A Tobit Model with Social Interactions under Incomplete Information*

Chao Yang†, Lung-fei Lee‡, and Xi Qu§

November 30, 2015

SUMMARY

We build a model to analyze censored behaviors for socially associated agents under a general form of incomplete information. Employing the analytical tools used by Yang and Lee(2014), we find that there is a unique equilibrium when parameters are within a reasonable range. We utilize the rate of censoring and the average observed outcomes to identify model parameters. We suggest estimating the model by the maximum likelihood method with inner-loop fixed point iterations, which performs well in Monte Carlo experiments. We apply the model to investigate municipal property tax rates in North Carolina and find significant competition effects.

JEL-Classification: C31, C35, C57

Keywords: Social Interactions, Tobit Model, Incomplete Information, Gaussian Quadrature

*We owe our thanks to Ohio Supercomputer Center for supporting high performance computation.

†School of Economics, Shanghai University of Finance and Economics, 325 The Building of the School of Economics, 777 Guoding Rd, Shanghai, 200433 China, *Tel.*+86(0)21-6590-2704. *Email*: yang.chao@mail.shufe.edu.cn

‡Department of Economics, The Ohio State University, 475 Arps hall, 1945 N. High Street, Columbus, OH 43210, USA. *Tel.* 614-292-5508. *Email*: lee.1777@osu.edu.

§Department of Economics, Antai College of Economics and Management, Shanghai Jiaotong University, 1954 Huashan Rd. Shanghai, 200030 China, *Tel.* +86(0)21-5230-1595. *Email*: xiqu@sjtu.edu.cn.

1 Introduction

Truncated and/or censored outcomes are frequently observed in practice, when agents face binding constraints in decision, in particular, the nonnegative constraint. As the most popular way to model this kind of limited dependent variables, the Tobit model has gained much attention for both empirical and theoretical reasons. For example, Kumar(2012) proposes an extension of nonparametric estimation methods for nonlinear budget-set models to censored dependent variables. Abreyava and Shen (2014) consider estimation of censored panel-data models with individual-specific slope heterogeneity. In this paper, we are interested in modeling the censored outcomes associated with social interactions where an agent's behavior may be affected by other group members. There are two types of models about social interactions. In one stream of the literature, agents who move simultaneously have complete information about all exogenous characteristics and shocks. As a result, each agent's behavior is influenced by the actual behaviors of others in that social group. See Lee(2007) and Boucher et al.(2014) for example. Other models are built on incomplete information, so an agent's actions are affected by her expectations on the behaviors of other agents in a social group. For instance, suppose that all exogenous characteristics are public information and idiosyncratic shocks are privately known, Manski (1993) studies linear models about socially interacted continuous choices; and Brock and Durlauf (2001) and Lee et al.(2014) investigate binary choices for socially linked agents. In their recent work, Yang and Lee (2014) extend previous research to a general form of incomplete information. They find that by allowing incomplete information in not only unobserved idiosyncratic shocks but also exogenous individual characteristics, conditional expectations about agents' behaviors can be functions of private information. Their discussions are mainly about continuous choices and binary choices. In this paper, we analyze social interactions of censored outcomes under a general form of incomplete information and study the tax rate competition in local government of North Carolina.

The basis of our model is a simultaneous-move Bayesian game when players' actions are bounded from below by no action so that a corner solution to the expected utility maximization problem is possible. We find that there is a one-to-one correspondence between a BNE and a profile of conditional expectation functions consistent with strategies. Similar to Yang and Lee(2014), we can embed those conditional expectation functions into a normed space and transform an equilibrium conditional expectation function into a fixed point of an operator.

Additionally, a sufficient condition that ensures the operator to be a contraction mapping holds for the Tobit model in terms of a reasonable range of the parameter of interest. That range corresponds to weak or moderate social interactions, but not strong ones. Strong social interactions might demonstrate multiple (expectation) equilibria which generate stable and unstable systems as it is shown for the binary choices models by Brock and Durlauf(2001). Focusing on the case of a unique equilibrium, we discuss computation, identification, and estimation of the model.

We explore the information which is unique in the Tobit model to help identify the variance of the idiosyncratic shocks. Different from models with continuous and binary choices where either continuous or discrete outcomes are observed, for the Tobit model with censored outcomes, two types of information are available from data: whether an outcome is censored or observed. With both types of information, in addition to coefficients of explanatory variables, it is possible to identify the variance of idiosyncratic shocks; while that is not possible for binary choice models.

By investigating interacted behaviors under a general form of incomplete information, in addition to models with networks and social interactions, our model is also related to the literature of estimation of games, such as the work of Bajari et al(2010) and Aradillas (2010). Bajari et al (2010) focus on estimation in a framework where all variables are public information except that each agent only gets the realization of her own shocks. Incomplete information about both exogenous characteristics and idiosyncratic shocks is discussed by Aradillas (2010). However, only a two-agent binary choice game is considered and only equilibria based on public information are discussed. Therefore, heterogeneity in social relations and in private information is missing there.

For estimation, similar to Rust (1987), we nest the fixed point iteration in the maximum likelihood estimation to derive parameter estimates. That nested fixed point algorithm works well in Monte Carlo experiments. For a sample with a large number groups, we also consider group common features (unobserved by econometricians) as random effects to investigate unobserved group heterogeneity. The incorporation of group common features is important in social interactions in order to capture correlation effects as in Manski(1993) and Moffitt(2000). Since group common features are observed by group members, given their realizations, equilibrium conditional expectations can be solved in a similar way. However, econometricians need to integrate over those random effects in estimation. Utilizing stochastic simulation estimation method, we can calculate a stochastic integration and estimate parameters using the nested fixed point maximum likelihood method.

As an empirical application, we study the property tax rates for contiguous municipalities in North Carolina. Tax competition among local governments has been theoretically and empirically studied (See Brueckner(2003) for a comprehensive review). Most research considers the tax rate as a continuous variable. However, it is more appropriate to adopt the Tobit model as tax rates are non-negative and local governments' choices are subject to this non-negative constraint. More recently, Porto and Revelli(2013) evaluate three empirical approaches to the analysis of spatially dependent limited tax policies. Their Tobit type models are based on interactions with latent variables and/or spatial time lags, which are different from ours. We model the property tax rate as an outcome of the Bayesian Nash Equilibrium from a static simultaneous-move game with incomplete information and estimate the corresponding parameters using the nested fixed point maximum likelihood method proposed in this paper. For the sample of municipal property tax rates, we find the existence of strong competition among near-by municipalities in North Carolina.

The paper proceeds as follows. In Section 2, we build the model and explain it as a reduced-form equilibrium in a simultaneous-move game under incomplete information with censored outcomes. In Section 3, we provide sufficient conditions for the existence of a unique equilibrium and discuss its computation under different forms of incomplete information. In the subsequent two sections, Section 4 and Section 5, we discuss identification and estimation of model parameters. Following an extension to allow group unobservables in Section 6, we investigate the finite sample performance of the nested fixed point maximum likelihood estimation by Monte Carlo experiments in Section 7 and study the tax competition among municipalities in North Carolina in Section 8. Section 9 concludes.

2 The Model

2.1 The Model Framework

Consider a group of n agents, $i = 1, \cdots, n$, who are socially linked. Their social relations are represented by an $n \times n$ weighting matrix, W_n, such that for all $i \neq j$, its (i, j) entry, $W_{n,ij} > 0$, if i connects with j; and $W_{n,ij} = 0$ otherwise. The diagonal elements, $W_{n,ii} = 0$ for all $i = 1, \cdots, n$.

W_n may be either symmetric or asymmetric. For example, W_n may represent a friendship network. Then $W_{n,ij} = 1$ if i views j as one of her friends; and $W_{n,ij} = 0$ otherwise. When friendship is mutual, the network is undirected and W_n is symmetric. However, if friendship is not mutual, it is possible that i regards j as one of her friends, but j does not think i is among her good friends. Then $W_{n,ij} \neq W_{n,ji}$ and W_n is asymmetric. In spatial econometrics, an example is the relative strength of spatial interactions among counties. In that case, $W_{n,ij}$ may be the reciprocal of the geographic distance between two different counties, $i \neq j$. Formulated in this way, W_n is symmetric. However, once it is row-normalized, W_n would in general be asymmetric.

We model the social interactions of agents' outcomes subject to nonnegative constraint, y_i's, under incomplete information:

$$y_i = \max\left\{\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j\neq i} W_{n,ij} E[y_j | X^p_{J_i}, Z] - \epsilon_i, 0\right\}. \tag{2.1}$$

This model distinguishes our work from the classical Tobit model by incorporating social interactions. From (2.1), we can see that j's behavior can affect i only if i links to j, i.e., $W_{n,ij} \neq 0$. Moreover, that impact is through i's expectation on j based on her information about various personal characteristics represented by X. That is suitable for the case when agents take actions simultaneously while some information is unknown. Therefore, similar to Yang and Lee (2014), we classify observable exogenous characteristics into three categories, the group features, X^g; commonly known individual characteristics, $X^c = (X_1^{c'}, \cdots, X_n^{c'})'$; and some personal traits which may be privately known, $X^p = (X_1^{p'}, \cdots, X_n^{p'})'$. To describe information structures in a general form, for any $i = 1, \cdots, n$, we use an $n \times 1$ vector J_i to represent her private information about X^p. That is, $J_i(j) = 1$ if X^p_j is known by i; and $J_i(j) = 0$ otherwise. $X^p_{j_i}$ then represents the random vector composed of those X^p_j's that i knows. In (2.1), for simplicity, Z collects all public information, including the group features, X^g, commonly known individual traits, X^c, the social relations, W_n, as well as the information structure, $J_1, \cdots, J_n$.

We assume that idiosyncratic shocks ϵ_i's are i.i.d. with the pdf, $f_\epsilon(\cdot)$, and the corresponding CDF, $F_\epsilon(\cdot)$. These idiosyncratic shocks are also independent of all exogenous characteristics and network connections.

2.2 Game Theoretical Foundation

Our model is related to a simultaneous-move game under incomplete information where the values of actions are bounded below by zero to satisfy the nonnegative constraint. For example, for investment on community services, although a resident may prefer to disinvest, a negative amount of investment is not possible. Similarly, for competitive stock brokers, portfolio choices will also be restricted if short-sales are not allowed. We use the $n \times n$ matrix W_n to represent social relations among a group of n players. Denote the action taken by agent i by a_i. Assume that $a_i \geq 0$. Her payoff is determined by the following equation:[1]

$$r(a_i, a_{-i}, X^g, X_i^c, X_i^p) = \alpha - \gamma(a_i - \beta_0 - X_i^{c'}\beta_1 - X_i^{p'}\beta_2 - X^{g'}\beta_3 - \lambda \sum_{j \neq i} W_{n,ij}a_j + \epsilon_i)^2. \quad (2.2)$$

As before, X^g denotes group features, X_i^c refers to publicly known individual characteristics, X_i^p stands for privately known personal traits. The idiosyncratic shocks, ϵ_i, is revealed only to i. They are i.i.d. and are independent of observable exogenous variables, i.e., X and W. The private information structure about X^p is also represented by vectors, $J_1, \cdots, J_n$. Following Harsanyi (1967a, 1967b), we interpret incomplete information by unknown "types". Let $\mathcal{S}_i$ represent the support of X_i^p and $\mathcal{T}$ denote the common support of ϵ_i's. The set of states, $\prod_i^n \mathcal{S}_i \times \mathcal{T}^n$, is defined as the set of possible values of X_i^p's and ϵ_i's for all players. In this case, player i's "type" is her private information about exogenous characteristics and shocks, $X_{J_i}^p$ and ϵ_i. Hence, her type set is the corresponding support, $\mathcal{R}_i = \prod_{k:J_i(k)=1} \mathcal{S}_k \times \mathcal{T}$. The signal function is a mapping from the states to her type, $\tau_i : \prod_i^n \mathcal{S}_i \times \mathcal{T}^n \to \prod_{k:J_i(k)=1} \mathcal{S}_k \times \mathcal{T}$. Her prior belief on the set of states is the joint distribution of X_i^p's, and the distribution for shocks. The prior belief is the same across all players. A strategy is a plan specifying an action for each possible realization of types. That is, $s_i : \mathcal{R}_i \to \mathcal{A}_i$, where $\mathcal{A}_i$ is i's set of actions.

Because players move at the same time, they do not know which actions others will take when they are making their own decisions. They can only form expectations based on their

[1] There are some other specifications which can imply linear equations for socially interacted outcomes, see Durlauf(2000), Ballester, Calvó-Armengol and Zenou(2006), and Tao and Lee(2014).

private information. The expected payoff by taking action a_i is as follows:

$$\begin{aligned} &E[r(a_i, a_{-i}, X^g, X_i^c, X_i^p)|X_{J_i}^p, Z, \epsilon_i] \\ =&\alpha - \gamma(a_i - \beta_0 - X_i^{c'}\beta_1 - X_i^{p'}\beta_2 - X^{g'}\beta_3 - \lambda \sum_{j\neq i} W_{n,ij}E[a_j|X_{J_i}^p, Z] + \epsilon_i)^2 \\ &+ \gamma\lambda^2\Big((\sum_{j\neq i} W_{n,ij}E[a_j|X_{J_i}^p, Z])^2 - E[(\sum_{j\neq i} W_{n,ij}a_j)^2|X_{J_i}^p, Z]\Big), \end{aligned} \tag{2.3}$$

where the expectation on a_j does not depends on ϵ_i, because ϵ_i's are independent of each other and also independent of X and W_n. Suppose that $\gamma > 0$, if there is no restrictions on a_i, the best response is to choose the ideal value,

$$a_i^* = \beta_0 + X^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j\neq i} W_{n,ij}E[a_j|X_{J_i}^p, Z] - \epsilon_i.$$

However, with the constraint, $a_i \geq 0$, it is possible to have corner solutions. The optimal action is $a_i = \max\{a_i^*, 0\}$. Thus, $(s_1(X_{J_1}^p, \epsilon_1), \cdots, s_n(X_{J_n}^p, \epsilon_n))$ is a Bayesian Nash Equilibrium (BNE) if

$$s_i(X_{J_i}^p, \epsilon_i) = \max\left\{\beta_0 + X^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j\neq i} W_{n,ij}E[s_j(X_{J_i}^p, \epsilon_i)|X_{J_i}^p, Z] - \epsilon_i, 0\right\}. \tag{2.4}$$

Therefore, y_i's in our model, (2.1), can be viewed as the realizations of actions in a BNE.

