Estimation of social-influence-dependent peer pressure in a large network game

ZHONGJIAN LIN† AND HAIQING XU‡

†Department of Economics, Emory University, Room 332, Rich Memorial Building, Atlanta, GA 30322, USA.
E-mail: zhongjian.lin@emory.edu
‡Department of Economics, University of Texas at Austin, 1 University Station, Austin, TX 78712, USA.
E-mail: h.xu@austin.utexas.edu

First version received: August 2015; final version accepted: August 2017

Summary  Research on peer effects in sociology has long been focused on social interactions and the associated social influence process. In this paper, we extend a large-network-based game model to a model that allows for the dependence of social interactions on social-influence status. In particular, we use the Katz–Bonacich centrality to measure individuals’ social influences, which are obtained directly from the observation of a social network. To solve the computational burden when the data come from the equilibrium of a large network, we extend a nested pseudo-likelihood estimation approach to our large-network-based game model. Using the National Longitudinal Study of Adolescent Health (Add Health) dataset, we investigate the peer effects of dangerous behaviour among high-school students. Our results show that the peer effects are statistically significant and positive. Moreover, students benefit more (statistically significant at the 5% level) from conformity or, equivalently, pay more for disobedience, in terms of peer pressure, if their friends have a higher status of social influence.

Keywords: Large network, Nested pseudo-likelihood estimation, Peer effects, Social influence, Social interactions.

1. INTRODUCTION

Game theoretic network models have been successfully used for studying social interactions. A leading example is network formation, which has been studied by, e.g. Jackson and Wolinsky (1996) and Bala and Goyal (2000) for the theory side, and by Christakis et al. (2010), Mele (2010), Badev (2013), Menzel (2016) and de Paula et al. (2017), for example, for the econometrics side. Another example considers social interactions in exogenously given large networks; see, e.g. Blume et al. (2015) and Xu (2017). In this paper, we extend the large-network-based social-interaction model of Xu (2017) by allowing peer effects to depend on social-influence status.

Models with strategic interactions have been widely used in industrial organization, where the literature mainly focuses on player-specific strategic effects; see, e.g. Bjorn and Vuong (1984) and Tamer (2003). Part of the rationale for employing such a specification is that the number of firms that interact with each other is small. In large-network-based social interactions,
homogeneous peer effects have been assumed for model tractability. However, empirical studies have documented that peer effects can be diluted by the presence of many uninfluential friends; see, e.g. Shi and Whinston (2013). Following the network approach in sociology, we use network centrality – a concept introduced in the late 1940s – to measure an individual’s social influence and prestige. Such a measure is motivated by the delineation of the multiple motives that spur interpersonal influence in the sociology literature – namely, to behave effectively, to build and maintain social relationships, and to manage self-concept.1 The first two goals suggest that social interactions are ‘not indiscriminate’ and that social-influence status matters (Allison, 1992), and also that people should be more influenced by those who have visible signs of success, such as wealth, power or social status (Cialdini and Trost, 1998). For example, Stack (1990) and Jonas (1992) found significant Werther effects, with juveniles killing themselves after both celebrity and non-celebrity suicides, but with significantly larger increases after celebrity suicides. In an empirical study of adolescent delinquent behaviour, Haynie (2001) found that friends with higher network centrality are more influential than those with lower centrality. Although emphasizing the role of social-influence status in peer effects, the empirical sociology literature does not employ a structural approach to deal with simultaneity in social interactions.

In this paper, we build upon the large-network-based social-interaction model of Xu (2017). In particular, we assume that an individual’s payoff from a decision depends on their own covariates, as well as on the choices of their direct friends. Following the fundamental principle in sociology, we assume that players benefit from choosing the same action as their friends. A key feature of our model is that we allow peer pressures for conformity to vary with friends’ (relative) social influence/prestige, as measured by the Katz–Bonacich (KB) centrality; see Katz (1953) and Bonacich (1987). In sociology, a network approach has been developed to use centrality measures to characterize an individual’s social status; see, e.g. Freeman (1978), Knoke and Burt (1983) and Borgatti and Everett (2006). All these centrality measures are consistent with the view that an individual’s social-influence status is inherently tied to the status of their associates. In our empirical application, the National Longitudinal Study of Adolescent Health (Add Health) dataset contains the KB centrality measure. Note that various measures of centrality are available in the social network literature and, in principle, one could use other centrality indices to measure social-influence status. As long as different centrality measures rank individuals’ social-influence status in the same order, it is unclear the extent to which an alternative measure would alter our empirical results. In a recent review, de Paula (2017) provides a detailed discussion on several measures of centrality (see also references therein).

