Big].\end{aligned} \tag{H.2$'$}$$

H.2 Deterministic Rule

In previous sections, we assume that an equilibrium is selected from the set of equilibria according to a stochastic rule. It is also possible to use a deterministic rule. To be specific, let $\mathfrak{E}(X, W_n)$ denote a set of equilibria for a group (X, W_n). It is equivalent to the set of solutions to a system of (generally nonlinear, functional) equations, $S(\xi; X, W_n) = 0$. Let $\Pi(\xi; X, W_n)$ denote a real-valued function of equilibria and group features (X, W_n). For instance, $\Pi(\xi; X, W_n)$ can be the expected total utility of the group, or the expected number of market entrants. We select an equilibrium to maximize the objective function:

$$\max_{\xi^e} \Pi(\xi^e; X, W_n) \;\; s.t. \;\; S(\xi^e; X, W_n) = 0. \tag{H.3}$$

When all exogenous covariates are known to the public, (H.3) is just an ordination optimization problem. According to our discussion in Section 3, the set of equilibria, or equivalently, the set of zeros for $S(\cdot; X, W_n)$, is finite. To solve (H.3) is to pick one of those finitely many points to maximize the criterion function. In general, however, the conditional expectation function, ξ^e, is a vector valued function defined on subsets of the Euclidean spaces, (H.3) is a problem of functional optimization. In that case, we may solve the optimization problem using optimal control.

For peer effects, especially, assuming that an equilibrium is selected under a fixed equilibrium selection rule, it is possible to derive a simpler estimation method. When both $\Pi(\xi^e; X, W_n)$ and $S(\xi^e; X, W_n)$ are continuous with exogenous characteristics, X, and the set of equilibria is compact, we can apply the Maximum theorem, which claims that the set of solutions to (H.3) is an upper-hemicontinuous correspondence of X.[26][27] If we can assure that there is a unique maximizer (by imposing some convexity conditiona, for example), the unique optimal equilibrium will be a continuous function of the group characteristics of X. So do the mean, $\overline{\xi}^*$. Therefore, when there is a large number of independent groups, we can non-parametrically estimate $\overline{\xi}^*$ from group means. By Proposition 5.1, with $\overline{\xi}^*$, we can recover conditional expectations on individual behaviors, the ξ_i^*'s. Then we can estimate model parameters either from sample likelihood function or moment conditions, conditioning on the equilibrium represented by $\overline{\xi}^*$.

The equilibrium conditional expectation acts like the group aggregate in the model built by Bisin et al.(2011). It shows that without assuming that the single equilibrium is played repeatedly over time periods or across markets, as long as the same criterion rule is used and there is a unique optimizing equilibrium, we can still use the two-step estimation, if the actions of individuals are influenced by the average of her peers.

[26]In our discussion on equilibria in the general information structure in Appendix E, we show that under some conditions, all equilibria is in a compact set. Since the set of zeros of the continuous function, $S(\cdot; X, W_n)$, is closed. The set of equilibria is itself compact.

[27]As for the Maximum Theorem, see Stokey et al.(1989) for a proof when criterion, $\Pi(\cdot, X, W_n)$, and the constraint, $S(\cdot; X, W_n)$, are functions defined on Euclidean spaces. A proof for general metric spaces can be found in Wikipedia. Seehttp://en.wikipedia.org/wiki/Maximum_theorem.

References

AGUIRREGABIRIA, V. and MIRA, P. 2002. "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models". *Econometrica*, 70(4):1519–1543.

AGUIRREGABIRIA, V. and MIRA, P. 2013. "Identification of Games of Incomplete Information with Multiple Equilibria and Common Observed Heterogeneity". *working paper.*

ARADILLA-LOPEZ, A. 2010. "Semiparametric Estimation of A Simultaneous Game with Incomplete Information". *Journal of Econometrics*, 157:409–431.

BAJARI, P., HONG, H., KRAINER, J., and NEKIPELOV, D. 2010a. "Estimating Static Models of Strategic Interactions". *Journal of Business and Economic Statistics*, 28(4):469–482.

BAJARI, P., HONG, H., KRAINER, J., and NEKIPELOV, D. 2010b. "Computing Equilibria in Static Games of Incomplete Information Using the All-Solution Homotopy". *working paper.*

BAJARI, P., HONG, H., and RYAN, S. 2010c. "Identification and Estimation of A Discrete Game of Complete Information". *Econometrica*, 78(5):1529–1568.

BALLESTER, C., CALVÓ-ARMENGOL, H., and ZENOU, Y. 2010. "Who is Who in Networks. Wanted: the Key Player". *Econometrica*, 74(5):1403–1417.

BISIN, A., MORO, A., and TOPA, G. 2011. "The Empirical Content of Models with Multiple Equilibria in Economies with Social Interactions". *working paper.*

BLEVINS, J. 2014. Nonparametric Identification of Dynamic Decision Processes with Discrete and Continuous Choices. *Quantitative Economics, forthcoming.*

BONSALL, F. F. 1962. *"Lectures on Some Fixed Point Theorems of Functional Analysis".* Tata Institute of Fundemental Research, Bombay.

BORKOVSKY, R., DORASZELSKI, U., and KRYUKOV, Y. 2010a. "A User's Guide to Solving Dynamic Stochastic Games Using the Homotopy Method". *Operations Research*, 58 (4):1116–1132.

BORKOVSKY, R., DORASZELSKI, U., KRYUKOV, Y., and SATTERTHWAITH, M. 2010b. “Learning-by-doing, Organizational Forgetting, and Industry Dynamics”. *Econometrica*, 78 (2):453–508.

BROCK, W. and DURLAUF, S. 2000. “Estimating Static Models of Strategic Interactions”. In HECKMAN, K. and LEAMER, E., editors, *Handbook of Econometrics: Edition 1*, pages 3297–3380. Elsevier.

BROCK, W. and DURLAUF, S. 2001. “Discrete Choice with Social Interactions”. *The Review of Economic Studies*, 42:430–445.

BROCK, W. and DURLAUF, S. 2003. “Multinomial Choice with Social Interactions”. *working paper.*

BROCK, W. and DURLAUF, S. 2007. “Identification of Binary Choice Models with Social Interactions”. *Journal of Econometrics*, 140:52–75.