3 Equilibrium Analysis

3.1 Equilibrium and Expectations

Similar to Yang and Lee(2014), we employ the conditional expectations on agents' behaviors as tools to analyze the model. From (2.1), fix $Z = z$, for any $k \neq i$, we have that

$$E[y_i|X_{J_k}^p, z] = E[H(\beta_0 + X^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j\neq i} W_{n,ij}E[y_j|X_{J_i}^p, Z])|X_{J_k}^p, z], \tag{3.1}$$

where $H(\cdot)$ is a real-valued function such that for any $x \in \Re^1$,

$$H(x) = \int_{-\infty}^{+\infty} I(x > c)(x - c)f_\epsilon(c)dc = xF_\epsilon(x) - \int_{c<x} cf_\epsilon(c)dc. \tag{3.2}$$

We can see that for any two different agents, $k \neq k'$, their predictions on the behavior of a third agent, i, with $i \neq k$ and $i \neq k'$, will be the same, if they have the same private information about X^p, i.e., $J_k = J_{k'}$. Given an information structure, J_k, $X_{J_k}^p$ is a random vector, composed of X_j^p's such that $J_k(j) = 1$. Denote the underlying sample space for X^p, by $(\Omega, \mathfrak{F}, \mathcal{P})$, each X_j^p

is a mapping from the sample space to $(\Re^{k_p}, \mathfrak{B}^{k_p})$.[2] So $X^p_{J_k}$ is a function from the sample space to $(\Re^{M_k}, \mathfrak{B}^{m_k})$, where $M_k = (\sum_{j=1}^n J_k(j))k_p$. Therefore, as a function of $X^p_{J_k}$, given public information $Z = z$, for any realization ω, $E[y_i|X^p_{J_k}, z]$ will return a real number, $E[y_i|X^p_{J_k} = X^p_{J_k}(\omega), z]$. That is to say, fixing an information structure, J_k, the conditional expectation is a random variable. Therefore, summarizing all possible information structures that are relevant to predict i's behaviors, we can describe conditional expectations on i as a mapping which returns a random variable for each relevant vector of privately known characteristics, $\psi_i : \mathfrak{A}_i \to \mathfrak{C}$, where $\mathfrak{A}_i = \left\{ X^p_{J_k} : W_{n,ki} \neq 0 \right\}$ and $\mathfrak{C}$ is the set of all random variables on $(\Omega, \mathfrak{F}, P)$. Collecting expectations for all group members, we get a vector of those functions, $\psi : \prod_i^n \mathfrak{A}_i \to \mathfrak{C}^n$, such that $\psi(A_1, \cdots, A_n) = (\psi_1(A_1), \cdots, \psi_n(A_n))'$, for any $(A_1, \cdots, A_n)' \in \prod_i^n \mathfrak{A}_i$. We call it the "conditional expectation function".[3] In a BNE, agents' predictions are consistent with others' strategies. As a result, (3.1) implies the following "consistency condition":

$$\psi_i(A_i) = E[H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j(X^p_{J_i}))|A_i, z], \tag{3.3}$$

for all $i = 1, \cdots, n$, $A = (A_1, \cdots, A_n)' \in \prod_i^n \mathfrak{A}_i$. If ψ satisfies (3.3), we say that it is an equilibrium conditional expectation function and denote it as ψ^e.

Now we related the conditional expectation function to the game on which our Tobit model is based.

Proposition 3.1 *Conditional on public information, $Z = z$, if $(s_1(X^p_{J_1}, \epsilon_1), \cdots, s_n(X^p_{J_n}, \epsilon_n))$ is a BNE of the Bayesian game in Section 2.2, i.e., it satisfies* (2.4), *there is a conditional expectation function, $\psi : \prod_i^n \mathfrak{A}_i \to \mathfrak{C}^n$, such that* (3.3) *holds. Conversely, if $\psi : \prod_i^n \mathfrak{A}_i \to \mathfrak{C}^n$ satisfies* (3.3), *there is a BNE, $(s_1(X^p_{J_1}, \epsilon_1), \cdots, s_n(X^p_{J_n}, \epsilon_n))$.*

Proof. See Appendix A. ■

Owing to Proposition 3.1, there is a one-to-one correspondence between a BNE and an equilibrium conditional expectation function. Hence, we can investigate the existence, uniqueness, and calculation of the BNE through inspecting conditional expectations.

[2] k_p is the dimension of X^p_i.

[3] If $W_{k,i} = 0$ for all i, the expected value of i's behavior does not affect those of other agents. Then expectations about y_i is redundant in determining the equilibrium. To facilitate discussion in general, we define conditional expectation function for every agent, keeping in mind that only agents who are connected with others are relevant in the system.

3.2 Unique Equilibrium

According to Yang and Lee (2014), we can view each conditional expectation function as a point in a function space and transform an equilibrium conditional expectation function into a fixed point of an operator in that space. Particularly, let Ξ be a set of functions such that any $\xi \in \Xi$ maps a profile of random vectors $A = (A_1, \cdots, A_n) \in \prod_{i=1}^n \mathfrak{A}_i$ to a random vector in $\mathfrak{C}^n$ with two properties,

$$\xi(A) = (\xi_1(A_1), \cdots, \xi_n(A_n))',$$

$$\max_{1 \leq i \leq n} \max_{\{j: W_{j,i} \neq 0\}} \int |\xi_i(x^p_{J_j})| dF_p(x^p) < \infty, \tag{3.4}$$

where F_p is a simplified notation for the conditional distribution of X^p given public information $Z = z$. In Yang and Lee(2014), it is shown that the function, $\|\cdot\| : \Xi \to \Re_+$, with

$$\|\xi\| = \max_{1 \leq i \leq n} \max_{\{j: W_{j,i} \neq 0\}} \int |\xi_i(x^p_{J_j})| dF_p(x^p_{J_i}),$$

for any $\xi \in \Xi$ is a well-defined norm and $(\Xi, \|\cdot\|)$ is a complete metric space.

Based on our model for censored behaviors, define an operator on Ξ, T, such that for any $\xi \in \Xi$, and $A = (A_1, \cdots, A_n) \in \prod_{i=1}^n \mathfrak{A}_i$,

$$T(\xi)(A)_i = T(\xi)_i(A_i) = E[H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\xi_j(X^p_{J_i}))|A_i, z], \tag{3.5}$$

where $H(\cdot)$ is given in (3.2). To disuss equilibrium conditional expectations in Ξ, we need the image of T to be always in Ξ. To ensure that property, we make the following assumption on the distribution of the idiosyncratic shocks.

Assumption 3.1 $E[\epsilon_i] < \infty$ *for any* $i = 1, \cdots, n$.

For any $i = 1, \cdots, n$, denote

$$g_i(X^p_{J_i}) = \beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\xi_j(X^p_{J_i}).$$

By (3.4), $\int |\xi_j(X^p_{J_i})| dF_p < \infty$ uniformly for all $j \neq i$, given public information $Z = z$. $g_i(X^p_{J_i})$ is a sum of constant and several integrable functions. Thus, $\int |g_i(X^p_{J_i})| dF_p < \infty$ uniformly for all i. For any $k \neq i$ with $W_{n,ki} \neq 0$, we have that

$$
\begin{aligned}
&\int |T(\xi)_i(X^p_{J_k})| dF_p \\
&= \int |g_i(X^p_{J_i}) F_\epsilon(g_i(X^p_{J_k})) - \int_{c<g_i(X^p_{J_i})} c f_\epsilon(c) dc| dF_p \\
&\leq \int |g_i(X^p_{J_i})| dF_p + \int | \int_{c<g_i(X^p_{J_i})} c f_\epsilon(c) dc | dF_p \\
&\leq \int |g_i(X^p_{J_i})| dF_p + E[|\epsilon|],
\end{aligned}
$$

where the second inequality follows from Assumption 3.1.[4] Therefore, $\max_i \max_{k \neq i} \int |T(\xi)_i(X^p_{J_k})| dF_p < \infty$. Thus, whenever $\xi \in \Xi$, $T(\xi) \in \Xi$.

We can see that if $\xi \in \Xi$ is a fixed point of T, it is an equilibrium conditional expectation function. Additionally, for equilibrium conditional expectation functions which are integrable with respect to the conditional distribution of X^p given public information Z, such a function must be a fixed point of T. That is, focusing on integrable functions, there is a one-to-one correspondence between a fixed point of T and an equilibrium conditional expectation function.

Proposition 3.2 *$T : \Xi \to \Xi$ is a contraction mapping if*

$$|\lambda| \|W_n\|_\infty < 1, \tag{3.6}$$

where $\|W_n\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n W_{n,ij}$. Then there is one and only one BNE.

Proof. See Appendix A. ■

If W_n is row-normalized, $\|W_n\|_\infty = 1$. Hence, a sufficient condition for the existence of a unique equilibrium is $|\lambda| < 1$. Namely, the magnitude of influence from socially associated agents is less than 1 (in absolute value). This assumption is conventional for linear spatial autoregressive models in the literature of spatial econometrics and social interactions. With (3.6) being satisfied, we investigate solution, identification, and estimation for the model.

3.3 Equilibrium Calculation

Owing to the properties of a contraction mapping in a complete metric space, when (3.6) holds, there is an algorithm to solve the unique equilibrium. To be specific, since T is a contraction mapping under (3.6), $\lim_{l \to \infty} T^l(\xi^0) = \psi^e$, for any $\xi^0 \in \Xi$, according to the norm of $(\Xi, \|\cdot\|)$. Thus, beginning with any initial guess, by iterating the operator T, we can approximate the

[4] $E[\epsilon_i] < \infty$ if and only if $E[|\epsilon_i|] < \infty$.

equilibrium conditional expectation function, ψ^e. Then by applying Proposition 3.1, we can derive the corresponding BNE. However, in general, since $\psi^e : \prod_{i=1}^n \mathfrak{A}_i \to \mathfrak{C}$ is a function of random vectors, to solve it means to derive its realizations for every point on the underlying sample space, which is by no way a trivial task.

Nonetheless, ψ^e changes with elements in the sample space indirectly through the random vector X^p and a predetermined information structure J. Therefore, if X_i^p's are discrete random vectors with a finite support, it suffices to characterize ψ^e by its values on those points, which makes it possible to represent ψ^e by a vector of finite dimensions. Although that representation is not applicable when X_i^p's vary continuously on a continuum support, due to the use of quadrature order stochastic integrals as an approximation of the integrals generated by expectations, we can approximate every possible realization of ψ^e by a finite-dimension vector. To elucidate the idea, we begin with the simplest case and then move on to more complicated ones.

3.3.1 Publicly Known Characteristics

When all exogenous characteristics, X^g, X^c, X^p, are public information in a group, there is no uncertainty other than the idiosyncratic shocks, which are i.i.d. and independent of all exogenous characteristics. Given $Z = z$, ψ^e reduces to an $n \times 1$ vector, satisfying

$$\psi_i^e = H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e), \tag{3.7}$$

for all $i = 1, \cdots, n$. Although we cannot get an analytical solution, due to the nonlinearity of the function $H(\cdot)$, we can solve ψ^e numerically by contraction mapping iteration at each given vector of parameters.

3.3.2 Self-Known Characteristics

If X_i^p is realized only to i(and the econometricians), we call the information structure as "self-known characteristics", in which case $J_i(k) = 0$ for all $k \neq i$. Then any two different agents do not share private information. For $i \neq k$, we have that

$$\psi_i^e(X_k^p) = E[H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e(X_i^p))|X_k^p, z]. \tag{3.8}$$

Inspecting (3.8), if all X_i^p's are independent of each other conditional on $Z = z$, the realization of X_k^p does not provide new information on X_i^p given $Z = z$. That is to say,

$$\psi_i^e(X_k^p) = E[H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e(X_i^p))|z],$$

for any $i \neq k$. Since the right-hand side does not depend on the random vector X_k^p, we can view expectations on i's behaviors as a constant. Therefore, with independent self-known characteristics, the equilibrium conditional expectation is still a vector, such that

$$\psi_i^e = E[H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e)|z], \tag{3.9}$$

Comparing (3.7) and (3.9), we can see that although in both cases every two agents $k \neq k'$ have the same expectation on the behavior of a third person, i, when all exogenous characteristics are public information, k and k' just integrate over unobserved idiosyncratic shocks in (3.7). In contrast, when they know just their only realizations for X^p, they have to integrate over X_i^p to predict i's behaviors in (3.9), for X_i^p is not included in Z.