To our knowledge, only a handful of papers have considered social-influence status (measured by network centralities) in a structural peer-effects analysis. In the spatial autoregressive model, Calvó-Armengol et al. (2009) show that the Nash equilibrium outcome of each individual in the network is proportional to the individual’s KB centrality measure, which is assumed to capture all the direct and indirect influences of the network on a given individual. An important empirical finding of their approach is that the more central (in terms of KB centrality) a person is in a network, the higher their outcome. In contrast, we do not construct the network centrality measure from direct and indirect peer effects, but rather we use the KB centrality as an exogenous observable. Observations of this measure are obtained directly from the Add Health dataset. Another important related paper is that of Liu and Lee (2010), who use the KB centrality

1 Note that the three motives are the goals of the influence target, instead of the goals of the influence agent, as ‘the more intriguing and instructive questions concerned not so much the reasons that someone would choose to influence another as the reasons that someone would choose to yield to influence from another’. See Cialdini and Trost (1998).
as an instrumental variable for peer effects in a linear social-interaction model. In our structural approach, we assume that the KB centralities of friends affect the level of peer pressure on a player and therefore affect the individual’s outcome directly.

To address the computational burden for solving the equilibrium of a large-network-based game, we apply the nested pseudo-likelihood estimation (NPLE) approach of Aguirregabiria and Mira (2007) to estimate our large-network-based game model. It is a natural idea to extend the NPLE approach to large network games. Similar to dynamic games, the large dimensionality of the equilibrium strategy profile in a large network game makes it costly to compute the equilibrium using fixed point algorithms. The NPLE approach starts with an arbitrary guess of the choice probabilities (e.g. the predicted choice probabilities from the standard Logit estimation without strategic interactions). We then conduct another Logit estimation by using the predicted choice probabilities of friends as an individual’s expectation on their friends’ equilibrium behaviour. After that, we obtain an update of the predicted choice probabilities. We repeat this updating procedure until it converges. Therefore, NPLE is an iterative algorithm that consists of a sequence of Logit estimations. In a large social network game, the NPLE is attractive to practitioners because of its simplicity of implementation and faster computation.

Using the Add Health dataset, we investigate the peer effects on the dangerous behaviour of high-school students. Our results show that peer effects are statistically significant and positive. Moreover, if friends choose ‘not conducting dangerous behaviours’, then a high-school student benefits more (statistically significant at the 5% level) from their own conformity or, equivalently, pays more for their own disobedience in terms of peer pressure, if their friends have higher social-influence status. We also compare the results from our model with Xu (2017) and the standard Logit model. In particular, the peer effects are insignificant in the approach of Xu (2017) as a result of attenuation bias. In the Logit model, the coefficient estimate for friends’ social-influence status is negative and statistically significant at the 5% level. This result suggests a negative correlation between players’ decisions and their friends’ social-influence status. However, it is implausible to give such a result a meaningful economics interpretation.

The rest of the paper is organized as follows. In Section 2, we introduce our model and the definition of the KB centrality. We also characterize the equilibrium and establish its uniqueness. In Section 3, we establish the identification of structural parameters and we define the NPLE algorithm. Asymptotic properties for NPLE are also established. In Section 4, we apply our estimation method to study peer effects on the dangerous behaviour of high-school students. Proofs of our results are collected in the Appendix. Moreover, Monte Carlo experiments are presented in an online Appendix, available on the publisher’s web site.

2. A MODEL OF SOCIAL INTERACTIONS IN LARGE NETWORKS

We consider a discrete game played on an existing large social network. The network is viewed as a random graph with vertices connected with directed edges. In the graph, each individual $i \in \mathcal{I} \equiv \{1, \ldots, n\}$ is represented by a vertex, who is connected to a group of best friends, represented by directed edges. Let $\ell_{ij} = 1$ if individual $i$ nominates $j$ as a best friend; $\ell_{ij} = 0$ otherwise. Following convention, let $\ell_{ii} = 0$ for all $i \in \mathcal{I}$. Moreover, we denote $F_i \equiv \{j \in \mathcal{I} : \ell_{ij} = 1\}$ as the group of $i$’s best friends. By definition, a best-friend relationship need not be symmetric in our directed network; in other words, $\ell_{ij} \neq \ell_{ji}$ is allowed. Furthermore, we denote the network graph by an $n \times n$ matrix $\mathbb{L}$, where the $ij$th entry is $\ell_{ij}$. © 2017 Royal Economic Society.
Using graph theory, different metrics are developed to quantify the influence of every node within a network; see, e.g. Borgatti and Everett (2006). In directed networks, for instance, Knoke and Burt (1983) use the number of outgoing links and the number of incoming links to measure influence and support, respectively. Such degree measures, however, are criticized for not taking into account the indirect connections to all the individuals in the network, but only immediate connections. Thus, we use the KB centrality measure as our social-influence metric, as suggested by, for example, Bonacich (1987) in the sociology literature. Specifically, for \( i = 1, \ldots, n \), let

\[
S_i = \sum_{k=1}^{\infty} \sum_{j=1}^{n} \lambda^k \times (L^k)_{ij}.
\]  

(2.1)

For given attenuation factor \( \lambda \in (0, 1) \), \( S_i \) in (2.1) is the KB centrality measure for individual \( i \). Note that \( \sum_{j=1}^{n} (L^k)_{ij} \) is the number of distinct paths of length \( k \) leading away from \( i \), where \( (L^k)_{ij} \) counts the number of paths of length \( k \) that connect \( i \) to \( j \); see Section 2.1.3 of Jackson (2010). By definition, \( S_i = 0 \) if \( F_i = \emptyset \).