BROOKS, J. and DINCULEANU, N. 1979. “Conditional Expectations and Weak and Strong Compactness in Spaces of Bochner Integrable Functions”. *Journal of Multivariate Analysis*, 9:420–427.

DEBREU, G. 1970. “Economies with a Finite Set of Equilibria”. *Econometrica*, 38:387–392.

DIERKER, E. 1972. “Two Remarks on the Number of Equilibria of an Economy”. *Econometrica*, 40:951–953.

DUNFORD, N. and SCHWARTZ, J. 1958. *“Linear Operators, Part I”*. Interscience, New York.

FOLLAND, G. B. 1999. *“Real Analysis: Modern Techniques and Their Applications”*. John Wiley & Sons, INC., New York, 2nd edition.

GARCIA, C. B. and ZANGWILL, W. I. 1981. *Pathways to Solutions, Fixed Points, and Equilibria.* Prentice-Hall Series in Computational Mathematics.

GASPAR, J. and JUDD, K. 1997. “Solving Large-Scale Rational Expectations Models”. *Macroeconomic Dynamics*, 68(2):235–260.

GUILLEMIN, V. and POLLACK, A. 1974. *"Differential Topology"*. Prentice-Hall, Inc., New Jersy.

HARSANYI, J. C. 1967a. "Games with Incomplete Information Played by 'Bayesian' players, parts i". *Management Science*, 14(3):159–182.

HARSANYI, J. C. 1967b. "Games with Incomplete Information Played by 'Bayesian' players, parts ii". *Management Science*, 14(3):320–334.

HOTZ, V. J. and MILLER, R. A. 1993. "Conditional Choice Probabilities and the Estimation of Dynamic Models". *The Review of Economic Studies*, 60:497–529.

HSIEH, C. S. and LEE, L. 2011. "A Social Interactions Model with Endogenous Friendship Formation and Selectivity". *working paper*.

JUDD, K. 1998. *"Numerical Methods in Economics"*. MIT Press, Cambridge, MA.

KHAN, M. A. and SUN, Y. N. 1995. "Pure Strategies in Games with Private Information". *Journal of Mathematical Economics*, 24:633–653.

KHAN, M. A. and SUN, Y. N. 2002. "Non-cooperative Games with Many Players". In AUMANN, R. J. and HART, S., editors, *Handbook of Game Theory with Economic Applications: Edition 1*, volume 3, pages 1761–1808. Elsevier.

KHAN, M. A. and ZHANG, Y. C. 2014. "On the Existence of Pure-strategy Equilibria in Games with Private Information: A Complete Characterization". *Journal of Mathematical Economics*, 50:197–202.

LAX, D., P. 2002. *"Functional Analysis"*. John Wilen&Sons, Inc., New York.

LEE, L. 2000. "A Numerically Stable Quadrature Procedure for the One-Factor Random Component Discrete Choice Model". *Journal of Econometrics*, 95(1):117–129.

LEE, L. 2001. "Interpolation, Quadrature and Stochastic Integration". *Econometric Theory*, 17:933–961.

LEE, L. 2007. "Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects". *Journal of Econometrics*, 140:333–374.

LEE, L. and LI, J. 2009. "Binary Choice under Social Interactions: An Empirical Study With and Without Subjective Data on Expectations". *Journal of Applied Econometrics*, 24: 257–281.

LEE, L., LI, J., and LIN, X. 2014. "Binary Choice Models with Social Network under Heterogeneous Rational Expectations". *Review of Economics and Statistics.* forthcoming.

LEUNG, M. 2015. "Two-Step Estimation of Network Formation Models with Incomplete Information". *Journal of Econometrics*, 188(1):182–195.

MANSKI, C. 1985. "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator". *Journal of Econometrics*, 27:313–333.

MANSKI, C. 1987. "Semiparametric Analysis of Random Effects: Linear Models from Binary Penel Data". *Econometrica*, 55:357–362.

MANSKI, C. 1993. "Identification of Endogenous Social Effects: The Reflection Problem". *The Review of Economic Studies*, 60:531–542.

MANSKI, C. 2000. "Economic Analysis of Social Interactions". *Journal of Economic Perspectives*, 14:115–136.

MANSKI, C. 2004. "Measuring Expectations". *Econometrica*, 72:1329–1376.

MAS-COLELL, A., D., W. M., and GREEN, J. 1995. *"Microeconomic Theory"*. Oxford University Press.

MELE, A. 2010. "A Structure Model of Segregation in Social Networks". *CeMMAP working paper.*

MILGROM, P. R. and WEBER, R. J. 1985. "Distributional Strategies for Games with Incomplete Information". *Mathematics of Operations Research*, 10:619–632.

OSBORNE, M. J. and RUBINSTEIN, A. 1994. *"A Course in Game Theory"*. MIT Press.

QU, X. and LEE, L. 2013. "LM Tests for Spatial Correlation in Spatial Models with Limited Dependent Variables". *Regional Science and Urban Economics*, 42:430–445.

RADNER, R. and ROSENTHAL, R. W. 1982. "Private Information and Pure-strategy Equilibria". *Mathematics of Operations Research*, 7:401–409.

ROTHENBERG, T. J. 1971. "Identification in Parametric Models". *Econometrica*, 39(3): 577–591.

RUST, J. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher". *Econometrica*, 55(5):999–1033.

RUSTON, A. F. 1986. *"Fredholm Theory in Banach Spaces"*. Cambridge University Press.

SIMON, J. 1987. "Commpact Sets in the Space $L^p(0,T;B)$". *Annali di Matematica para ed applicata*, CXLVI(IV):65–96.

STOKEY, N. L., LUCAS, E. J., R., and PRESCOTT, E. C. 1989. *"Recursive Methods in Economic Dynamics"*. Harvard University Press.

TAMER, E. 2003. "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria". *The Review of Economic Studies*, 70(1):147–165.

van NEERVEN, J. 2014. "Compactness in the Lebesgue-Boucher spaces $L^p(\mu;X)$". *Indagationes Mathematicae*, 25:389–394.