If X_i^p's are correlated, however, conditional expectations depend on the private information used to make predictions. Scrutinizing (3.8), if there are two agents k and k' with $W_{n,ki} \neq 0$ and $W_{n,k'i} \neq 0$, their private information influences predictions on i's behaviors through the conditional distributions $F_p(X_i^p|X_k^p, z)$ and $F_p(X_i^p|X_{k'}^p, z)$. Therefore, when the two conditional distributions differ, k and k' will form different conditional expectations. However, if we can provide conditions assuring that the two conditional distributions are the same, those two agents' predictions on i's behaviors will be identical once they get the same realizations. One sufficient condition is "exchangeability", cited below from Yang and Lee (2014):

Assumption 3.2 *Conditional on public information, $Z = z$, X_i^p's have the same support, $\mathfrak{S}_p$. Their conditional joint distribution, $f^p(X_1^p, \cdots, X_n^p|Z = z)$,*[5] *is exchangeable, i.e., for any permutation, $s : \{1, \cdots, n\} \to \{1, \cdots, n\}$,*

$$f^p(X_1^p, \cdots, X_n^p|Z = z) = f^p(X_{s(1)}^p, \cdots, X_{s(n)}^p|Z = z).$$

[5] When X_i^p's are discrete random variables, $f^p(\cdot|Z = z)$ is the probability mass function. If X_i^p's are continuous, $f^p(\cdot|Z = z)$ represents the density.

Under "exchangeability", if $X_k^p = X_{k'}^p = x$,

$$\begin{aligned}
\psi_i^e(X_k^p = x) =& E[H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e(X_i^p))|X_k^p = x, z] \\
=& E[H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e(X_i^p))|X_{k'}^p = x, z] \\
=& \psi_i^e(X_{k'}^p = x).
\end{aligned} \tag{3.10}$$

for any $k, k^{'}$ with $W_{n,ki} \neq 0$ and $W_{n,k^{'}i} \neq 0$. Therefore, we can directly define ψ_i^e on the common support of X_i^p's $\mathcal{S}_p$, and characterize ψ^e by

$$\psi_i^e(x) = E[H(\beta_0 + X_i^{c'}\beta_1 + y^{'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e(y))|x, z], \tag{3.11}$$

for all $i = q, \cdots, n$ and $x \in \mathcal{S}_p$. Now we consider two classes of joint distributions that satisfy Assumption 3.2.

1. (Discrete X^p) Suppose that X_i^p can only take one of m values in $\{x^l : 1 \leq l \leq m\}$, where each x^l is a vector with specific values. Given public information $Z = z$, the conditional probability function, $f_p(y|x, z)$, is fully captured by the transition matrix,

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix},$$

   where $p_{ll'} = prob(X_i^p = x^{l^{'}}|X_k^p = x^l, Z = z)$, for $k \neq i$. We can represent ψ^e by an $(nm) \times 1$ vector,

$$\psi^e = (\psi_1^e(x^1), \cdots, \psi_1^e(x^m), \cdots, \psi_n^e(x^1), \cdots, \psi_n^e(x^m))^{'},$$

   and characterize it by the following system of nonlinear equations:

$$\psi_i^e(x^l) = \sum_{\widetilde{l}=1}^{m} p_{l\widetilde{l}} H(\beta_0 + X_i^{c'}\beta_1 + x^{\widetilde{l}'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e(x^{\widetilde{l}})), \tag{3.12}$$

   for $i = 1, \cdots, n$ and $l, \widetilde{l} = 1, \cdots, m$. Beginning with any $(nm) \times 1$ vector and iterating the contraction mapping, we can derive ψ^e.

2. (Continuous X^p) Consider the case when $X_1^p, \cdots, X_n^p$ are jointly normal with mean

$(\mu', \cdots, \mu')'$, and variance-covariance matrix

$$\begin{pmatrix} \Sigma_1 & \Sigma_2 & \cdots & \Sigma_2 \\ \Sigma_2' & \Sigma_1 & \cdots & \Sigma_2 \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_2' & \Sigma_2' & \cdots & \Sigma_1 \end{pmatrix}.$$

Then for any $i \neq j$, conditional on $X_j^p = X$, X_i^p is normal with mean $\mu + \Sigma_2 \Sigma_1^{-1}(X - \mu)$ and variance $\Sigma_1 - \Sigma_2 \Sigma_1^{-1} \Sigma_2'$. In this case, each X_i^p can take any value in $\Re^{k_p}$. Thus, it is impossible to represent ψ^e by a finite-dimension vector. However, scrutinizing (3.11), we find that for any i and $x \in \Re^{k_p}$, $\psi_i^e(x)$ is determined by an integral, which can be approximated by the values of the function on a fixed number of values (quadrature points) using the quadrature method. To illustrate the idea, let us consider the special case that each X_i^p is a single random variable, i.e., its dimension $k_p = 1$. In this case, denote $\Sigma_1 = \eta^2$ and $\Sigma_2 = \rho\eta^2$. We first transform integration in $\Re^1$ into integration over a finite interval, $[-1, 1]$, and then apply the Gauss-Legendre quadrature.

$$\begin{aligned}
&\psi_i^e(x) \\
=&\int_{-\infty}^{+\infty} H(\beta_0 + X_i^{c'}\beta_1 + \widetilde{x}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e(\widetilde{x})) \\
&\cdot \frac{1}{\sqrt{2\pi(1-\rho^2)\eta^2}} \exp(-\frac{(\widetilde{x} - \rho x - (1-\rho)\mu)^2}{2(1-\rho^2)\eta^2}) d\widetilde{x} \\
=&\int_0^1 H(\beta_0 + X_i^{c'}\beta_1 + \log(\frac{z}{1-z})\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e(\log(\frac{z}{1-z}))) \\
&\cdot \frac{1}{\sqrt{2\pi(1-\rho^2)\eta^2}} \exp(-\frac{(\log(\frac{z}{1-z}) - \rho x - (1-\rho)\mu)^2}{2(1-\rho^2)\eta^2}) \frac{1}{z(1-z)} dz \\
=&\sqrt{\frac{2}{\pi(1-\rho^2)\eta^2}} \int_{-1}^1 H(\beta_0 + X_i^{c'}\beta_1 + \log(\frac{\widetilde{z}+1}{1-\widetilde{z}})\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e(\log(\frac{\widetilde{z}+1}{1-\widetilde{z}}))) \\
&\cdot \exp(-\frac{(\log(\frac{\widetilde{z}+1}{1-\widetilde{z}}) - \rho x - (1-\rho)\mu)^2}{2(1-\rho^2)\eta^2}) \frac{1}{(\widetilde{z}+1)(1-\widetilde{z})} d\widetilde{z} \\
\approx&\sqrt{\frac{2}{\pi(1-\rho^2)\eta^2}} \sum_{k=1}^K \omega_k H(\beta_0 + X_i^{c'}\beta_1 + \log(\frac{\widetilde{z}_k+1}{1-\widetilde{z}_k})\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e(\log(\frac{\widetilde{z}_k+1}{1-\widetilde{z}_k}))) \\
&\cdot \exp(-\frac{(\log(\frac{\widetilde{z}_k+1}{1-\widetilde{z}_k}) - \rho x - (1-\rho)\mu)^2}{2(1-\rho^2)\eta^2}) \frac{1}{(\widetilde{z}_k+1)(1-\widetilde{z}_k)}.
\end{aligned} \tag{3.13}$$

In (3.13), the second equality is derived by a change of integration variable, $\widetilde{x} = \log(\frac{z}{1-z})$

and the third equality comes from a transformation of $z = \frac{\widetilde{z}+1}{2}$. At last, the approximation is based on standard Gauss-Legendre quadrature, where ω_k's are the weights, $\widetilde{z}_k$'s are the corresponding abscissae, and K is the number of abscissae. Define accordingly, $x_k^p = \log(\frac{\widetilde{z}_k+1}{1-\widetilde{z}_k})$, for $k = 1, \cdots, K$, we get nK equalities,

$$\psi_i^e(x_{k'}^p) = \sqrt{\frac{2}{\pi(1-\rho^2)\eta^2}} \sum_{k=1}^{K} \omega_k H(\beta_0 + X_i^{c'}\beta_1 + x_k^p\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j\neq i} W_{n,ij}\psi_j^e(x_k^p)) \cdot \exp(-\frac{(x_k^p - \rho x_{k'}^p - (1-\rho)\mu)^2}{2(1-\rho^2)\eta^2}) \frac{1}{(\widetilde{z}_k+1)(1-\widetilde{z}_k)},$$

for all $i = 1, \cdots, n$ and $k' = 1, \cdots, K$. This is very similar to (3.12). Hence, we can solve $\psi_i^e(x_{k'}^p)$'s by contraction mapping iterations. After that, for any $x \in \Re^1$, we can approximate $\psi_i^e(x)$ by (3.13). Owing to the fast convergence of the Gauss-Legendre quadrature, we only need to take a small number of abscissae. In our Monte Carlo experiments, it is shown that good performance can be achieved in estimation when choosing $K = 8$.

When X_i^p's are of multiple dimensions, multiple-dimension quadrature methods are available but quite computationally intensive. Alternatively, we can use the stochastic integral approximation with importance sampling. Let $h(a_i)$ be a density with its support containing the support of X_i^p such that $f_p(x^p|x)/h(x^p)$ is well defined. Then we can generate K random draws, x_k^p, from $h(\cdot)$. The stochastic approximation will be

$$\psi_i^e(x) \approx \frac{1}{K}\sum_{k=1}^{K} H(\beta_0 + X_i^{c'}\beta_1 + x_k^p\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j\neq i} W_{n,ij}\psi_j^e(x_k^p)) \frac{f_p(x_k^p|x)}{h(x_k^p)}.$$

Analogous to previous discussion, we first solve $\psi^e(x_k^p)$'s by contraction mapping and then approximate the function $\psi_i^e(x)$ at any point x.

For general information structures, the unique equilibrium can be calculated in a similar way. When all X_i^p's are discrete random vectors with a finite support, we can fully solve ψ^e directly via contraction mapping iteration. When X_i^p's are continuous random variables, we choose a finite number of points and approximate the integration in conditional expectations by a weighted sum. If continuous X_i^p's are not exchangeable, for the stochastic simulation, the simulated values from important densities can be different for each agent. However, if the number of simulated points is the same, say K, for each i, the total number of equations for solution can remain to be nK. Since values of ψ^e on those finite number of variables can be solved as a vector by contraction mapping iteration, the values of ψ^e for all realizations can be approximated.

4 Identification

For identification, we aim at recovering the model primitives, β, λ, F_x and F_ϵ, from observations on X, Y, and W_n.

Definition 4.1 *$(\beta, \lambda, F_x, F_\epsilon)$ is observationally equivalent to $(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{F}_x, \widetilde{F}_\epsilon)$ for social connections W_n at $\overline{W}_n$ and information structure J at $\overline{J}$, if they generate the same distribution of the observables,*

$$F_{Y,X|\overline{W}_n,\overline{J}}(\cdot,\cdot|\beta,\lambda,F_x,F_\epsilon) = F_{Y,X|\overline{W}_n,\overline{J}}(\cdot,\cdot|\widetilde{\beta},\widetilde{\lambda},\widetilde{F}_x,\widetilde{F}_\epsilon). \tag{4.1}$$

If (4.1) holds for X at $\overline{X}$, $(\beta, \lambda, F_x, F_\epsilon)$ is observationally equivalently to $(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{F}_x, \widetilde{F}_\epsilon)$ at $\overline{W}_n$, $\overline{J}$ and $\overline{X}$.

The observational equivalence in terms of distribution, (4.1) implies, in particular, the same conditional expectations, i.e.,

$$E[y_i|X^p_{\overline{J}_k}, z, \beta, \lambda, F_x, F_\epsilon] = E[y_i|X^p_{\overline{J}_k}, z, \widetilde{\beta}, \widetilde{\lambda}, \widetilde{F}_x, \widetilde{F}_\epsilon]. \tag{4.2}$$

Definition 4.2 *Suppose that $(\beta^*, \lambda^*, F^*_x, F^*_\epsilon)$ is the true parameters for Y, X at $\overline{W}_n$ and $\overline{J}$. $(\beta^*, \lambda^*, F^*_x, F^*_\epsilon)$ is identified if any $(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{F}_x, \widetilde{F}_\epsilon) \neq (\beta^*, \lambda^*, F^*_x, F^*_\epsilon)$ cannot be observationally equivalent to $(\beta^*, \lambda^*, F^*_x, F^*_\epsilon)$ at $\overline{W}_n$ and $\overline{J}$.*

Our discussion about identification is based on the hypothesis:

Assumption 4.1 *The distribution of exogenous characteristics, $F_x(\cdot)$, can be inferred from data about X.*

So we can focus on β, λ, and F_ϵ. Additionally, we parametrize the distribution of ϵ_i's.

Assumption 4.2 *ϵ_i's are i.i.d. with the full support, $\Re^1$, according to a parametric pdf, $f_\epsilon(\cdot;\sigma)$, where the functional form, $f_\epsilon(\cdot;\cdot)$ is known but the parameter value of σ is unknown. The corresponding CDF is $F_\epsilon(\cdot;\sigma)$, which is strictly increasing in its argument.*

In the following Lemma 4.1, we show that σ can be identified from the relationship between the mean outcomes, $E[y_i|X^p, z]$, and the average amount of censoring, $E[I(y_i > 0)|X^p, z]$, where X^p refers to the matrix of all privately known characteristics, $(X_1^{p'}, \cdots, X_n^{p'})'$.

Lemma 4.1 *Given public information, $Z = z$, for all $i = 1, \cdots, n$,*

$$E[y_i|X^p, z] = E[I(y_i > 0)|X^p, z]F_\epsilon^{-1}(E[I(y_i > 0)|X^p, z]; \sigma) - \int_{c<F^{-1}(E[I(y_i>0)|X^p,z];\sigma)} c f_\epsilon(c; \sigma) dc. \tag{4.3}$$

Proof. See Appendix B. ■

Since (4.3) holds for an arbitrarily chosen agent, we can suppress the subscript and simply write

$$E[y|X^p, z] = E[I(y > 0)|X^p, z]F_\epsilon^{-1}(E[I(y > 0)|X^p, z]; \sigma) - \int_{c<F_\epsilon^{-1}(E[I(y>0)|X^p,z];\sigma)} c f_\epsilon(c; \sigma) dc. \tag{4.3$'$}$$

In addition to the relation (4.3$'$), we impose the following two assumptions for the identification of σ.