For our empirical application, the Add Health dataset contains such a measure with \( \lambda = 0.1 \). Figure 1 provides a probability distribution of \( S_1 \) (conditional on \( S_i > 0 \)). In particular, the shadow area is the (pointwise) 95% confidence interval.

In our network game, each individual simultaneously chooses \( Y_i \in \{0, 1\} \). In our empirical application, \( Y_i = 1 \) refers to student \( i \) conducting dangerous behaviour in a recent period. Then, the utility function of \( i \) is given by

\[
U_i = \begin{cases} 
X_i' \beta_1 + \frac{1}{Q_i} \sum_{j \in F_i} \alpha_1(S_j - S_i) \times 1(Y_j = 1) - \epsilon_{0i}, & \text{if } Y_i = 1; \\
X_i' \beta_0 + \frac{1}{Q_i} \sum_{j \in F_i} \alpha_0(S_j - S_i) \times 1(Y_j = 0) - \epsilon_{0i}, & \text{if } Y_i = 0.
\end{cases}
\]  

(2.2)

Here, \( X_i \in \mathbb{R}^d \) includes a constant and a vector of individual \( i \)’s demographic characteristics, \( \epsilon_{0i}, \epsilon_{1i} \in \mathbb{R} \) are unobserved action-dependent utility shocks and \( Q_i = \sum_{j=1}^{n} \ell_{ji} = \sum_{j=1}^{n} 1(j \in F_i) \) denotes the total number of friends. For expositional simplicity, we assume \( Q_i \geq 1 \) in (2.2).\(^2\) Moreover, \( \beta_0, \beta_1 \in \mathbb{R}^d \) are payoff coefficients and \( \alpha_0(\cdot) \) and \( \alpha_1(\cdot) \) are unknown structural

\(^2\) The case of \( Q_i = 0 \) can be accommodated simply by letting \( U_i = X_i' \beta_d \) if \( Y_i = d \).
functions. In particular, $\alpha_1$ measures the peer pressure on player $i$ from their friend $j$ when choosing the same action (i.e. $Y_i = Y_j = 1$). A similar interpretation applies to $\alpha_0$. Furthermore, $\alpha^\dagger(S_j - S_i) \equiv \alpha_1(S_j - S_i) + \alpha_0(S_j - S_i)$ measures the total peer effects that lead to conformity among friends in a social network. To achieve identification, we need to normalize the payoff coefficient of $X_i$ for action 0 (i.e. $\beta_0 = 0$), because the covariates $X_i$ are included in the payoff indices of both actions.

In the above payoff function, a key feature is that $\alpha_1(S_j - S_i)$ and $\alpha_0(S_j - S_i)$ depend on friend $j$’s relative social influence $S_j - S_i$. Such a specification on peer effects is related to social-influence models used in sociology; see, e.g. Friedkin and Johnsen (1990). In our empirical application, we investigate the following question. If an individual’s friends choose (not) to conduct dangerous behaviour, does the amount of peer pressure from friends increase with their social-influence status?

Next, we specify the information structure. Let $X_i$ and $L$ be publicly observed state variables and let $(\epsilon_{0i}, \epsilon_{1i})$ be player $i$’s private information. Note that because $F_i$ and $S_i$ are derived from $L$, they are also publicly observed state variables. Let $\mathbb{W} \equiv \{(X'_1, \ldots, X'_n); L\}$ be all the public information in the game. According to the Bayesian Nash equilibrium (BNE) solution concept, the best response function is given by

$$R_i(\mathbb{W}, \epsilon_i) = 1 \left\{ X'_i(\beta_1 - \beta_0) - \sum_{j \in F_i} \frac{\alpha_0(S_j - S_i)}{Q_i} + \sum_{j \in F_i} \frac{\alpha^\dagger(S_j - S_i) \Pr(Y_j = 1|\mathbb{W})}{Q_i} - \epsilon^*_i \geq 0 \right\},$$

where $\epsilon^*_i = \epsilon_1 - \epsilon_0$. In equilibrium, players’ decisions can be written by