WATSON, L., BILLUPS, S., and MORGAN, A. 1987. "HOMPACK: A Suite of Codes for Globally Convergent Homotopy Algorithms". *ACM Transactions on Mathematical Software*, 13:281–310.

WATSON, L., SOSONKINA, M., MELVILLE, R., MORGAN, A., and WALKER, H. 1997. "Algorithm777: HOMPACK90:A Suite of Fortran 90 Codes for Globally Convergent Homotopy Algorithms". *ACM Transactions on Mathematical Software*, 23:514–549.

YANG, C. 2014. "Sample Selection with Social Interactions". *working paper.*

YANG, C. and LEE, L. 2017. "Social Interactions under Incomplete Information with Heterogeneous Expectations". *Joumal of Econometrics*, forthcoming.

YANG, C., QU, X., and LEE, L. 2016. "Tobit Models with Social Interactions: Complete vs Incomplete Information". *working paper.*

Table 1: Binary Choice I: Estimation Comparison

		I		II		III	
True parameters		$G = 100$	$G = 200$	$G = 100$	$G = 200$	$G = 100$	$G = 200$
β_0	0	0.9376	0.9409	0.1851	0.1792	0.0030	0.0056
		(0.2174)	(0.1566)	(0.2127)	(0.1470)	(0.2537)	(0.1773)
β_1	1	0.7397	0.7470	0.9237	0.9281	1.0210	1.0165
		(0.0957)	(0.0662)	(0.1050)	(0.0674)	(0.1299)	(0.0848)
β_2	1	0.7957	0.7859	0.9103	0.9280	0.9935	1.0116
		(0.4130)	(0.3037)	(0.4138)	(0.2889)	(0.4648)	(0.3134)
λ	0.8			0.6263	0.6266	0.8259	0.8113
				(0.0037)	(0.0004)	(0.1047)	(0.0833)
$m \log L$		-0.3288	-0.3288	-0.2401	-0.2407	-0.2356	-0.2371
		(0.0258)	(0.0180)	(0.0218)	(0.0150)	(0.0231)	(0.0165)
$\overline{\lvert\lambda\rvert}$				-	-	0.6267	0.6267
						(4.4465×10^{-16})	(4.4465×10^{-16})
r_{unctr}						0.6744	0.6735
						(0.0496)	(0.0343)

Note: Regression I corresponds to the conventional regression without social interactions. Regression II and II allow for interactions through social relations. Regression II imposes a sufficient condition on λ, assumes equilibrium uniqueness, and uses the method of contraction mapping iteration for equilibrium computation. Regression III does not impose restrictions on the interaction intensity, λ, and uses the homotopy continuation method for equilibrium computation. $\overline{\lvert\lambda\rvert}$ is the upper bound on the intensity of social interactions, corresponding to the sufficient condition for contraction mapping in Yang and Lee(2017). r_{unctr} denotes the proportion of groups in which that condition is violated under the true parameter values.

Table 2: Binary Choice II: Estimation Comparison for Moderate Interactions

True parameters		I	II	III	IV
β_0	0	0.0298	-0.0010	-0.0011	0.0014
		(0.0741)	(0.0483)	(0.0484)	(0.0493)
β_1	1	1.0322	1.0219	1.0219	1.0241
		(0.0976)	(0.1004)	(0.1005)	(0.0979)
β_2	1	1.2152	0.9906	0.9894	1.0175
		(0.3356)	(0.2666)	(0.2678)	(0.2848)
λ	0.2		0.1962	0.1962	0.1707
			(0.0294)	(0.0294)	(0.0695)
$m \log L$		-0.4804	-0.4594	-0.4594	-0.4620
		(0.0247)	(0.0246)	(0.0246)	(0.0264)
$\overline{\lvert\lambda\rvert}$					0.3133
					(1.6758×10^{-16})
r_{unctr}					0
					(0)
n_e					1
					(0)
r_m					0
					(0)
$\widehat{n}_e$					1
					(0)
$\widehat{r}_m$					0
					(0)

Note: In each simulation, the number of independent groups is $G = 100$. The population of every group is $n = 5$. Regression I corresponds to the conventional regression without social interactions. Regressions II, III, and IV take social interactions into account. Regressions II and III assumes equilibrium uniqueness. Regression II restricts the interaction intensity to satisfy the sufficient condition in Yang and Lee(2017) and uses contraction mapping iterations to solve for the equilibrium. Regression III does not restrict the interaction intensity and computes the equilibrium through solving nonlinear equations by the Newton's method. Regression IV allows for equilibrium multiplicity, uses the homotopy continuation method to solve for the equilibrium set, and select an equilibrium according to the expected total utilities to complete the model. $\overline{|\lambda|}$ stands for the upper bound on interaction intensity which ensures equilibrium uniqueness in a sample according to Yang and Lee(2017). r_{unctr} represents the proportion of groups which violate that sufficient condition. n_e and $\widehat{n}_e$ refer to respectively the average sample and estimated number of equilibria in a group. r_m and $\widehat{r}_m$ report the sample and estimated proportion of groups with multiple equilibria respectively. Numbers in parentheses are standard deviations.

Table 3: Binary Choice II: Estimation Comparison for Large Interactions

True parameters		I	II	III	IV
β_0	0	0.0899	0.0021	-0.0042	0.0439
		(0.1339)	(0.0431)	(0.0525)	(0.0694)
β_1	1	0.4767	0.5191	0.5748	1.0544
		(0.0525)	(0.0636)	(0.2164)	(0.2492)
β_2	1	1.1328	0.6596	0.6197	0.9346
		(0.4606)	(0.2650)	(0.3941)	(0.3442)
λ	0.8		0.3109	0.3900	0.8325
			(0.0039)	(0.0277)	(0.0968)
α	1				1.0776
					(0.6221)
$m \log L$		-0.6031	-0.4312	-0.3989	-0.1218
		(0.0200)	(0.0349)	(0.0477)	(0.0235)
$\overline{\lvert\lambda\rvert}$					0.3133
					(2.2232×10^{-16})
r_{unctr}					1
					(0)
n_e					2.2918
					(0.0784)
r_m					0.7811
					(0.0407)
$\widehat{n}_e$					2.2487
					(0.1716)
$\widehat{r}_m$					0.7642
					(0.1015)