Assumption 4.3 *$f_\epsilon(c; \sigma)$ is differentiable with respect to σ. Moreover,* $\lim_{c\to-\infty} c\frac{\partial F_\epsilon(c;\sigma)}{d\sigma} = 0$.

Assumption 4.4 *The ratio, $\frac{\frac{\partial F_\epsilon(c;\sigma)}{\partial\sigma}}{f_\epsilon(c;\sigma)}$, is strictly monotonic with respect to c.*

Proposition 4.1 *For any network, W_n, and information structure, J, if Assumptions 4.1 to 4.4 are satisfied, σ can be identified from moments, $E[y|X^p, z]$ and $E[I(y > 0)|X^p, z]$.*

Proof. See Appendix B. ■

The proof of Proposition 4.1 depends on the relationship, (4.3$'$), which is valid with any information structure on X. In principle, with appropriate empirical observations, $E[y|X^p, z]$ and $E[I(y^* > 0)|X^p, z]$ can be identified nonparameterically from empirical observations. Then we can identify σ for any information structure.

As an example of Assumptions 4.3 and 4.4, consider the case that ϵ_i is normally distributed with zero mean and standard deviation σ. Then, $F_\epsilon(c; \sigma) = \Phi(c/\sigma)$ and $f_\epsilon(c; \sigma) = \frac{1}{\sigma}\phi(\frac{c}{\sigma})$, where $\Phi(\cdot)$ and $\phi(\cdot)$ are respectively the CDF and pdf of the standard normal distribution. $\lim_{c\to-\infty} c\frac{\partial F_\epsilon(c;\sigma)}{\partial\sigma} = \lim_{c\to-\infty} -(\frac{c}{\sigma})^2\phi(\frac{c}{\sigma}) = 0$ and $\frac{\frac{\partial F_\epsilon(c;\sigma)}{\partial\sigma}}{f_\epsilon(c;\sigma)} = \frac{-\frac{c}{\sigma^2}\phi(\frac{c}{\sigma})}{\frac{1}{\sigma}\phi(\frac{c}{\sigma})} = -\frac{c}{\sigma}$, which is decreasing in c. Therefore, the sufficient conditions in Assumptions 4.3 and 4.4 are satisfied.

Actually, it is more transparent to see the identification of the standard deviation in the normal disturbance case through (4.3$'$). By calculation, we can get that
$E[y|X^p, z] = \sigma\Big(\Phi^{-1}(E[I(y > 0)|X^p, z])E[I(y > 0)|X^p, z] + \phi(\Phi^{-1}(E[I(y > 0)|X^p, z]))\Big)$,
which implies that

$$\sigma = \frac{E[y|X^p, z]}{\Phi^{-1}(E[I(y > 0)|X^p, z])E[I(y > 0)|X^p, z] + \phi(\Phi^{-1}(E[I(y > 0)|X^p, z]))}. \tag{4.4}$$

Now we turn to identification of other parameters. For a single group, any group characteristics, X^g, is absorbed by the constant term. So we focus on the identification of β_0, β_1, β_2 and λ. We impose two additional assumptions.

Assumption 4.5 *Given an information structure, $J = \overline{J}$, $X^c = \overline{X}^c$, $X^p = \overline{X}^p$, and $W = \overline{W}_n$, $E[Y_i|J = \overline{J}, W = \overline{W}_n, X^c = \overline{X}^c, X^p_{J_k} = \overline{X}^p_{J_k}]$ can be identified (nonparametrically), for any $i, k = 1, \cdots, n$, $i \neq k$, and $W_{n,ki} \neq 0$.*

Assumption 4.6 $\begin{pmatrix} l_n, & \overline{X}^c, & \overline{X}^p, & E^p \end{pmatrix}$ *has full column rank, where l_n is an $n \times 1$ vector of 1's and E^p is an $n \times 1$ vector, whose i-th component is $\sum_{j\neq i} \overline{W}_{n,ij} E[Y_j|J = \overline{J}, W = \overline{W}_n, X^c = \overline{X}^c, X^p_{J_i} = \overline{X}^p_{J_i}]$.*

The full rank condition in Assumption 4.6 is essential to rule out multicollinearity of regressors, similar to conventional linear regressions.

Proposition 4.2 *For a single group, for a social network matrix, $\overline{W}_n$, information structure, $\overline{J}$, and personal characteristics, $\overline{X}^c$ and $\overline{X}^p$, if Assumptions 4.1 to 4.6 hold, we can identify β_0, β_1, β_2, λ, and σ.*

Proof. See Appendix B. ■

When there are multiple independent groups in the sample, we can identify β_3 from the variation of group features, X^g.

5 Estimation

Because all exogenous characteristics are observed in the data, from econometricians' point of view, randomness in agents' behaviors come from the idiosyncratic shocks, ϵ_i's. As those shocks are independent across agents, the likelihood for behaviors of agents are independent of each other. Therefore, the sample log likelihood can be written as follows:

$$\begin{aligned} &\log L(Y|X^c, X^p, X^g, W_n) \\ =&\sum_{i=1}^{n} \big(I(y_i > 0) \log f_\epsilon(y_i - (\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j\neq i} W_{n,ij} E[y_j|X^p_{J_i}, z]); \sigma) \\ &+ I(y_i = 0) \log F_\epsilon(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j\neq i} W_{n,ij} E[y_j|X^p_{J_i}, z]; \sigma)\big). \end{aligned} \tag{5.1}$$

In (5.1), the conditional expectations, $E[y_j|X^p_{J_i}, z]$'s, are determined by the equilibrium condition, (3.1). Therefore, we need first solve those conditional expectations in order to calculate the sample likelihood. Since the conditional expectation for the whole group is the fixed point of a contraction mapping, we can solve it by contraction mapping iteration for every parameter vector and then choose the parameter vector to maximize the sample log likelihood.[6] That is, we nest a fixed point solution algorithm in the maximum likelihood estimation, similar to Rust (1987).

Because the conditional likelihoods of agents' behaviors are independent of each other and the joint likelihood of a whole sample is the product of marginal likelihoods, we can derive large sample properties of the estimator in a conventional way for independent observations.

6 Extension

In practice, there are some possible common factors which can influence all group members but are unknown to econometricians. For example, when making decisions on tax rates, municipal officers know the lobbying power of different parties. Nonetheless, there might not be a measure about that from data sets. Such factors are group unobservables. In this paper, we model group unobservables as random effects.

To be specific, consider G independent groups. For a group g, in addition to X^g, $X^{c,g}$, and $X^{p,g}$, there is another set of group features lumped together into a variable, ω^g, which is unobserved. Assume that ω_g is independent of other variables and has a zero mean. The observed censored outcomes satisfy

$$y_{i,g} = \max\left\{\beta_0 + X^{c'}_{i,g}\beta_1 + X^{p'}_{i,g}\beta_2 + X^{g'}\beta_3 + \omega^g + \lambda\sum_{j\neq i} W_{n_g,ij}E[y_{j,g}|X^{p,g}_{J_{i,g}}, Z] - \epsilon_{i,g}, 0\right\}. \tag{2.1$'$}$$

Because ω^g is public information for agents, its presence will be similar to X^g in the analysis of equilibrium expectation and behaviors. The distribution of ω^g can be identified through variation across different groups. However, when unobservable group random effects are taken into account, our estimation method will need to be modified. Since ω^g is unobservable to

[6] In Section 3.3, we discuss in detail how to solve the unique equilibrium under various circumstances.

econometricians, we integrate over it to construct the sample likelihood function:

$$\begin{aligned}
&\log L(Y|X^c, X^p, X^g, W) \\
=&\sum_{g=1}^{G} \log \Big[\int \prod_{i=1}^{n_g} \big(f_\epsilon(y_{i,g} - (\beta_0 + X_{i,g}^{c'}\beta_1 + X_{i,g}^{p'}\beta_2 + X^{g'}\beta_3 + \omega^g + \lambda \sum_{j\neq i} W_{n_g,ij} E[y_{j,g}|X_{J_{i,g}}^{p,g}, z]); \sigma)^{I(y_{i,g}>0)} \\
&\cdot F_\epsilon(\beta_0 + X_{i,g}^{c'}\beta_1 + X_{i,g}^{p'}\beta_2 + X^{g'}\beta_3 + \omega^g + \lambda \sum_{j\neq i} W_{n_g,ij} E[y_{j,g}|X_{J_{i,g}}^{p,g}, z]; \sigma)^{I(y_{i,g})=0} \big) f_\omega(\omega^g) d\omega^g \Big].
\end{aligned} \tag{6.1}$$

In estimation, we use stochastic integration to approximate the integration over the unobserved group random effects, ω^g's. That is, we derive S independent draws, $\omega^{g,s}$, for $s = 1, \cdots, S$ for each group $g = 1, \cdots, G$ from the density $f_\omega(\cdot; \gamma)$, to construct a simulated sample log likelihood:

$$\begin{aligned}
&\log \overline{L}(Y|X^c, X^p, X^g, W) \\
=&\sum_{g=1}^{G} \log \Big[\frac{1}{S} \sum_{s=1}^{S} \prod_{i=1}^{n_g} \big(f_\epsilon(y_{i,g} - (\beta_0 + X_{i,g}^{c'}\beta_1 + X_{i,g}^{p'}\beta_2 + X^{g'}\beta_3 + \omega^{g,s} + \lambda \sum_{j\neq i} W_{n_g,ij} E[y_{j,g}|X_{J_{i,g}}^{p,g}, z]); \sigma)^{I(y_{i,g}>0)} \\
&\cdot F_\epsilon(\beta_0 + X_{i,g}^{c'}\beta_1 + X_{i,g}^{p'}\beta_2 + X^{g'}\beta_3 + \omega^{g,s} + \lambda \sum_{j\neq i} W_{n_g,ij} E[y_{j,g}|X_{J_{i,g}}^{p,g}, z]; \sigma)^{I(y_{i,g})=0} \big) \Big].
\end{aligned} \tag{6.1$'$}$$

In estimation, we will still nest fixed point iteration in a maximum likelihood estimation algorithm, replacing the true likelihood, (6.1), with the simulated one according to (6.1$'$).

7 Monte Carlo Experiments

We investigate finite sample performance of the nested fixed point maximum likelihood estimation via Monte Carlo experiments. In our experiments, the observed group feature, X^g, is absent, for simplicity. The idiosyncratic shocks, ϵ_i's, are i.i.d. with a pdf, $(1/\sigma)\phi(\cdot/\sigma)$, where $\phi(\cdot)$ is the standard normal density. We focus on the estimation of the coefficient of the intercept, β_0, the individual commonly known characteristics, β_1, private personal features, β_2, the interactions from socially associated agents, λ, and the standard deviation of the idiosyncratic shocks, σ. Their true values are $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda = 0.3$, and $\sigma = 1$.

We suppose agents are linked in social networks, which are represented by the matrix W_n, where n is the population size of the whole network. For any two agents, $i \neq j$, $W_{n,ij} = 1$ if i links to j; and $W_{n,ij} = 0$ otherwise. $W_{n,ii} = 0$ for all $i = 1, \cdots, n$. Then we row-normalize W_n

such that the sum of each row is equal to 1. Two types of network designs are considered. In the first case, the whole sample is composed of a collection of independent groups. Consequently, W_n can be organized as a block-diagonal matrix, with each block representing social relations in one group. We assume that all groups have the same size, 20. Within a group, for every agent, 3 other agents are randomly selected to be linked to her, which is represented as $F = 3$ in reported tables of estimates. The number of groups, G, is either 100 or 500. In the second network design, there is a single group of socially associated agents with its population size being either $n = 200$ or $n = 1000$. As for social relations, we make experiments on two settings. First, we consider the case that the number of friends an agent can make, F, is fixed. We begin with $F = 30$ for the network of size $n = 200$. When the population size increases to $n = 1000$, we look at $F = 30$ and $F = 150$. That is, we compare the estimates when the number of social links is kept constant with that when social links increase proportionally with population size. After that, we turn to the case where the number of social links an agent can make is random. It can take any integer value between 0 and an upper bound UF with equal probabilities. That is, we use social links generated by a uniform discrete distribution. For population size $n = 200$, the upper limit is $UF = 59$ so that the mean is near 30, the same as the fixed number of social links for group size $n = 200$ in the previous case. For $n = 1000$, we investigate two cases, $UF = 59$ and $UF = 299$, corresponding to the settings for fixed number of social links with a group of size 1000. In this way, we try to find how the randomness of social links influences the correlation among agents' behaviors and the performance of estimators.