$$Y_i = R_i(\mathbb{W}, \epsilon_i), \quad \forall i \leq n.$$

It is worth pointing out that we maintain a controversial assumption throughout the paper that the network connections are exogenous. In our empirical application on peer effects of the dangerous behaviour of high-school students, for instance, the social-influence measure (i.e. KB centrality) could be affected by students’ participation in risky behaviour. In other words, besides maximizing a player’s ‘action-specific’ payoff, another reason for a player choosing dangerous behaviour is to build up network connections (i.e. ‘to build and maintain social relationships’ for future social interactions, and/or ‘to manage self-concept’). Clearly, such a feedback effect from the action to the network raises an issue of endogeneity, which is an intriguing problem for further research. In practice, it is helpful to avoid/mitigate such an issue, whereby a researcher uses observations on network connections measured before the decisions of outcome variables are made by players.

### 2.1. Equilibrium characterization

To characterize the equilibrium, we first make an assumption on the distribution of $\epsilon_i$.

**Assumption 2.1.** The error terms $\{(\epsilon_{0i}, \epsilon_{1i}) : i \leq n\}$ are independent and identically distributed (i.i.d.) across both actions and players. Furthermore, the error term has an extreme value type I distribution with density

$$f(t) = \exp(-t) \exp(-\exp(-t)).$$
Assumption 2.1 is fairly standard in the discrete choice model literature; see, e.g. Bajari et al. (2012). Assumption 2.1 provides a closed-form expression for players’ best responses in terms of choice probabilities.

Let $\sigma^*_i(W) = \Pr(Y_i = 1|W)$ be the equilibrium choice probability of choosing action 1. Further, let $\tilde{\alpha}_i(\cdot) = \alpha_0(\cdot)/Q_i$ and $\tilde{\alpha}^*_i(\cdot) = \alpha^*(\cdot)/Q_i$. Under Assumption 2.1, the best-response function can be written in terms of equilibrium choice probabilities. For $i = 1, \ldots, n$,

$$
\sigma^*_i(W) = \frac{\exp(X_i'\beta_1 - \beta_0) - \sum_{j \in F_i} \tilde{\alpha}_i(S_j - S_i) + \sum_{j \in F_i} \tilde{\alpha}^*_i(S_j - S_i)\sigma^*_j(W)}{1 + \exp(X_i'\beta_1 - \beta_0) - \sum_{j \in F_i} \tilde{\alpha}_i(S_j - S_i) + \sum_{j \in F_i} \tilde{\alpha}^*_i(S_j - S_i)\sigma^*_j(W)}.
$$

To ensure that equation (2.3) admits a unique solution, we next introduce an assumption on the strength of peer effects. Let $S_0$ be the support of $S_j - S_i$ where $j \in F_i$.

**Assumption 2.2.** Let $\sup_{s \in [\alpha_0(s) + \alpha_1(s)]} < 4$.

Under Assumption 2.2, the dependence of the equilibrium choices satisfies the mixing conditions, which serve as a key to dependent data analysis. Similar assumptions for equilibrium uniqueness in Bayesian games can also be found in, for example, Brock and Durlauf (2001), Horst and Scheinkman (2006) and Xu (2017).

**Theorem 2.1.** Under Assumptions 2.1 and 2.2, there exists a unique pure strategy BNE for any $n$.

Theorem 2.1 is important for statistical inference on large network games. When there are multiple equilibria, an obvious obstacle for statistical inference is the incompleteness of the econometric model. For more discussion on the issues of multiple equilibria, see, for example, Tamer (2010, 2003) and de Paula (2013).

### 3. IDENTIFICATION AND ESTIMATION

For tractability, we linearize $\alpha(\cdot)$ for our empirical analysis. Let $\phi_0(s) = \phi_0 + \phi_1 \times s$ and $\alpha_1(s) = \psi_0 + \psi_1 \times s$. By definition, $\alpha^*(s) = \alpha_0(s) + \alpha_1(s) = \phi_0 + \psi_0 + (\phi_1 + \psi_1) \times s$. By the equilibrium condition (2.3), $\beta_1$ and $\beta_0$ cannot be separately identified in the structural model as only their difference $\beta_1 - \beta_0$ matters for the equilibrium. Therefore, we set $\beta_0 = 0$ as a normalization.