Note: In each simulation, the number of independent groups is $G = 100$. The population of every group is $n = 5$. Regression I corresponds to the conventional regression without social interactions. Regressions II, III, and IV take social interactions into account. Regressions II and III assumes equilibrium uniqueness. Regression II restricts the interaction intensity to satisfy the sufficient condition in Yang and Lee(2017) and uses contraction mapping iterations to solve for the equilibrium. Regression III does not restrict the interaction intensity and computes the equilibrium through solving nonlinear equations by the Newton's method. Regression IV allows for equilibrium multiplicity, uses the homotopy continuation method to solve for the equilibrium set, and select an equilibrium according to the expected total utilities to complete the model. $\overline{\lvert\lambda\rvert}$ stands for the upper bound on interaction intensity which ensures equilibrium uniqueness in a sample according to Yang and Lee(2017). r_{unctr} represents the proportion of groups which violate that sufficient condition. n_e and $\widehat{n}_e$ refer to respectively the average sample and estimated number of equilibria in a group. r_m and $\widehat{r}_m$ report the sample and estimated proportion of groups with multiple equilibria respectively. Numbers in parentheses are standard deviations.

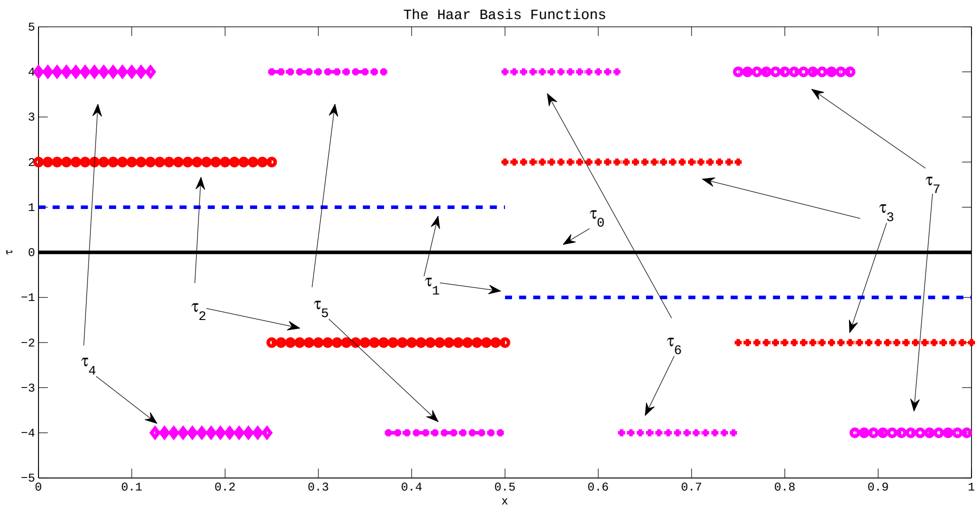

Figure 1: The Haar Basis Functions

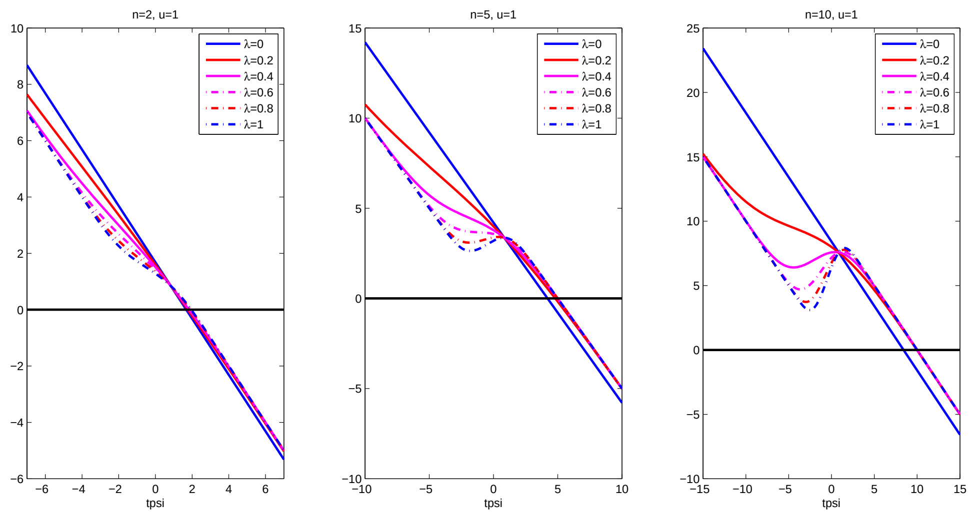

Figure 2: Equilibrium Illustration for Binary Choice I with Influences from Peers A

Note: In this figure, "tpsi" refers to the expected total outcomes in a group of symmetric members for the Binary Choice Model I with influences from peers. For each value of λ, an equilibrium expected total outcome is a zero of a nonlinear function, whose graph is depicted as a curve. The characteristics of the equilibrium set may differ as the group population, n, and homogeneous individual utility, u, vary. The left, middle, and right diagrams respectively correspond to three cases: $n = 2$, $u = 2$; $n = 5$, $u = 2$; and $n = 10$, $u = 2$.

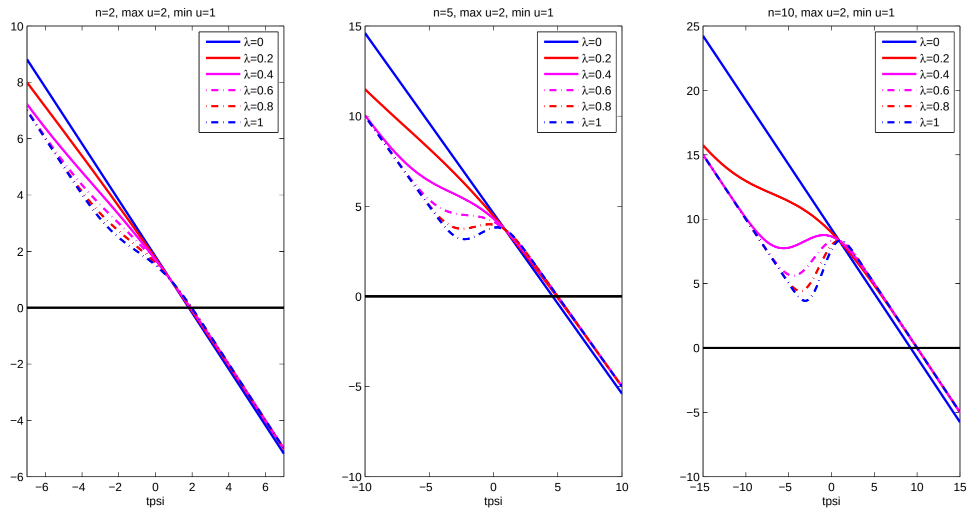

Figure 3: Equilibrium Illustration for Binary Choice I with Influences from Peers B