The commonly known individual characteristics, X^c, are generated as independent variables. For X_i^p, we consider two cases, corresponding to our discussion in Section 3.3. That is, they are either discretely distributed with a finite support or continuously distributed with a continuum support. In both cases, X_i^p's are correlated with an "exchangeable" joint distribution. For the first case, we adopt the simplest setting that each X_i^p is dichotomous, taking a value of 0 or 1. The realizations of X_i^p's are determined as follows: 40% of the agents in a group are picked randomly. People who are picked have $X_i^p = 1$; and those who are not selected have $X_i^p = 0$. For this design, the distribution of X_i^p's does not depends on the identities of agents, with a transition (conditional) probability matrix:

$$P = \begin{pmatrix} P(X_2^p = 1|X_1^p = 1) & Pr(X_2^p = 0|X_1^p = 1) \\ P(X_2^p = 1|X_1^p = 0) & Pr(X_2^p = 0|X_1^p = 0) \end{pmatrix} = \begin{pmatrix} \frac{n^1-1}{n-1} & \frac{n-n^1}{n-1} \\ \frac{n^1}{n-1} & \frac{n-n^1-1}{n-1} \end{pmatrix},$$

where n is the number of agents in the group and $n^1 = 0.4n$ is the number of agents who are picked in the group. Members know the joint distribution of X_i^p's. Because both n and n^1 are observable, econometricians can infer X_i^p's joint distribution from data. So we focus on the likelihood of y_i's conditional on regressors for estimation.

When X_i^p's are continuously distributed, we adopt the framework that $X_{i,g}^p = \alpha_g + \epsilon_{i,g}^p$, for $g = 1, \cdots, G$ and $i = 1, \cdots, n$. Suppose that α_g's are i.i.d. normal with mean μ and variable σ_1^2. $\epsilon_{i,g}^p$'s are i.i.d. normal with zero mean and variance σ_2^2. They are also independent of α_g's. Then within a group g, $(X_{1,g}^p, \cdots, X_{n,g}^p)'$ is jointly normal with mean $(\mu, \cdots, \mu)'$ and variance-covariance matrix,

$$\eta^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix},$$

where $\eta^2 = \sigma_1^2 + \sigma_2^2$ and $\rho = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$. For any $i \neq j$, given $X_{j,g}^p = x$, $X_{i,g}^p$ is normally distributed with mean $\rho x + (1 - \rho)\mu$ and variance $(1 - \rho^2)\eta^2$. We can see this is just the example we discussed in Section 3.3. Hence, we can apply the Gauss-Legendre quadrature approximation in calculating an equilibrium. For the Monte Carlo study, we choose $\mu = 1$, $\eta = 2$, and $\rho = 0.4$. While values of those parameters are known to agents, econometricians need to estimate them from observed data. In this situation, we adopt a two-step algorithm in estimation. We first estimate μ, η, and ρ from $X_{i,g}^p$'s. Those estimates are consistent, when the number of independent groups, G, increases to ∞. We then plug those estimates into the likelihood of y and use the nested fixed point maximum likelihood method to estimate model parameters.[7]

The experiment results are tabulated in Tables 1 to 6. We can see that the nested fixed point maximum likelihood algorithm works well in general. Estimates have small biases and their (empirical) standard deviations decrease as sample size increases. Comparing the estimates when the sample is composed of many independent groups with those when there is just a single group of socially related agents, we find that the former ones are better than the later ones in terms of smaller biases and standard deviations. That is intuitive. Although agents make decisions independently, their expectations are correlated with each other. Hence, the

[7]When the first stage estimator is consistent, estimates of other parameters will be consistent. However, the standard errors of the first stage estimates would have effects on the standard errors of the second stage estimates. Then we need to make adjustments on standard errors of the second stage estimates.

more intensive the social relationships, the higher the collinearity of the regressors and the less variant the generated regressors. When there are many independent groups in the sample, as the number of group increases, the independence among the sample regressors increases, which implies better performance of estimators. Additionally, when we look at the Monte Carlo experiment results for a single group, we see that when the number of social links is fixed, the standard deviations of the estimates for the intercept, β_0, and the social relation intensity, λ, do not decrease when the sample size and the number of social links increase proportionally. But that situation alleviates if the number of social links is randomly determined. That is because the additional randomness helps reduces collinearity and increases variations of the generated regressors. In Table 7, we present the results when there is unobserved group random effects. Due to the presence of group unobservables, we need to use simulated likelihood for estimation(with simulation size $S = 500$). As a result, estimation biases are bigger than those without any unobserved group random effect. The performance of the nested fixed point ML estimator improves when the number of independent groups is raised.

For all the experiments, we tried two different information structures. In the first case, all X_i^p's are assumed to be public information to group members. In the second case, however, X_i^p is only revealed to i herself. We generate data under one information structure and estimate the model under both the correct one and the corresponding misspecified one. We can see that the estimation under the true information structure implies in general a higher maximized sample log likelihood than that under the misspecified one. Therefore, maximized sample likelihood is useful for model selection.

8 Empirical Application

We apply our model to property tax for municipalities in North Carolina. Municipal tax revenue depends on both tax rates and the amount of properties upon which tax is levied. A higher rate of property tax can increase tax revenue per unit properties owned by its residents. At the same time, it may provide incentives for residents to move out to nearby municipalities which offer lower tax rates. Therefore, there is a trade-off for setting a high tax rate due to the competition between nearby municipalities. We model this tax competition by a game with incomplete information.

This tax competition can be modeled as a simultaneous-move game with incomplete information in Section 2.2. Consider n municipalities in a state. They are related according to geographic vicinity, represented by an $n \times n$ matrix, W_n. For any two different cities, $i \neq j$, $W_{n,ij} = W_{n,ji} = 1$ if the distance between them is less than some cutoff value, $e > 0$; and $W_{n,ij} = W_{n,ji} = 0$ otherwise. as usual, $W_{n,ii} = 0$, for all i.[8] We use a_i to denote the rate of property tax set by city i. A tax rate must be non-negative, i.e., $a_i \geq 0$ for all $i = 1, \cdots, n$. i's payoffs by choosing a_i when other municipalities choose a_{-i} is given by (2.2). When a city chooses tax rates without knowing others' decisions, the expected payoff by choosing a_i is given by (2.3). In this equation, if $\lambda > 0$, the higher the tax rates of near-by municipalities, the higher the ideal tax rate that i wants to set, showing competition between cities. Without the non-negative constraint, the ideal tax rate is determined by some city-level demographics known to all local governments, such as the population, some city features which may not be known by other cities, say how rich residents in the city are in the current period, and the tax rates set by contiguous municipalities. For policy-making on property tax, the tax rates are non-negative.[9] Therefore, if the ideal rate is negative, we will get a corner solution. When the observed tax rates, $(a_1, \cdots, a_n)$, is an outcome of such an equilibrium, we have that

$$a_i = \max \left\{ \beta_0 + X^{c'}\beta_1 + X_i^{p'}\beta_2 + \lambda \sum_{j \neq i} W_{n,ij} E[a_j | X_{J_i}^p, Z] - \epsilon_i, 0 \right\}. \tag{8.1}$$

Therefore, we may use our model to investigate the problem of tax competition. Assume that condition (3.6) is satisfied. Hence, the observed data comes from the unique equilibrium in this model, which can be solved as a fixed point.

We look at municipalities in North Carolina, collecting data on property tax rates, government finance, and demographics in the 2012 fiscal year as well as geographic statistics (latitudes and longitudes). Since the total property tax a household pays is the sum of city tax and county tax, we also collect data on county property tax rates in 2012. Data of county and city property tax rates are from North Carolina Department of Revenue. Information about municipal government finance comes from North Carolina Department of State Treasurer. Data about city median household income is found from "Find theData.org", which is based on the American

[8] We can see that in this special case, W_n is a symmetric matrix with zero diagonal elements. In estimation, we row-normalize W_n such that the sum of each row is equal to 1.

[9] Local governments have other ways to subsidize residents. However, for property tax, the rates are non-negative.

Community Survey. Latitudes and longitudes are found from "CityLatitudeLongitude.com"[10] We calculate distance between any two cities based on latitudes and longitudes, using the Haversine formula[11]. Sample statistics are summarized in Table 8. From the table, we can find a big variety among the 506 municipalities in the sample in terms of demographics and financial status. Among those municipalities, the rates of property tax is strictly positive except for 29 of them who levy no property taxes.

In defining geographic vicinity, we tried two different cutoff values for distance between two municipalities, 30 kilometers and 50 kilometers, and estimate model parameters under the two associated social weighting matrices respectively. As for the public information about a municipality, we include the population and the property tax rates of related counties.[12] Since it is possible that a municipal government knows more about the financial situation of people living in its own territory than other governments do, it is reasonable to include in X^p some of the residents' financial data in the current period. Here we choose the median household income. Specifically, we assume that city median household income depends on two factors, state average level income and city idiosyncratic shocks.

$$MHI_i = \alpha + \epsilon_i^p, \tag{8.2}$$

where MHI_i is the median household income of municipality i in the sate (North Carolina). The random variable, α, represents state average median household income and ϵ_i^p, is a municipal-specific shock. Suppose that α is normal with mean μ and standard deviation, ω; ϵ_i^p's are i.i.d. with zero mean and standard deviation, γ; and ϵ_i^p is independent of α. Then X_i^p's have an exchangeable joint normal distribution, with mean μ, variance $\eta^2 = \omega^2 + \gamma^2$, and correlation coefficient, $\rho = \frac{\omega^2}{\omega^2+\gamma^2}$. [13] We estimate the model under two different information structures. The median household income is publicly known to related cities in the first scenario and is self-known in the second case.

In Table 9, we report regression results of the model under different vicinity cutoffs and

[10]The latitudes and longitudes of most municipalities in our sample are listed on the webpage, "CityLatitude-Longitude.com". Data about the rest 14 cities are found by searching on Google individually.

[11]

$$d = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\iota_2 - \iota_1}{2}\right) + \cos(\iota_1)\cos(\iota_2)\sin^2\left(\frac{\xi_2 - \xi_1}{2}\right)}\right),$$

where d is the distance, r is the radius, ι_1 and ι_2 are latitudes, and ξ_1 and ξ_2 are longitudes.

[12]When a municipality shares its border with several counties, we use the population-weighted average property tax rate.

[13]We collect the time series of the median household income for the state of North Carolina, α_t, from 1984 to 2012, and estimate μ and ω by sample mean and standard deviation.

information structures. We see that all estimates are significant at the 5% level with signs consistent across different regressions. The tax rate of a municipality is positively related to the tax rate set by the county to which it is affiliated and is negatively related to the median household income of the residents. The wealthier the residents in a city, the lower the rate of property tax. More importantly, the intensity of interactions between near-by cities, λ, is siginificantly positive, which supports the competition effects among municipalities. Comparing the maximized sample likelihood, we can see that models with spatial interactions outperform the traditional Tobit model. Moreover, among the four regressions with spatial interactions, we can see the magnitudes of parameters are similar to each other. The estimated competition effects between neighboring municipalities when the median household income is self-known are a little bit stronger than those under the hypothesis of public information.

9 Conclusion

We consider social interactions for censored outcomes under incomplete information, which incorporates various types of nonlinearity. First, outcomes of a social group depends nonlinearly on exogenous variables and model parameters due to expectations of peer's performances among socially linked agents. Second, the outcome of an agent changes with her own features in a nonlinear way due to censoring. Applying the theoretical results about socially interacted behaviors under incomplete information in Yang and Lee (2014), we relate the model with a simultaneous-move game under incomplete information with binding nonnegative constraint. We transform solution of an equilibrium conditional expectation function into calculating a fixed point of a function mapping and derive sufficient conditions for that mapping to be a contraction, which ensures the existence of a unique equilibrium. Under this scenario, we can solve the unique equilibrium, and identify and estimate model parameters.

In this paper, we solve and estimate the model under the hypothesis that parameters are within the range that ensures a unique equilibrium. This corresponds to weak or moderate social interaction scenario, under which it is possible to derive clear implications from model estimation. The situation of multiple equilibria corresponds to strong social interaction. In the literature, there are also methods which allows for multiple equilibria. Cases in point are the nonparametric two-step method by Bisin et al (2011) and Leung (2013). Their basic idea is to assume that agents of identical characteristics play the same strategy ex ante and exactly the same equilibrium is played for any repetition of the game. In that way, it is possible to identify choice probabilities consistently by non-parametric method. Plugging those choice probabilities back into the likelihood function, we can derive consistent estimates.[14]. For social interactions with large group sizes, a similar approach is also possible such as in Shang and Lee(2011). However, with small or moderate group sizes or in a general network setting without repetitions, computationally tractable estimation approaches remain to be found. All those are of interest for future research.

In the Tobit model, there are social interactions for only one type of behaviors. The value of observed outcomes and the censoring result are determined by just a single equation. A natural extension will be the sample selection model with social interactions incorporated. With social interactions in both the outcome equation and selection equation, as well as connections between the two equations, it is possible to analyze social interactions for two classes of related choices, one continuous and another discrete. That is a concise case with multiple equilibria for future research.

Appendices

A Equilibrium Analysis

Proof of Proposition 3.1. Suppose that if $(s_1(X^p_{J_1}, \epsilon_1), \cdots, s_n(X^p_{J_n}, \epsilon_n))$ satisfies (2.4), define $\psi : \prod_{i=1}^n \mathfrak{A}_i \to \mathfrak{C}^n$ as follows. For any $A = (A_1 \cdots, A_n)' \in \prod_{i=1}^n \mathfrak{A}_i$, there exists $k_1, \cdots, k_n \in \{1, \cdots, n\}$ such that $W_{n,k_i i} \neq 0$ and $A_i = X^p_{J_{k_i}}$. Set $\psi(A) = (\psi_1(A_1), \cdots, \psi_n(A_n))'$, where $\psi_i(A_i) = \psi_i(X^p_{J_{k_i}}) = E[s_i(X^p_{J_i}, \epsilon_1)|X^p_{J_{k_i}}, z]$, for $i = 1, \cdots, n$. Then, we have that

$$
\begin{aligned}
\psi_i(A_i) =& E[\max\left\{\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda\sum_{j\neq i} W_{n,ij}E[s_j(X^p_{J_i}, \epsilon_i)|X^p_{J_i}, Z] - \epsilon_i, 0\right\}|A_i, z] \\
=& E[H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda\sum_{j\neq i} W_{n,ij}E[s_j(X^p_{J_i}, \epsilon_i)|X^p_{J_i}, Z])|A_i, z] \\
=& E[H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda\sum_{j\neq i} W_{n,ij}\psi_j(X^p_{J_i}))|A_i, z].
\end{aligned}
$$

[14]In Bisin et al (2011), an alternative estimation method is to solve all equilibrium for each parameter value and choose the one to maximize log likelihood.