#### 3.1. Identification

First, note that $\sigma^*_i(W)$ can be obtained directly from the distribution of observables. Following the definition of identification – see, e.g. Hurwicz (1950) – we treat $\sigma^*_i(W)$ as a known object. Let $T_i = \ln \sigma^*_i(W) - \ln(1 - \sigma^*_i(W))$. It follows by (2.3) that

$$
T_i = X_i'\beta_1 - \phi_0 - \phi_1 \sum_{j \in F_i} \tilde{S}_{ij} Q_i + (\phi_0 + \psi_0) \sum_{j \in F_i} \sigma^*_j(W) Q_i + (\phi_1 + \psi_1) \sum_{j \in F_i} \tilde{S}_{ij} \sigma^*_j(W) Q_i,
$$

3 See Xu (2017) for a detailed discussion on the definition of identification in a single large network game model.
where $\tilde{S}_j = S_j - S_i$. Note that $\phi_0$ cannot be separately identified from the constant term of $X'\beta_1$. Therefore, let $\phi_0 = 0$ as a normalization. It follows that

$$T_i = X'_i\beta_1 - \phi_1 \frac{\sum_{j \in F_i} \tilde{S}_j (1 - \sigma_j^*(\mathbb{W}))}{Q_i} + \psi_0 \frac{\sum_{j \in F_i} \sigma_j^*(\mathbb{W})}{Q_i} + \psi_1 \frac{\sum_{j \in F_i} \tilde{S}_j \sigma_j^*(\mathbb{W})}{Q_i},$$

which takes a linear expression of structural parameters.

Let $\theta \equiv (\beta_1, \phi_1, \psi_0, \psi_1) \in \Theta \subseteq \mathbb{R}^d \times \mathbb{R}_+^3$, where $\Theta$ is the parameter space. The positiveness of $\phi_1$, $\psi_0$ and $\psi_1$ reflects the fundamental principle in sociology that friends benefit from conformity. Let $\theta_0$ and $\theta$ be the true parameters for the data-generating process and a generic value in $\Theta$, respectively. Moreover, we denote

$$Z_i \equiv \left(X'_i, \frac{\sum_{j \in F_i} \tilde{S}_j (\sigma_j^*(\mathbb{W}) - 1)}{Q_i}, \frac{\sum_{j \in F_i} \sigma_j^*(\mathbb{W})}{Q_i}, \frac{\sum_{j \in F_i} \tilde{S}_j \sigma_j^*(\mathbb{W})}{Q_i}\right) \in \mathbb{R}^{d + 3}.
$$

**Assumption 3.1.** $E[Z_i Z'_i]$ has full rank, i.e. $\text{Rank}(E[Z_i Z'_i]) = d + 3$.

Assumption 3.1 requires no perfect collinearity of $Z_i$. This assumption is essentially a full rank condition. As is pointed out in Bajari et al. (2012), it is the exclusive payoff shifters of other players that induce independent variations in player $i$’s beliefs, which render the rank condition meaningful. In our model, such a full rank condition holds if: (a) $X_i$ has no perfect collinearity; (b) conditional on $(X_i, S_i, Q_i, \{S_j : j \in F_i\})$, $\{\sigma_j^*(\mathbb{W}) : j \in F_i\}$ has no perfect collinearity; and (c) for every $j \in F_i$, conditional on $(X_i, F_i/\{j\})$, we have $0 < \mathbb{P}(j \in F_i) < 1$. In particular, the last condition requires variations in $Q_i$, given $X_i$. It should also be noted that Assumption 3.1 is testable as $Z_i$ can be non-parametrically estimated (see Xu, 2017).

**Lemma 3.1.** Suppose Assumptions 2.1, 2.2 and 3.1 hold. Then $\theta_0$ is identified by

$$(E[Z_i Z'_i])^{-1} E[Z_i T_i].$$

The proof directly follows our discussions above and is therefore omitted.

It is worth pointing out that if the relative social influence of friends has a strictly positive effect on peer pressure (i.e. $\phi_1 > 0$ and/or $\psi_1 > 0$), then ignoring such an effect will necessarily induce omitted variable bias to the estimation of peer effects. To see this, suppose equilibrium beliefs $\{\sigma_j^*(\mathbb{W}) : j = 1, \ldots, n\}$ are observed in the data. Then a Logit estimation without including $(\sum_{j \in F_i} \tilde{S}_j (\sigma_j^*(\mathbb{W}) - 1))/Q_i$ and $(\sum_{j \in F_i} \tilde{S}_j \sigma_j^*(\mathbb{W}))/Q_i$ as regressors would be inconsistent because of their correlation with $(\sum_{j \in F_i} \sigma_j^*(\mathbb{W}))/Q_i$, which is the regressor for the constant peer effect coefficient $\psi_0$.

### 3.2. Estimation

Our estimation follows the NPLE approach of Aguirregabiria and Mira (2007). Similar to their dynamic setting, the difficulties in large network games arise from the computational burden of solving the equilibrium. Using an iterative algorithm, the NPLE significantly reduces the computational burden, but it is less efficient than the maximum likelihood estimation (MLE) approach of Xu (2017). More importantly, the proposed approach is essentially a sequence of Logit estimations, which is easy to implement.