Note: In this figure, "tpsi" refers to the expected total outcomes in a group of asymmetric members for the Binary Choice Model I with Influences from peers. For each value of λ, an equilibrium expected total outcome is a zero of a nonlinear function, whose graph is depicted as a curve. The characteristics of the equilibrium set may differ as the group population, n, and heterogeneous individual utilities, u_i's, vary. The left, middle, and right diagrams respectively correspond to three cases: $n = 2$, $\max_i u_i = 2$, $\min_i u_i = 1$; $n = 5$, $\max_i u_i = 2$, $\min_i u_i = 1$; and $n = 10$, $\max_i u_i = 2$, $\min_i u_i = 1$.

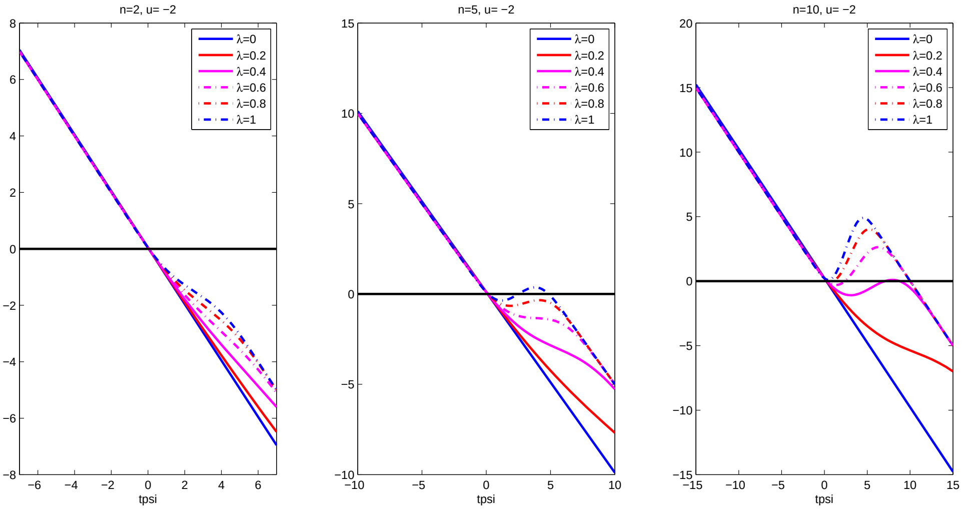

Figure 4: Equilibrium Illustration for Binary Choice I with Influences from Peers C

Note: In this figure, "tpsi" refers to the expected total outcomes in a group of symmetric members for the Binary Choice Model I with influences from peers. For each value of λ, an equilibrium expected total outcome is a zero of a nonlinear function, whose graph is depicted as a curve. The characteristics of the equilibrium set may differ as the group population, n, and homogeneous individual utility, u, vary. The left, middle, and right diagrams respectively correspond to three cases: $n = 2$, $u = -2$; $n = 5$, $u = -2$; and $n = 10$, $u = -2$.

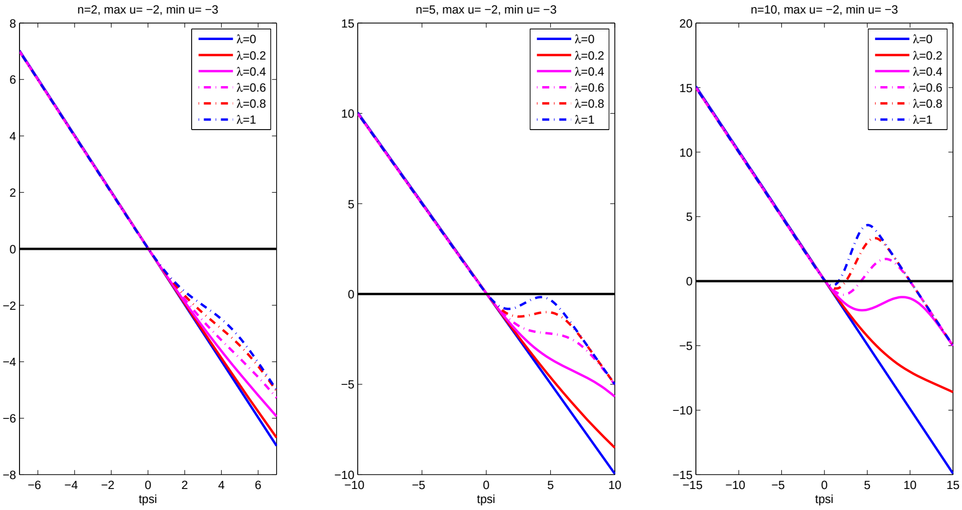

Figure 5: Equilibrium Illustration for Binary Choice I with Influences from Peers D

Note: In this figure, "tpsi" refers to the expected total outcomes in a group of asymmetric members for the Binary Choice Model I with Influences from peers. For each value of λ, an equilibrium expected total outcome is a zero of a nonlinear function, whose graph is depicted as a curve. The characteristics of the equilibrium set may differ as the group population, n, and heterogeneous individual utilities, u_i's, vary. The left, middle, and right diagrams respectively correspond to three cases: $n = 2$, $\max_i u_i = -2$, $\min_i u_i = -3$; $n = 5$, $\max_i u_i = -2$, $\min_i u_i = -3$; and $n = 10$, $\max_i u_i = -2$, $\min_i u_i = -3$.