In the above equation, the first equality holds because ϵ_i's are independent of each other and also independent of X and W_n. The second equality follows from the definition of ψ. Thus, (3.3) holds. Conversely, suppose that a function $\psi : \prod_i^n \mathfrak{A}_i \to \mathfrak{C}^n$ satisfies (3.3). Define a profile of strategies, $(s_1(X^p_{J_1}, \epsilon_1), \cdots, s_n(X^p_{J_n}, \epsilon_n))$, by

$$s_i(X^p_{J_i}, \epsilon_i) = \max \left\{ \beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j(X^p_{J_i}) - \epsilon_i, 0 \right\},$$

for $i, j = 1, \cdots, n$ and $W_{j,i} \neq 0$. Taking expectations over $s_j(X^p_{J_i}, \epsilon_j)$ conditional on $X^p_{J_i}$ and $Z = z$, we derive that

$$\begin{aligned} E[s_j(X^p_{J_i}, \epsilon_j)|X^p_{J_i}, z] =& E\Big[\max \left\{ \beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j(X^p_{J_i}) - \epsilon_i, 0 \right\} \Big| X^p_{J_i}, z\Big] \\ =& E[H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j(X^p_{J_i}))|X^p_{J_i}, z] \\ =& \psi_j(X^p_i), \end{aligned}$$

where the second equality follows from (3.3). Hence, we have that

$$s_i(X^p_{J_i}, \epsilon_i) = \max \left\{ \beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}E[s_j(X^p_{J_i}, \epsilon_j)|X^p_{J_i}, z] - \epsilon_i, 0 \right\}.$$

Therefore, (2.4) is satisfied. ■

Proof of Proposition 3.2. For any two functions, $\xi, \xi' \in \Xi$,

$$\begin{aligned} \left\|T(\xi) - T(\xi')\right\| =& \max_{1 \le i \le n} \max_{\{k: W_{n,ki} \neq 0\}} \int |E[\Big(H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\xi_j(X^p_{J_i})) \\ &- H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\xi'_j(X^p_{J_i}))\Big)|x^p_{J_k}, z]|dF_p(x^p) \\ \le& \max_{1 \le i \le n} \max_{\{k: W_{n,ki} \neq 0\}} |\lambda| \sup_x |\frac{dH(x)}{dx}| \sum_{j \neq i} W_{n,ij} \int E[|\xi_j(X^p_{J_i}) - \xi'_j(X^p_{J_i})||x^p_{J_k}, z]dF_p(x^p) \\ =& \max_{1 \le i \le n} \max_{\{k: W_{n,ki} \neq 0\}} |\lambda| \sup_x |\frac{dH(x)}{dx}| \sum_{j \neq i} W_{n,ij} \int |\xi_j(X^p_{J_i}) - \xi'_j(X^p_{J_i})|dF_p(X^p) \\ \le& |\lambda| \, \|W_n\|_\infty \sup_x |\frac{dH(x)}{dx}|| \left\|\psi - \psi'\right\|. \end{aligned}$$

Recall that $F_p(\cdot)$ is a simplified notation of the distribution of X_i^p's conditional on public information $Z = z$. By calculation, $dH(x)/dx = F_\epsilon(x)$. Therefore, if $|\lambda| \, \|W_n\|_\infty < 1$, $T : \Xi \to \Xi$ is a contraction mapping on the complete metric space, $(\Xi, \| \cdot \|)$. Then, it admits a unique fixed point. Because there is a one-to-one correspondence between a BNE and a fixed point of T, there is a unique BNE. ■

B Identification Proofs

Proof of Lemma 4.1. Define $c_i = \beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda\sum_{j\neq i} W_{n,ij}E[y_j|X_{J_i}^p, z]$. According to (2.1), we have that $E[I(y_i > 0)|X^p, z] = Pr(c_i - \epsilon_i > 0|X^p, z) = F_\epsilon(c_i;\sigma)$, where the second equality follows from the independence between the idiosyncratic shocks and the exogenous characteristics. Since $F_\epsilon(c;\sigma)$ is strictly increasing with respect to c, we have the inversion, $c_i = F_\epsilon^{-1}(E[I(y_i > 0)|X^p, z];\sigma)$. Because

$$E[y_i|X^p, z] = \int_{c<c_i}(c_i - c)f_\epsilon(c;\sigma)dc = c_i F_\epsilon(c_i;\sigma) - \int_{c<c_i} cf_\epsilon(c;\sigma)dc,$$

we derive (4.3). ■

Proof of Proposition 4.1. For X^p and z, we define a function S such that

$$S(E[y|X^p, z], E[I(y > 0)|X^p, z], \sigma) = E[y|X^p, z] - E[I(y > 0)|X^p, z]F_\epsilon^{-1}(E[I(y > 0)|X^p, z];\sigma) + \int_{c<F_\epsilon^{-1}(E[I(y>0)|X^p,z];\sigma)} cf_\epsilon(c;\sigma)dc.$$

By (4.3′), when data are generated from the true parameter σ_0, $S(E[y|X^p, z], E[I(y > 0)|X^p, z], \sigma_0) = 0$. We would like to show that given the observed data $(E[y|X^p, z], E[I(y > 0)|X^p, z])$, there is a unique σ that satisfies the above equation. That is, σ can be identified from the observed data. To achieve this, we first take the partial derivative,

$$\begin{aligned}&\frac{\partial S(E[y|X^p, z], E[I(y > 0)|X^p, z], \sigma)}{\partial\sigma}\\ =&\big(F_\epsilon^{-1}(E[I(y > 0)|X^p, z];\sigma)f_\epsilon(F_\epsilon^{-1}(E[I(y > 0)|X^p, z];\sigma);\sigma) - E[I(y > 0)|X^p, z]\big)\\ &\cdot\frac{\partial F_\epsilon^{-1}(E[I(y > 0)|X^p, z];\sigma)}{\partial\sigma}\\ &+\int_{c<F_\epsilon^{-1}(E[I(y>0)|X^p,z];\sigma)} c\frac{\partial f_\epsilon(c;\sigma)}{\partial\sigma}dc.\end{aligned}$$

Since $F_\epsilon(F_\epsilon^{-1}(E[I(y > 0)|X^p, z];\sigma);\sigma) = E[I(y^* > 0)|X^p, z]$,
$f_\epsilon(F_\epsilon^{-1}(E[I(y > 0)|X^p, z];\sigma);\sigma)\frac{\partial F_\epsilon^{-1}(E[I(y>0)|X^p,z];\sigma)}{\partial\sigma} + \frac{\partial F_\epsilon(F_\epsilon^{-1}(E[I(y>0)|X^p,z];\sigma);\sigma)}{\partial\sigma} = 0$,
where the derivative with respect to σ in the second term is the derivative of $F_\epsilon(c;\sigma)$ with respect to σ. Thus,

$$\frac{\partial F_\epsilon^{-1}(E[I(y > 0)|X^p, z];\sigma)}{\partial\sigma} = -\frac{\frac{\partial F_\epsilon(F_\epsilon^{-1}(E[I(y>0)|X^p,z];\sigma);\sigma)}{\partial\sigma}}{f_\epsilon(F_\epsilon^{-1}(E[I(y > 0)|X^p, z];\sigma);\sigma)}.$$

Additionally, by using integration by parts,

$$\int_{c<F_\epsilon^{-1}(E[I(y>0)|X^p,z];\sigma)} c\frac{\partial f_\epsilon(c;\sigma)}{\partial\sigma}dc$$

$$=F_\epsilon^{-1}(E[I(y0)|X^p,z];\sigma)\frac{\partial F_\epsilon(F_\epsilon^{-1}(E[I(y>0)|X^p,z];\sigma);\sigma)}{\partial\sigma}-\lim_{c\to-\infty}c\frac{\partial F_\epsilon(c;\sigma)}{\partial\sigma}$$

$$-\int_{c<F_\epsilon^{-1}(E[I(y>0)|X^p,z];\sigma)}\frac{\partial F_\epsilon(c;\sigma)}{\partial\sigma}dc.$$

Hence,

$$\frac{\partial S(E[y|X^p,z],E[I(y>0)|X^p,z],\sigma)}{\partial\sigma}$$

$$=E[I(y>0)|X^p,z]\frac{\frac{\partial F_\epsilon(F_\epsilon^{-1}(E[I(y>0)|X^p,z];\sigma);\sigma)}{\partial\sigma}}{f_\epsilon(F_\epsilon^{-1}(E[I(y>0)|X^p,z];\sigma);\sigma)}$$

$$-\int_{c<F_\epsilon^{-1}(E[I(y>0)|X^p,z];\sigma)}\frac{\partial F_\epsilon(c;\sigma)}{\partial\sigma}dc$$

$$=\int_{c<F_\epsilon^{-1}(E[I(y>0)|X^p,z];\sigma)}(\frac{\frac{\partial F_\epsilon(F_\epsilon^{-1}(E[I(y>0)|X^p,z];\sigma);\sigma)}{\partial\sigma}}{f_\epsilon(F_\epsilon^{-1}(E[I(y>0)|X^p,z];\sigma);\sigma)}-\frac{\frac{\partial F_\epsilon(c;\sigma)}{\partial\sigma}}{f_\epsilon(c;\sigma)})f_\epsilon(c;\sigma)dc.$$

Therefore, when Assumption 4.4 is satisfied, $\frac{\partial S(E[y|X^p,z],E[I(y>0)|X^p,z],\sigma)}{\partial\sigma}$ is either positive or negative. As a result, by the implicit function theorem, for $(E[y|X^p,z],E[I(y^*>0)|X^p,z])$, there is a unique σ such that (4.3′) holds. That is, from the moments, $(E[y|X^p,z],E[I(y^*>0)|X^p,z])$, we can identify σ. ■

Proof of Proposition 4.2. σ is identified from Proposition 4.1. As $H(x;\sigma)=xF_\epsilon(x,\sigma)-\int_{c<x}cf_\epsilon(c;\sigma)dc$, $\frac{\partial H(x;\sigma)}{\partial x}=F_\epsilon(x;\sigma)>0$ for any $x\in\Re^1$. Thus, $H(x;\sigma)$ is strictly increasing in x and $H^{-1}(\cdot;\sigma)$ is well-defined. Therefore, we have

$$H^{-1}(E[y_i|X^p,z])=\beta_0+X_i^{c'}\beta_1+X_i^{p'}\beta_2+\lambda\sum_{j\neq i}W_{n,ij}E[y_j|X_{J_i}^p,z].$$

Therefore, if $(\beta_0,\beta_1',\beta_2',\lambda)'$ and $(\widetilde{\beta}_0,\widetilde{\beta}_1',\widetilde{\beta}_2',\widetilde{\lambda}')'$ are observationally equivalent at $\overline{W}$, $\overline{J}$ and $\overline{X}$,

$$\begin{pmatrix} l_n, & \overline{X}^c, & \overline{X}^p, & E^p\end{pmatrix}(\beta_0-\widetilde{\beta}_0,\beta_1'-\widetilde{\beta}_1',\beta_2'-\widetilde{\beta}_2',\lambda-\widetilde{\lambda})'=0.$$

As $\begin{pmatrix} l_n, & \overline{X}^c, & \overline{X}^p, & E^p\end{pmatrix}$ has full column rank, $(\beta_0,\beta_1',\beta_2',\lambda)'=(\widetilde{\beta}_0,\widetilde{\beta}_1',\widetilde{\beta}_2',\widetilde{\lambda}')'$. ■

References

Abrevaya, J. and Shen, S. 2014. Estimation of censored panel-data models with slope heterogeneity. *Journal of Applied Econometrics*, 29:523–548.

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.

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. 2010. Estimating static models of strategic interactions. *Journal of Business and Economic Statistics*, 28(4):469–482.

Ballester, C., Calvó-Armengol, H., and Zenou, Y. 2006. Who is who in networks. wanted: the key player. *Econometrica*, 74(5):1403–1417.

Bisin, A., Mora, A., and Topa, G. 2011. The empirical content of models with multiple equilibria in economies with social interactions. *working paper.*

Boucher, V., Bramoullé, Y., Djebbari, H., and Fortin, B. 2014. Do peer affect student achievement? Evidence from Canada using group size variation. *Journal of Applied Econometrics*, 29:91–109.

Brock, W. and Durlauf, S. 2000. Interactions-based models. 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.

Brueckner, J. 2003. Strategic interaction among governments: an overview of empirical studies. *International Regional Science Review*, 26:175–188.

de Finetti, B. 1975. *Theory of probability*, volume 2. John Wiley& Sons, New York.

Durlauf, S. 2000. A framework for the study of individual behavior and social interactions. *working paper 16, Wisconsin Madison, Social Systems.*

Gaspar, J. and Judd, K. 1997. Solving large-scale rational expectations models. *Macroeconomic Dynamics*, 68(2):235–260.

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.

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

Kumar, A. 2012. Nonparametric estimation of the impact of taxes on female labor supply. *Journal of Applied Econometrics*, 27:415–439.

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. 2013. Two-step estimation of network formation models with incomplete information. *working paper*.

Li, S., Liu, Y., and Deininger, K. 2013. How important are endogenous peer effects in group lending? estimating a static game of incomplete information. *Journal of Applied Econometrics*, 28:864–882.