Consider a random sample $\{(Y_i, X_i, F_i) : i = 1, \ldots, n\}$ from a single large social network. It is worth noting that our approach can be easily extended to applications where observations...
come from a small number of networks but each network has a large size. In both cases, our asymptotic analysis relies on the number of players going to infinity.

Under this parametric specification, we are particularly interested in testing $H_0: \phi_1 = \psi_1 = 0$ versus $H_1: \phi_1 \neq 0 \text{ or } \psi_1 \neq 0$. Rejection of the null hypothesis provides evidence for causal effects on peer pressure from friends’ social influence.

3.3. NPLE algorithm

Let $\Sigma^*(\mathbb{W}) = (\sigma_1^*(\mathbb{W}), \ldots, \sigma_n^*(\mathbb{W}))'$ and $\Sigma = (\sigma_1, \ldots, \sigma_n)' \in [0, 1]^n$ be the equilibrium choice probability profile and a generic probability profile, respectively. For arbitrary $\Sigma \in [0, 1]^n$, let

$$Z_i(\Sigma) \equiv \left( \frac{X_i', \sum_{j \in F_i} \tilde{S}_{ij}(\sigma_j - 1)}{Q_i}, \frac{\sum_{j \in F_i} \sigma_j}{Q_i}, \frac{\sum_{j \in F_i} \tilde{S}_{ij}\sigma_j}{Q_i} \right)'$$

and

$$\Gamma_i(\Sigma, \theta; \mathbb{W}) = \frac{\exp(Z_i'(\Sigma)\theta)}{1 + \exp(Z_i'(\Sigma)\theta)}.
$$

Moreover, we denote $\Sigma(\theta; \mathbb{W})$ as the solution to the equation system:

$$\Gamma_i(\Sigma, \theta; \mathbb{W}) = \sigma_i, \quad \forall \ i \leq n.$$ 

By definition, $\Sigma^*(\mathbb{W}) = \Sigma(\theta_0; \mathbb{W})$. Recall that under the conditions in Theorem 2.1, the above equation system has a unique solution $\Sigma(\theta_0; \mathbb{W})$ for the given structural parameter $\theta_0 \in \Theta$.

Furthermore, we define a pseudo log-likelihood function:

$$\hat{L}_n(\theta, \Sigma) = \frac{1}{n} \sum_{i=1}^n \left( Y_i \ln \Gamma_i(\Sigma, \theta; \mathbb{W}) + (1 - Y_i) \ln(1 - \Gamma_i(\Sigma, \theta; \mathbb{W})) \right), \quad \theta \in \Theta.
$$

Note that $\hat{L}_n(\cdot, \Sigma)$ becomes the true log-likelihood function if $\Sigma$ is set to be $\Sigma(\cdot; \mathbb{W})$ as a function of $\theta$. As the network size $n$ goes to infinity, we define further the limit of the pseudo log-likelihood function by

$$L(\theta, \Sigma) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n E[Y_i \ln \Gamma_i(\Sigma, \theta; \mathbb{W}) + (1 - Y_i) \ln(1 - \Gamma_i(\Sigma, \theta; \mathbb{W}))].$$

Given the above notation, we are ready to describe our estimation procedure. First, we start with an arbitrary initial value $\Sigma^{[0]} \in [0, 1]^n$, without loss of generality, let $\Sigma^{[0]} = (0, \ldots, 0) \in [0, 1]^n$. Next, we iterate the following two steps.

**STEP 1.** Given $\Sigma^{[j-1]}$, let

$$\hat{\theta}^{[j]} = \arg \max_{\theta \in \Theta} \hat{L}_n(\theta, \Sigma^{[j-1]}).$$

**STEP 2.** Given $\hat{\theta}^{[j]}$, let

$$\Sigma^{[j]} = \Gamma(\Sigma^{[j-1]}, \hat{\theta}^{[j]}; \mathbb{W}),$$

© 2017 Royal Economic Society.
where \( \Gamma(\Sigma, \theta; \mathbb{W}) = (\Gamma_1(\Sigma, \theta; \mathbb{W}), \ldots, \Gamma_n(\Sigma, \theta; \mathbb{W}))' \). This procedure stops at the \( K \)th iteration when \( \|\hat{\theta}^{[K]} - \hat{\theta}^{[K-1]}\| \) is less than a predetermined tolerance (e.g. \( 10^{-6} \)). We then define our estimator by \( \hat{\theta}_{\text{NPLE}} = \hat{\theta}^{[K]} \).

The local convergence of the NPLE algorithm is ensured by the local contraction condition established in Kasahara and Shimotsu (2012). In our framework, however, it is difficult to verify their conditions for the convergence of the NPLE algorithm. It has been noticed in the literature, and we have experienced in our Monte Carlo experiments and empirical application, that the NPLE algorithm typically converges to the same fixed point, regardless of the initial values; see also, e.g. Aguirregabiria and Mira (2007).