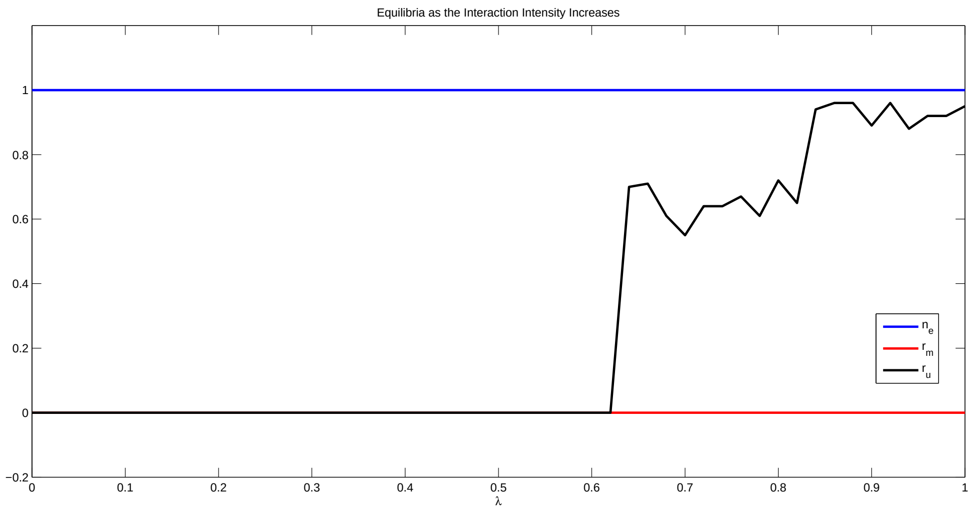

Figure 6: Equilibrium Illustration for Binary Choice I with General Social Relations

Note: This figure shows the features of the equilibrium set in Binary Choice Model I as the interaction intensity, λ, increases, for a sample of $G = 100$ groups with homogeneous group population $n = 5$. n_e represents the average number of equilibria of the sample. r_m is the ratio of groups with more than one equilibria. r_u stands for the proportion of groups whose social relation matrix does not satisfy the sufficient condition for contraction mapping in Yang and Lee(2017).

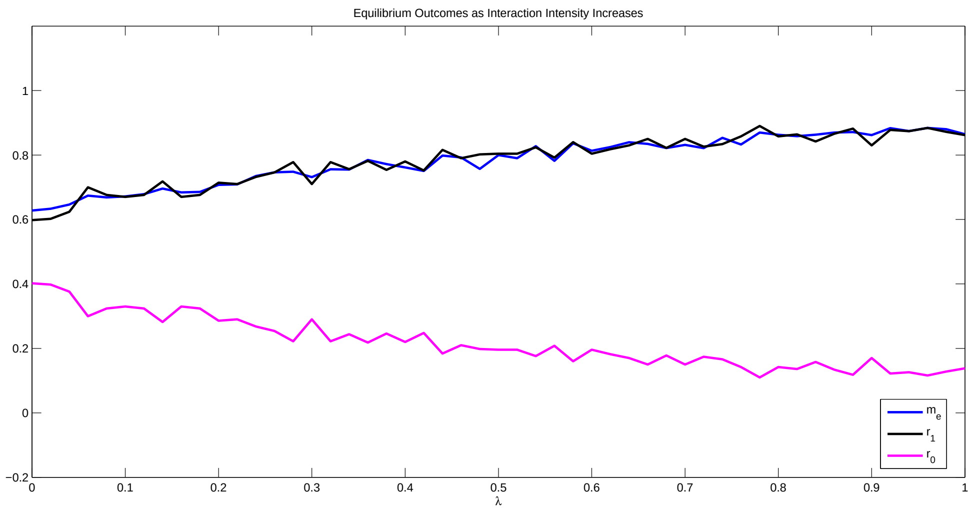

Figure 7: Equilibrium Outcomes for Binary Choice I with General Social Relations

Note: This figure shows the features of the equilibrium outcomes in Binary Choice Model I as the interaction intensity, λ, increases, for a sample of $G = 100$ groups with homogeneous group population $n = 5$. m_e, r_1, and r_0, respectively represent the average expected (individual) outcomes, the ratio of agents who choose "1", and the ratio of agents who choose "0", of the unique equilibrium.

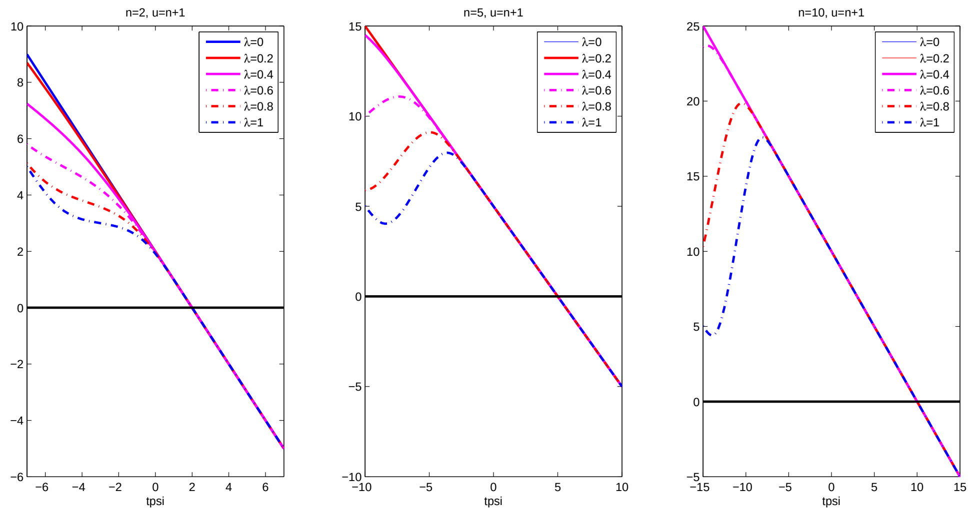

Figure 8: Equilibrium Illustration for Binary Choice II with Influences from Peers A

Note: In this figure, "tpsi" refers to the expected total outcomes in a group of symmetric members for the Binary Choice Model II with influences from peers. For each value of λ, an equilibrium expected total outcome is a zero of a nonlinear function, whose graph is depicted as a curve. The characteristics of the equilibrium set may differ as the group population, n, and homogeneous individual utility, u, vary. The left, middle, and right diagrams respectively correspond to three cases: $n = 2$, $u = n + 1$; $n = 5$, $u = n + 1$; and $n = 10$, $u = n + 1$.