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 panel 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, New York.

Moffitt, R. A. 2000. Policy interventions, low-level equilibria, and social interactions. *Social Dynamics*, 97:45–82.

Osborne, M. J. and Rubinstein, A. 1994. *A Course in Game Theory*. MIT Press, Cambridge, Massachusetts; London, England.

Porto, E. D. and Revelli, F. 2013. Tax-limited reaction functions. *Journal of Applied Econometrics*, 28:823–839.

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.

Rust, J. 1987. Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher. *Econometrica*, 55(5):999–1033.

Shang, Q. and Lee, L. 2011. Two-step estimation of endogenous and exogenous group effects. *Econometric Reviews*, 30(2):173–207.

Tao, J. and Lee, L. 2014. A social interaction model with an extreme order statistics. *The Econometrics Journal*, 17(3):197–240.

Yang, C. and Lee, L. 2014. Social interactions under incomplete information with heterogeneous expectations. *working paper*.

Table 1: Tobit Model with Discrete Characteristics for Independent Groups

		Characteristics are all Publicly Known			
		Publicly Known		Self-Known	
Model Specification		$n = 20$, $G = 100$	$n = 20$, $G = 500$	$n = 20$, $G = 100$	$n = 20$, $G = 500$
β_0	0	0.0017 (0.0599)	0.0004 (0.0261)	0.0014 (0.0647)	-0.0000 (0.0285)
β_1	1	1.0020 (0.0274)	0.9999 (0.0122)	1.0021 (0.0276)	0.9999 (0.0122)
β_2	1	0.9987 (0.0483)	1.0003 (0.0217)	0.9988 (0.0487)	1.0005 (0.0219)
λ	0.3	0.2983 (0.0486)	0.2997 (0.0214)	0.2985 (0.0536)	0.3000 (0.0238)
σ	1	0.9997 (0.0194)	0.9998 (0.0090)	1.0016 (0.0195)	1.0017 (0.0090)
$m \log L$		-1.1611 (0.0163)	-1.1618 (0.0072)	-1.1627 (0.0163)	-1.1633 (0.0072)
r_{true}		0.8920	0.9990	-	-
r_{censor}		0.3209	0.3206	0.3209	0.3206

		Characteristics are Self-Known			
		Self-Known		Publicly Known	
Model Specification		$n = 20$, $G = 100$	$n = 20$, $G = 500$	$n = 20$, $G = 100$	$n = 20$, $G = 500$
β_0	0	0.0016 (0.0643)	0.0003 (0.0283)	0.0588 (0.0605)	0.0578 (0.0265)
β_1	1	1.0020 (0.0278)	0.9999 (0.0122)	1.0051 (0.0275)	1.0032 (0.0122)
β_2	1	0.9988 (0.0484)	1.0005 (0.0219)	0.9864 (0.0483)	0.9880 (0.0219)
λ	0.3	0.2983 (0.0535)	0.2997 (0.0235)	0.2469 (0.0510)	0.2481 (0.0225)
σ	1	0.9998 (0.0194)	0.9999 (0.0090)	1.0014 (0.0194)	1.0014 (0.0090)
$m \log L$		-1.1614 (0.0163)	-1.1620 (0.0072)	-1.1627 (0.0163)	-1.1633 (0.0072)
r_{true}		0.8750	0.9970	-	-
r_{censor}		0.3208	0.3205	0.3208	0.3205

Note: G is the number of groups. n is the population for each group. $m \log L$ is the estimated sample average log likelihood. r_{true} is the proportion of simulations for which estimated log likelihood is bigger than that under wrong information structure. r_{censor} is the censoring rate. The numbers in parentheses are standard deviation.

Table 2: Tobit Model with Continuous Characteristics for Independent Groups

		Characteristics are All Publicly Known			
		Publicly Known		Self-Known	
Model Specification		$n=20$, $G=100$	$n=20$, $G=500$	$n=20$, $G=100$	$n=20$, $G=500$
β_0	0	-0.0030 (0.0385)	0.0002 (0.0172)	-0.0636 (0.1122)	-0.0592 (0.0494)
β_1	1	1.0008 (0.0248)	0.9998 (0.0114)	1.0067 (0.0272)	1.0061 (0.0124)
β_2	1	0.9995 (0.0166)	0.9997 (0.0077)	1.0365 (0.0270)	1.0369 (0.0125)
λ	0.3	0.3012 (0.0167)	0.3001 (0.0074)	0.3022 (0.0612)	0.3002 (0.0271)
σ	1	0.9988 (0.0191)	0.9998 (0.0085)	1.0759 (0.0222)	1.0769 (0.0097)
$m \log L$		-1.1361 (0.0306)	-1.1368 (0.0133)	-1.1937 (0.0330)	-1.1942 (0.0142)
r_{true}		1.0000	1.0000	-	-
r_{censor}		0.2746	0.2749	0.2746	0.2749

		Characteristics are Self-Known			
		Self-Known		Publicly Known	
Model Specification		$n=20$, $G=100$	$n=20$, $G=500$	$n=20$, $G=100$	$n=20$, $G=500$
β_0	0	-0.0037 (0.0901)	0.0003 (0.0390)	0.4473 (0.0474)	0.4528 (0.0218)
β_1	1	1.0010 (0.0255)	0.9999 (0.0115)	1.0068 (0.0256)	1.0057 (0.0117)
β_2	1	1.0014 (0.0210)	0.9998 (0.0094)	1.0777 (0.0167)	1.0774 (0.0077)
λ	0.3	0.3003 (0.0487)	0.3000 (0.0209)	0.0431 (0.0209)	0.0414 (0.0094)
σ	1	0.9989 (0.0191)	0.9998 (0.0084)	1.0077 (0.0193)	1.0087 (0.0085)
$m \log L$		-1.1452 (0.0280)	-1.1459 (0.0121)	-1.1525 (0.0283)	-1.1533 (0.0121)
r_{true}		0.9990	1.0000	-	-
r_{censor}		0.2682	0.2683	0.2682	0.2683

		Distribution of Self-Known Characteristics	
	True parameters	$n=20$, $G=100$	$n=20$, $G=500$
μ	1	1.0054 (0.0037)	0.9993 (0.0570)
η	2	1.9964 (0.0637)	2.0000 (0.0280)
ρ	0.4	0.3969 (0.0384)	0.4000 (0.0164)

Note: G is the number of groups. n is the population for each group. $m \log L$ is the estimated sample average log likelihood. r_{true} is the proportion of simulations for which estimated log likelihood is bigger than that under wrong information structure. r_{censor} is the censoring rate. The numbers in parentheses are standard deviation.

Table 3: Tobit Model with Discrete Characteristics and Constant Friend Number for A Single Group

		Characteristics are Publicly Known					
		Publicly Known			Self-Known		
Model Specification		$n = 200$ $F = 30$	$n = 1000$ $F = 150$	$n = 1000$ $F = 30$	$n = 200$ $F = 30$	$n = 1000$ $F = 150$	$n = 1000$ $F = 30$
β_0	0	0.0561	0.0217	-0.0100	0.0651	0.0039	-0.0091
		(0.5469)	(0.5333)	(0.2483)	(0.5794)	(0.5602)	(0.2818)
β_1	1	0.9963	1.0013	1.0012	0.9964	1.0014	1.0012
		(0.0835)	(0.0380)	(0.0381)	(0.0832)	(0.0379)	(0.0381)
β_2	1	1.0023	0.9997	0.9994	1.0009	0.9995	0.9994
		(0.1600)	(0.0700)	(0.0699)	(0.1600)	(0.0701)	(0.0699)
λ	0.3	0.2416	0.2777	0.3094	0.2330	0.2758	0.3085
		(0.5368)	(0.5242)	(0.2410)	(0.5713)	(0.5516)	(0.2740)
σ	1	0.9881	0.9980	0.9981	0.9884	0.9981	0.9983
		(0.0630)	(0.0270)	(0.0271)	(0.0632)	(0.0270)	(0.0270)
$m \log L$		-1.1512	-1.1607	-1.1606	-1.1515	-1.1607	-1.1608
		(0.0532)	(0.0228)	(0.0228)	(0.0533)	(0.0228)	(0.0228)
r_{true}		0.5400	0.5730	0.5930	-	-	-
r_{censor}		0.3202	0.3203	0.3204	0.3202	0.3203	0.3204

		Characteristics are Self-Known					
		Self-Known			Publicly Known		
Model Specification		$n = 200$ $F = 30$	$n = 1000$ $F = 150$	$n = 1000$ $F = 30$	$n = 200$ $F = 30$	$n = 1000$ $F = 150$	$n = 1000$ $F = 30$
β_0	0	0.0648	0.0254	-0.0094	0.1136	0.0794	0.0560
		(0.5785)	(0.5605)	(0.2825)	(0.5529)	(0.5410)	(0.2490)
β_1	1	0.9967	1.0012	1.0012	0.9964	1.0012	1.0013
		(0.0836)	(0.0379)	(0.0380)	(0.0840)	(0.0380)	(0.0381)
β_2	1	1.0012	0.9993	0.9994	1.0015	0.9993	0.9982
		(0.1603)	(0.0700)	(0.0701)	(0.1604)	(0.0699)	(0.0701)
λ	0.3	0.2331	0.2746	0.3088	0.1850	0.2211	0.2448
		(0.5706)	(0.5518)	(0.2745)	(0.5441)	(0.5323)	(0.2423)
σ	1	0.9884	0.9980	0.9980	0.9884	0.9980	0.9982
		(0.0633)	(0.0270)	(0.0270)	(0.0632)	(0.0270)	(0.0270)
$m \log L$		-1.1513	-1.1607	-1.1606	-1.1513	-1.1607	-1.1608
		(0.0533)	(0.0229)	(0.0228)	(0.0532)	(0.0229)	(0.0228)
r_{true}		0.4890	0.4690	0.5780	-	-	-
r_{censor}		0.3202	0.3202	0.3203	0.3202	0.3202	0.3203

Note: n is the number of agents in the group. F is the constant number of friends a person can make. $m \log L$ is the estimated sample average log likelihood. r_{true} is the proportion of simulations for which estimated log likelihood is bigger than that under wrong information structure. r_{censor} is the censoring rate. The numbers in parentheses are standard deviation.

Table 4: Tobit Model with Discrete Characteristics and Random Friend Number for A Single Group

		Characteristics are Publicly Known					
		Publicly Known			Self-Known		
Model Specification		$n = 200$ $UF = 59$	$n = 1000$ $UF = 299$	$n = 1000$ $UF = 59$	$n = 200$ $UF = 59$	$n = 1000$ $UF = 299$	$n = 1000$ $UF = 59$
β_0	0	0.0066	0.0070	0.0015	0.0077	0.0045	0.0051
		(0.1596)	(0.2746)	(0.1402)	(0.1729)	(0.3041)	(0.1532)
β_1	1	1.0042	1.0014	1.0016	1.0045	1.0014	1.0016
		(0.0714)	(0.0385)	(0.0384)	(0.0714)	(0.0385)	(0.0384)
β_2	1	1.0002	0.9996	1.0001	1.0002	0.9997	0.9997
		(0.1488)	(0.0700)	(0.0702)	(0.1481)	(0.0700)	(0.0701)
λ	0.3	0.2910	0.2927	0.2978	0.2888	0.2949	0.2943
		(0.2459)	(0.2698)	(0.1354)	(0.2692)	(0.2983)	(0.1487)
σ	1	0.9894	0.9993	0.9993	0.9903	0.9994	0.9998
		(0.0501)	(0.0286)	(0.0287)	(0.0503)	(0.0286)	(0.0287)
$m \log L$		-1.4070	-1.1616	-1.1599	-1.4079	-1.1617	-1.1604
		(0.0508)	(0.0236)	(0.0236)	(0.0509)	(0.0236)	(0.0236)
r_{true}		0.5760	0.5440	0.6740	-	-	-
r_{censor}		0.3214	0.3206	0.3219	0.3214	0.3206	0.3219
mF		29.4344	149.6204	29.5245	29.4344	149.6204	29.5245

		Characteristics are Self-Known					
		Self-Known			Publicly Known		
Model Specification		$n = 200$ $UF = 59$	$n = 1000$ $UF = 299$	$n = 1000$ $UF = 59$	$n = 200$ $UF = 59$	$n = 1000$ $UF = 299$	$n = 1000$ $UF = 59$
β_0	0	0.0162	0.0043	0.0056	0.0633	0.0590	0.0516
		(0.3633)	(0.3044)	(0.1532)	(0.3233)	(0.2742)	(0.1394)
β_1	1	1.0058	1.0014	1.0015	1.0060	1.0015	1.0019
		(0.0847)	(0.0384)	(0.0383)	(0.0849)	(0.0385)	(0.0384)
β_2	1	1.0010	0.9997	0.9998	0.9983	0.9989	0.9974
		(0.1575)	(0.0701)	(0.0704)	(0.1584)	(0.0702)	(0.0707)
λ	0.3	0.2830	0.2951	0.2939	0.2370	0.2416	0.2487
		(0.3491)	(0.2984)	(0.1485)	(0.3211)	(0.2699)	(0.1354)
σ	1	0.9894	0.9992	0.9992	0.9897	0.9993	0.9995
		(0.0631)	(0.0286)	(0.0286)	(0.0631)	(0.0287)	(0.0286)
$m \log L$		-1.1498	-1.1616	-1.1600	-1.1501	-1.1617	-1.1603
		(0.0534)	(0.0236)	(0.0236)	(0.0535)	(0.0236)	(0.0236)
r_{true}		0.5630	0.5900	0.6250	-	-	-
r_{censor}		0.3215	0.3205	0.3218	0.3215	0.3205	0.3218
mF		29.4344	149.6204	29.5245	29.4344	149.6204	29.5245

Note: n is the number of agents in the group. UF and mF are respectively the maximum and average number of friends. $m \log L$ is the estimated sample average log likelihood. r_{true} is the proportion of simulations for which estimated log likelihood is bigger than that under wrong information structure. r_{censor} is the censoring rate. The numbers in parentheses are standard deviation.