By definition, the above NPLE is essentially a fixed point solution to maximize the log-likelihood function, which can be equivalently defined by

\[
\hat{\theta}_{\text{NPLE}} = \arg\max_{\theta \in \Theta} \hat{L}_n(\theta, \Sigma) \quad \text{and} \quad \Sigma = \Gamma(\Sigma, \theta; \mathbb{W}).
\]

See Aguirregabiria and Mira (2007) for a more detailed discussion.

3.4. Asymptotic analysis

We make further assumptions to establish the asymptotic properties of \( \hat{\theta}_{\text{NPLE}} \).

**Assumption 3.2.** The underlying parameter \( \theta_0 \) uniquely solves the following equation:

\[
\theta = \arg\max_{c \in \Theta} L(c, \Sigma(\theta; \mathbb{W})). \tag{3.2}
\]

Assumption 3.2 is essentially an identification assumption for using the NPLE algorithm. It is straightforward to show that \( \theta_0 \) solves (3.2): \( \theta_0 \) maximizes \( L(\cdot, \Sigma^*(\mathbb{W})) \) by the standard argument for the MLE method, and \( \Sigma^*(\mathbb{W}) = \Sigma(\theta_0; \mathbb{W}) \). Without Assumption 3.2, however, (3.2) might admit multiple solutions, and each of these represents a fixed point of the NPLE algorithm.

As is also suggested by Aguirregabiria and Mira (2007), one should try different initial values of \( \Sigma^{[0]} \) to see whether the algorithm converges to the same fixed point. Assumption 3.2 could be invalidated if there are multiple fixed points. When this happens, one should always select the fixed point that maximizes the value of the pseudo log-likelihood. Not surprisingly, the computation time of the algorithm relies on the number of starting values the researcher initiates. In general, when the game structure admits multiple equilibria, it might be difficult to detect some fixed point where beliefs come from an unstable equilibrium by using such a method. See Pesendorfer and Schmidt-Dengler (2010) for a numerical example.

**Assumption 3.3.** \( \mathcal{S}_X \) is bounded and \( \Theta \) is compact.

Assumption 3.3 ensures that choice probabilities derived from the model are uniformly bounded away from zero, which implies that \( \hat{L}_n(\cdot, \Sigma^{[j]}) \) is also uniformly bounded for all \( j \).

**Assumption 3.4.** Let \( \max_{j \in \{1, \ldots, n\}} \sum_{j=1}^{n} \ell_{ij} \leq M \) for some constant \( M \in \mathbb{N}^+ \).

Assumption 3.4 is needed to limit the dependence among all the observations. In our Add Health dataset, \( M = 10 \), which comes from the fact that at most 10 nominations were permitted in the survey design.
For any $h \in \mathbb{N}$ and $i \in \mathcal{I}$, let $N_{(i,h)} = \{ j \in \mathcal{I} : (L^k)_{ij} \geq 1 \text{ for some } k \leq h \}$. Moreover, let $L^{(i,h)}$ be a $\#N_{(i,h)} \times \#N_{(i,h)}$ submatrix of $L$, which describes the graph for the subnetwork among $N_{(i,h)}$.

**Assumption 3.5.** Fix arbitrary $h \in \mathbb{N}$. The probability distribution of $L^{(i,h)}$ converges to a limiting distribution as $n \to \infty$ for all $i$; $L^{(i,h)}$ is independent of $L^{(j,h)}$ if $N_{(i,h)} \cap N_{(j,h)} = \emptyset$. Moreover, the payoff covariates $X_i$ are i.i.d. across players given the exogenous random network.

In the large network asymptotics, Assumption 3.5 is also made in Xu (2017) for the consistency of an MLE-type estimator. In particular, the first part of Assumption 3.5 requires that the distribution of subgraphs converge to a limit as the network size goes to infinity. The second part means two non-overlapping subgraphs have independent connecting structures. This condition generally holds in the random graph literature because, conditional on $L^{(i,h)}$ and $L^{(j,h)}$ not overlapping, the graph structure of $L^{(i,h)}$ does not provide additional information on the graph of $L^{(j,h)}$ if each link is determined independently. The last part of Assumption 3.5 is a strong assumption. In practice, it has been emphasized in the sociology literature that friends’ demographic variables (e.g. age, education, race) are usually positively correlated; see, e.g. Easley and Kleinberg (2010). Following the spatial econometrics literature, it is possible to relax this assumption by allowing for some degree of dependence, as long as the statistical dependence of two players’ covariates decays sufficiently fast (e.g. at some exponential rate) with their network distance.