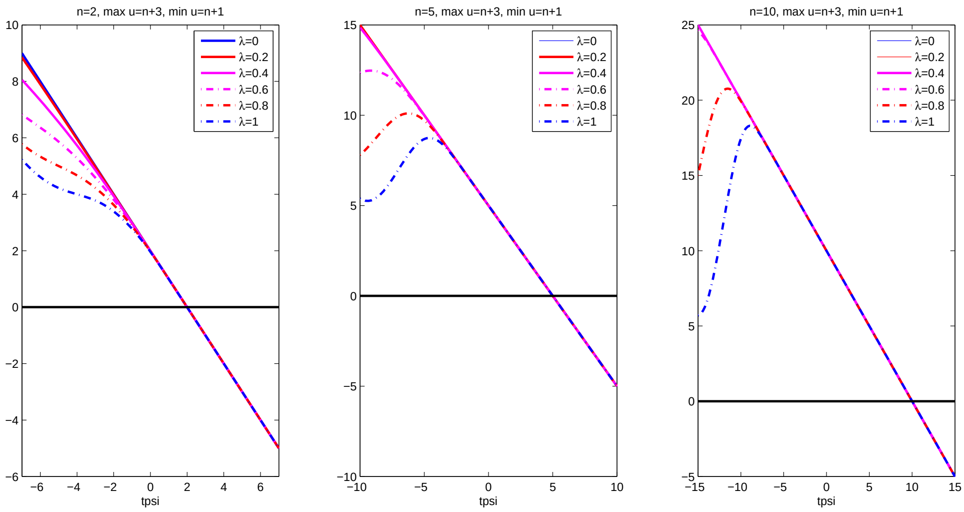

Figure 9: Equilibrium Illustration for Binary Choice II with Influences from Peers B

Note: In this figure, "tpsi" refers to the expected total outcomes in a group of asymmetric members for the Binary Choice Model II with Influences from peers. For each value of λ, an equilibrium expected total outcome is a zero of a nonlinear function, whose graph is depicted as a curve. The characteristics of the equilibrium set may differ as the group population, n, and heterogeneous individual utilities, u_i's, vary. The left, middle, and right diagrams respectively correspond to three cases: $n = 2$, $\max_i u_i = n + 3$, $\min_i u_i = n + 1$; $n = 5$, $\max_i u_i = n + 3$, $\min_i u_i = n + 1$; and $n = 10$, $\max_i u_i = n + 3$, $\min_i u_i = n + 1$.

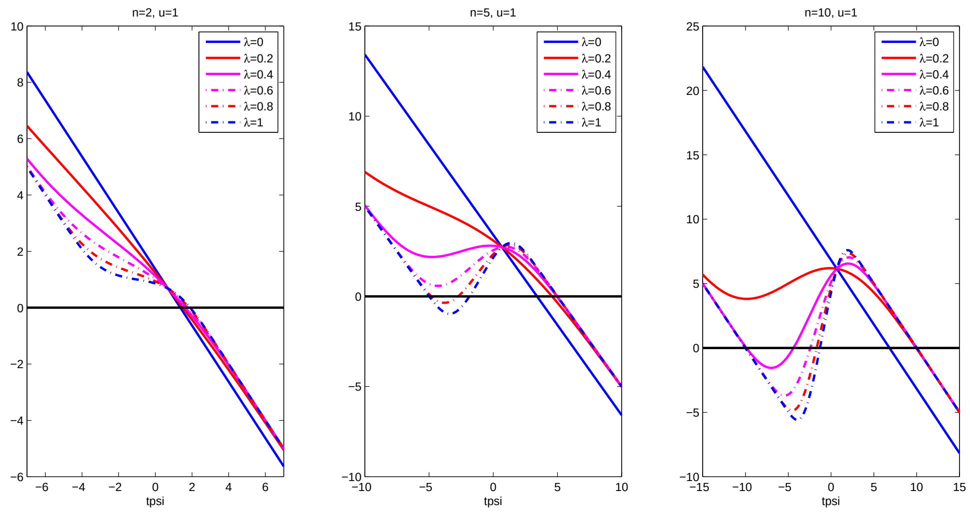

Figure 10: Equilibrium Illustration for Binary Choice II with Influences from Peers C

Note: In this figure, "tpsi" refers to the expected total outcomes in a group of symmetric members for the Binary Choice Model II with influences from peers. For each value of λ, an equilibrium expected total outcome is a zero of a nonlinear function, whose graph is depicted as a curve. The characteristics of the equilibrium set may differ as the group population, n, and homogeneous individual utility, u, vary. The left, middle, and right diagrams respectively correspond to three cases: $n = 2$, $u = 1$; $n = 5$, $u = 1$; and $n = 10$, $u = 1$.

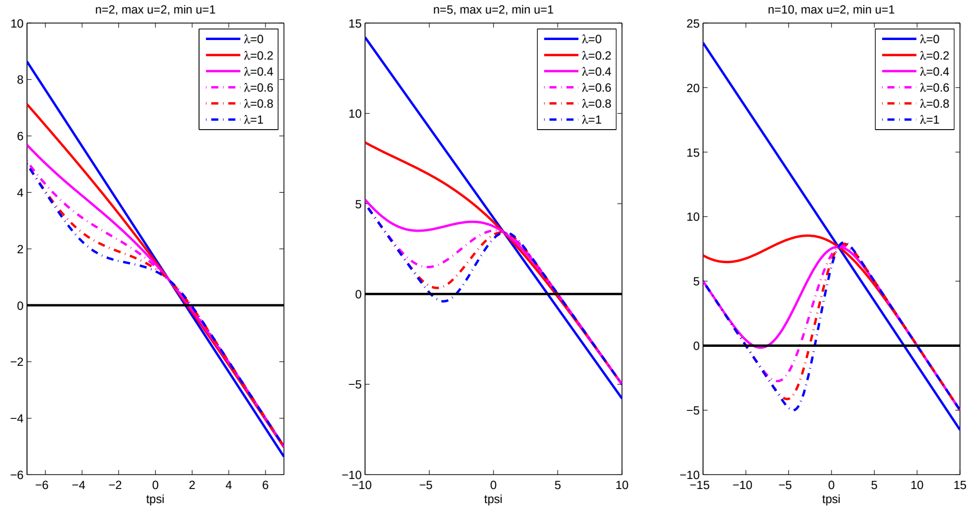

Figure 11: Equilibrium Illustration for Binary Choice II with Influences from Peers D