Table 5: Tobit Model with Continuous Characteristics and Constant Friend Number for A Single Group

Model Specification		Characteristics are Publicly Known: Publicly Known $n=200$ $F=30$	$n=1000$ $F=150$	$n=1000$ $F=30$	Self-Known $n=200$ $F=30$	$n=1000$ $F=150$	$n=1000$ $F=30$
β_0	0	0.0106	0.0319	-0.0101	0.1998	0.2659	0.1371
		(0.6553)	(0.6319)	(0.2668)	(0.9211)	(0.8726)	(0.5054)
β_1	1	0.9980	1.0011	1.0012	0.9980	1.0010	1.0009
		(0.0848)	(0.0385)	(0.0386)	(0.0849)	(0.0385)	(0.0387)
β_2	1	0.9987	1.0009	1.0009	0.9298	0.9425	0.9228
		(0.0632)	(0.0278)	(0.0279)	(0.1620)	(0.1414)	(0.0786)
λ	0.3	0.2863	0.2713	0.3026	0.2342	0.1996	0.2708
		(0.3790)	(0.3876)	(0.1673)	(0.5352)	(0.4943)	(0.2672)
σ	1	0.9859	0.9972	0.9973	0.9878	0.9976	0.9993
		(0.0657)	(0.0284)	(0.0285)	(0.0661)	(0.0285)	(0.0286)
$m \log L$		-1.1261	-1.1346	-1.1345	-1.1280	-1.1350	-1.1365
		(0.2576)	(0.2501)	(0.2499)	(0.2587)	(0.2502)	(0.2499)
r_{true}		0.5990	0.6220	0.8120	-	-	-
r_{censor}		0.2751	0.2755	0.2756	0.2751	0.2755	0.2756

Model Specification		Characteristics are Self-Known: Self-Known $n=200$ $F=30$	$n=1000$ $F=150$	$n=1000$ $F=30$	Publicly Known $n=200$ $F=30$	$n=1000$ $F=150$	$n=1000$ $F=30$
β_0	0	0.0643	0.0442	-0.0171	0.3756	0.3959	0.3649
		(0.9197)	(0.8261)	(0.4353)	(0.6291)	(0.6078)	(0.2574)
β_1	1	0.9996	1.0011	1.0010	0.9990	1.0011	1.0016
		(0.0821)	(0.0370)	(0.0370)	(0.0818)	(0.0370)	(0.0371)
β_2	1	1.0092	1.0085	0.9979	1.0919	1.0936	1.0936
		(0.1738)	(0.1506)	(0.0809)	(0.0615)	(0.0282)	(0.0281)
λ	0.3	0.2652	0.2733	0.3089	0.0811	0.0595	0.0824
		(0.5229)	(0.4756)	(0.2487)	(0.3527)	(0.3577)	(0.1464)
σ	1	0.9864	0.9975	0.9974	0.9871	0.9977	0.9984
		(0.0628)	(0.0275)	(0.0275)	(0.0627)	(0.0275)	(0.0276)
$m \log L$		-1.1335	-1.1430	-1.1429	-1.1338	-1.1431	-1.1436
		(0.2258)	(0.2163)	(0.2163)	(0.2257)	(0.2164)	(0.2165)
r_{true}		0.5100	0.5270	0.6810	-	-	-
r_{censor}		0.2696	0.2694	0.2694	0.2696	0.2694	0.2694

Note: n is the number of agents in the group. F is the constant number of friends a person can make. $m \log L$ is the estimated sample average log likelihood. r_{true} is the proportion of simulations for which estimated log likelihood is bigger than that under wrong information structure. r_{censor} is the censoring rate. The numbers in parentheses are standard deviation.

Table 6: Tobit Model with Continuous Characteristics and Random Friend Number for A Single Group

		Characteristics are Publicly Known					
		Publicly Known			Self-Known		
Model Specification		$n = 200$ $UF = 59$	$n = 1000$ $UF = 299$	$n = 1000$ $UF = 59$	$n = 200$ $UF = 59$	$n = 1000$ $UF = 299$	$n = 1000$ $UF = 59$
β_0	0	0.0009	-0.0056	-0.0063	0.0976	0.1504	0.1288
		(0.3353)	(0.2956)	(0.1515)	(0.4870)	(0.4572)	(0.2588)
β_1	1	0.9978	0.9981	0.9984	0.9976	0.9980	0.9981
		(0.0905)	(0.0405)	(0.0406)	(0.0916)	(0.0405)	(0.0407)
β_2	1	1.0007	0.9991	0.9992	0.9200	0.9216	0.9203
		(0.0661)	(0.0276)	(0.0276)	(0.1025)	(0.0770)	(0.0460)
λ	0.3	0.2919	0.2978	0.3068	0.2793	0.2703	0.2791
		(0.2771)	(0.2090)	(0.1120)	(0.2976)	(0.2574)	(0.1382)
σ	1	0.9836	0.9960	0.9962	0.9880	0.9974	1.0015
		(0.0655)	(0.0285)	(0.0285)	(0.0662)	(0.0287)	(0.0290)
$m \log L$		-1.1039	-1.1495	-1.1480	-1.1081	-1.1508	-1.1529
		(0.2721)	(0.2506)	(0.2503)	(0.2744)	(0.2513)	(0.2526)
r_{true}		0.7090	0.7710	0.9170	-	-	-
r_{censor}		0.2895	0.2620	0.2634	0.2895	0.2620	0.2634
mF		29.4867	149.5960	29.5193	29.4867	149.5960	29.5193

		Characteristics are Self-Known					
		Self-Known			Publicly Known		
Model Specification		$n = 200$ $UF = 59$	$n = 1000$ $UF = 299$	$n = 1000$ $UF = 59$	$n = 200$ $UF = 59$	$n = 1000$ $UF = 299$	$n = 1000$ $UF = 59$
β_0	0	-0.0258	-0.0055	-0.0102	0.2464	0.2614	0.2296
		(0.4592)	(0.4146)	(0.1962)	(0.3522)	(0.3208)	(0.1945)
β_1	1	0.9969	0.9980	0.9981	0.9973	0.9985	0.9993
		(0.0887)	(0.0400)	(0.0401)	(0.0884)	(0.0401)	(0.0404)
β_2	1	0.9954	0.9975	0.9969	1.0921	1.0911	1.0890
		(0.0977)	(0.0788)	(0.0420)	(0.0648)	(0.0277)	(0.0279)
λ	0.3	0.3160	0.3041	0.3070	0.1415	0.1255	0.1444
		(0.2631)	(0.2395)	(0.1120)	(0.2384)	(0.1900)	(0.1028)
σ	1	0.9835	0.9962	0.9963	0.9873	0.9973	0.9999
		(0.0621)	(0.0275)	(0.0277)	(0.0625)	(0.0275)	(0.0279)
$m \log L$		-1.1134	-1.1551	-1.1537	-1.1161	-1.1558	-1.1565
		(0.2394)	(0.2189)	(0.2193)	(0.2398)	(0.2191)	(0.2199)
r_{true}		0.6490	0.7180	0.8810	-	-	-
r_{censor}		0.2828	0.2582	0.2595	0.2828	0.2582	0.2595
mF		29.4867	149.5960	29.5193	29.4867	149.5960	29.5193

Note: n is the number of agents in the group. UF and mF are respectively the maximum and average number of friends. $m \log L$ is the estimated sample average log likelihood. r_{true} is the proportion of simulations for which estimated log likelihood is bigger than that under wrong information structure. r_{censor} is the censoring rate. The numbers in parentheses are standard deviation.

Table 7: Tobit Model with Discrete Characteristics and Unobserved Group Random Effects

		Characteristics are Publicly Known			
		Publicly Known		Self-Known	
Model Specification		$n = 20$, $G = 100$	$n = 20$, $G = 500$	$n = 20$, $G = 100$	$n = 20$, $G = 500$
β_0	0	0.0323 (0.1733)	0.0127 (0.0612)	0.0287 (0.1776)	0.0121 (0.0630)
β_1	1	1.0023 (0.0287)	1.0001 (0.0128)	1.0021 (0.0286)	0.9999 (0.0129)
β_2	1	0.9982 (0.0517)	1.0007 (0.0226)	0.9980 (0.0523)	1.0002 (0.0228)
λ	0.3	0.3051 (0.0613)	0.3023 (0.0259)	0.3057 (0.0702)	0.3025 (0.0298)
σ	1	1.0008 (0.0194)	1.0005 (0.0094)	1.0029 (0.0194)	1.0025 (0.0095)
γ	1	1.0514 (0.1364)	1.0194 (0.0513)	1.0477 (0.1371)	1.0190 (0.0516)
$m \log L$		-1.1667 (0.0344)	-1.1655 (0.0149)	-1.1682 (0.0344)	-1.1669 (0.0149)
r_{true}		0.8880	0.9960	-	-
r_{censor}		0.3509	0.3516	0.3509	0.3516

		Characteristics are Self-Known			
		Self-Known		Publicly Known	
Model Specification		$n = 20$, $G = 100$	$n = 20$, $G = 500$	$n = 20$, $G = 100$	$n = 20$, $G = 500$
β_0	0	0.0289 (0.1781)	0.0124 (0.0630)	0.1106 (0.1830)	0.0920 (0.0637)
β_1	1	1.0024 (0.0289)	1.0001 (0.0129)	1.0030 (0.0291)	1.0006 (0.0129)
β_2	1	0.9985 (0.0520)	1.0006 (0.0229)	0.9870 (0.0519)	0.9895 (0.0228)
λ	0.3	0.3051 (0.0707)	0.3026 (0.0297)	0.2458 (0.0629)	0.2426 (0.0265)
σ	1	1.0009 (0.0194)	1.0005 (0.0093)	1.0026 (0.0195)	1.0022 (0.0093)
γ	1	1.0481 (0.1362)	1.0192 (0.0520)	1.1095 (0.1454)	1.0756 (0.0553)
$m \log L$		-1.1669 (0.0345)	-1.1657 (0.0149)	-1.1680 (0.0346)	-1.1668 (0.0149)
r_{true}		0.8560	0.9950	-	-
r_{censor}		0.3509	0.3515	0.3509	0.3515

Note: G is the number of groups. n is the population for each group. $m \log L$ is the estimated sample average log likelihood. r_{true} is the proportion of simulations for which estimated log likelihood is bigger than that under wrong information structure. r_{censor} is the censoring rate. The numbers in parentheses are standard deviation.

Table 8: Sample Statistics

Variables		Mean	Standard Deviation	Min	Max
Property Tax Rate	Per $100 Valuation	0.3676	(0.1972)	0	0.8200
Population	$\times 10^3$	10.2842	(45.0819)	0.0250	751.9990
Utility Revenue	$ $\times 10^6$	7.3150	(29.6843)	0	344.9110
Median Household Income	$ $\times 10^4$	4.2092	(1.8504)	1.1750	15.7297
Related County Tax Rate	Per $100 Valuation	0.6380	(0.1518)	0.1162	0.9900
No. of Related County		1.1008	(0.3326)	1	4
Distance	Kilometers	216.7976	(125.4035)	1.2133	767.1423
No. of Observations		506			

Table 9: Tobit Model for Property Tax Competition

Regression		Regression for City Property Tax Rates (1)	(2)	(3)	(4)	(5)
Constant	β_0	0.3436***	0.2232***	0.1534***	0.1610***	0.0852**
		(0.0511)	(0.0488)	(0.0535)	(0.0442)	(0.0364)
Population	β_1	0.0005***	0.0006***	0.0004***	0.0007***	0.0004***
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Related County Tax Rate	β_3	0.2822***	0.1574**	0.1454**	0.1652***	0.1387**
		(0.0648)	(0.0647)	(0.0658)	(0.0641)	(0.0585)
Median Household Income	β_4	-0.0393***	-0.0334***	-0.0327***	-0.0311***	-0.0301***
		(0.0043)	(0.0046)	(0.0044)	(0.0044)	(0.0041)
Iteraction Intensity	λ		0.4754***	0.6843***	0.6020***	0.8517***
			(0.1401)	(0.2023)	(0.1311)	(0.1493)
Shock Variance	σ	0.1848***	0.1833***	0.1835***	0.1849***	0.1839***
		(0.0065)	(0.0064)	(0.0064)	(0.0066)	(0.0065)
Estimated log Likelihood		89.6458	93.5224	92.7961	93.7981	96.5846
No. of Observations		506	506	506	506	506
Number of "Neighbors"			10.8656	10.8656	28.3636	28.3636
			(4.3634)	(4.3634)	(9.0854)	(9.0854)
Cutoff Distance (kilometers)			30	30	50	50

Variables		Distribution of Median Household Income μ	ω		η	ρ
Estimates		3.3971	0.7742		2.0059	0.1490

Note: Regression (1) is ordinary Tobit regression without social interactions. Regressions (2) and (4) correspond to the Tobit model when all characteristics are public information. Municipal median household income is assumed to be self-known for municipalities in Regressions (3) and (5). Two municipalities are viewed as close "neighbors" if the distance between them is less than 30 kilometers for Regressions (2) and (3), or less than 50 kilometers for Regressions (4) and (5). Numbers in parentheses are standard deviations. Estimates that are significant at the %10, %5, and %1 levels are marked by "*", "**", and "***", respectively.