Note: In this figure, "tpsi" refers to the expected total outcomes in a group of asymmetric members for the Binary Choice Model II with Influences from peers. For each value of λ, an equilibrium expected total outcome is a zero of a nonlinear function, whose graph is depicted as a curve. The characteristics of the equilibrium set may differ as the group population, n, and heterogeneous individual utilities, u_i's, vary. The left, middle, and right diagrams respectively correspond to three cases: $n = 2$, $\max_i u_i = 2$, $\min_i u_i = 1$; $n = 5$, $\max_i u_i = 2$, $\min_i u_i = 1$; and $n = 10$, $\max_i u_i = 2$, $\min_i u_i = 1$.

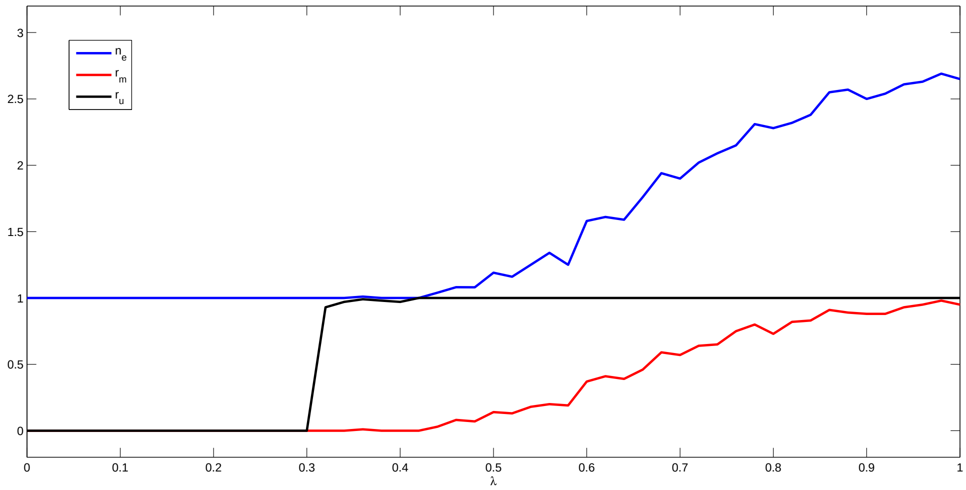

Figure 12: Equilibrium Illustration for Binary Choice II with General Social Relations

Note: This figure shows the features of the equilibrium set in Binary Choice Model II as the interaction intensity, λ, increases, for a sample of $G = 100$ groups with homogeneous group population $n = 5$. n_e represents the average number of equilibria of the sample. r_m is the ratio of groups with more than one equilibria. r_u stands for the proportion of groups whose social relation matrix does not satisfy the sufficient condition for contraction mapping in Yang and Lee(2017).

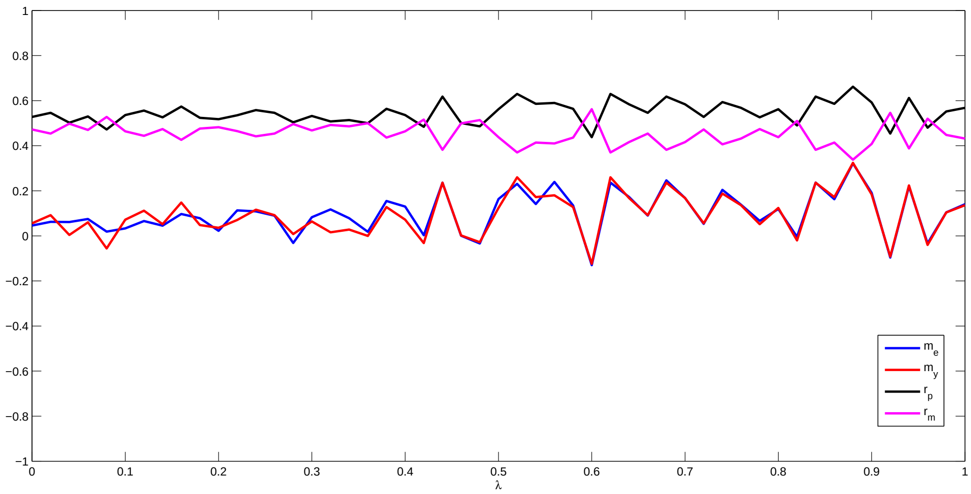

Figure 13: Equilibrium Outcomes for Binary Choice II with General Social Relations

Note: This figure shows the features of the equilibrium outcomes in Binary Choice Model II as the interaction intensity, λ, increases, for a sample of $G = 100$ groups with homogeneous group population $n = 5$. m_e, m_y, r_p, and r_m, respectively represent the average expected (individual) outcomes, the average individual choices, the ratio of agents who choose "1", and the ratio of agents who choose "-1", of the unique/selected equilibrium.

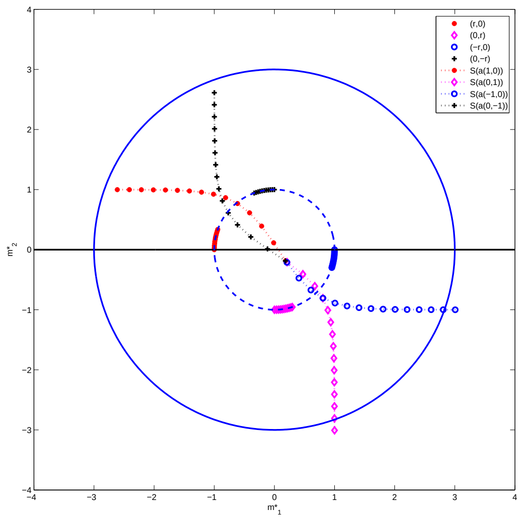

Figure 14: Homotopic Mappings on Sphere for Binary Choices