**Theorem 3.1.** Suppose Assumptions 2.1, 2.2 and 3.1–3.5 hold. In particular, Assumption 2.2 holds for all $\theta \in \Theta_1$. Then

$$\hat{\theta}_{\text{NPLE}} \xrightarrow{p} \theta_0.$$ 

In Theorem 3.1, we need to restrict the parameter space for $\alpha(\cdot)$ such that Assumption 2.2 holds for all $\theta \in \Theta$. Similar to the stationarity restriction in the autoregressive model, such a condition imposes restrictions on $(\phi_1, \psi_0, \psi_1)$ that depend on the support of $S_j - S_i$.

Following Aguirregabiria and Mira (2007), we now derive the limiting distribution of $\hat{\theta}_{\text{NPLE}}$. Let

$$A_n = \frac{1}{n} \sum_{i=1}^{n} E[Z_i Z_i' \sigma_i^*(\mathbb{W})(1 - \sigma_i^*(\mathbb{W})) + \frac{1}{Q_i} \sum_{j \in F_i} Z_i Z_j' \sigma_j^*(\mathbb{W})(1 - \sigma_j^*(\mathbb{W}))(\psi_0 + (\phi_1 + \psi_1) \tilde{S}_{ji})]$$

and

$$B_n = \frac{1}{n} \sum_{i=1}^{n} E[Z_i Z_i'(Y_i - \sigma_i^*(\mathbb{W}))^2].$$

Note that $A_n$ and $B_n$ depend on index $n$ through $\mathbb{W}$.

**Assumption 3.6.** $\theta_0$ belongs to the interior of $\Theta$.

**Assumption 3.7.** There exist non-singular $(d + 3) \times (d + 3)$ matrices $A$ and $B$ such that $A_n \to A$ and $B_n \to B$.

Assumption 3.6 is standard. Assumption 3.7 is a high-level condition that requires $A_n$ and $B_n$ to converge to some non-singular limiting matrices, respectively, as the network size goes.
to infinity. Such a condition could be derived by specifying a network growing mechanism. Moreover, the non-degeneracy of $A$ and $B$ requires that all the determinants of $A_n$ and $B_n$ are outside an open ball of zero for all $n$, which is essentially a rank condition.

**Theorem 3.2.** Suppose that all the conditions in Theorem 3.1 and Assumptions 3.6 and 3.7 hold. We then have

$$\sqrt{n}(\hat{\theta}_{\text{NPLE}} - \theta_0) \overset{d}{\to} N(0, \Omega_0),$$

where $\Omega_0 = A^{-1}BA^{-1}$.

We have also investigated the finite sample performance of $\hat{\theta}_{\text{NPLE}}$ by using Monte Carlo experiments. In all of our designs, the estimator behaves well in terms of average bias, standard deviation and mean square error. See our online Appendix for more details.

## 4. EMPIRICAL APPLICATION: DANGEROUS BEHAVIOUR OF HIGH-SCHOOL STUDENTS

In this section, we apply our method to study peer effects of dangerous behaviour on high-school students. Adolescent risky behaviour has been studied in terms of peer effects; see, e.g. Nakajima (2007) and Gaviria and Raphael (2001). To the best of our knowledge, however, there has been no structural analysis in the literature on peer effects that are dependent on social influence. In particular, the research question we ask is the following. How do students of high social-influence status, who typically are less likely to behave dangerously, affect their peers through the network? There is no doubt that they should affect more people given their high centrality in the network, but do they impose more peer pressure on their followers than ordinary peers? In this paper, we use the self-report questionnaires from the Add Health dataset to study this empirical question.

### 4.1. Add Health Dataset

The Add Health study is a longitudinal study of a nationally representative sample of adolescents in grades 7–12 in the United States during the 1994–1995 school year. Add Health combines longitudinal survey data on respondents’ social and economic features with contextual data on family, friendships and peer groups. In the dataset, each student has nominations of at most five male friends and at most five female friends, which allows us to construct a social network among observations. From the Wave I survey, we have 85,627 students from more than 100 representative schools in all regions of the United States. In this study, we pick a pair of sister schools (i.e. No. 62 and No. 162), with a significant proportion of inter-school friend nominations. Our sample contains 2,460 students. Table 1 provides summary statistics of the observables.

Observed demographic characteristics include age and gender, as well as the KB centrality measure. The average KB centrality measure of friends is constructed from the friends’ nominations. The dependent variable is constructed using the self-report questionnaires in the Add Health dataset. Specifically, the survey question is as follows. ‘During the past twelve months, how often did you do something dangerous because you were dared to?’
Table 1. Summary of statistics of key variables from the data.

| Variable                        | Min | Max | Mean | Std deviation |
|---------------------------------|-----|-----|------|---------------|
| Risky behaviour                 | 0   | 1   | 0.44 | 0.50          |
| Age                             | 10  | 19  | 15.18| 1.64          |
| Female                          | 0   | 1   | 0.52 | 0.54          |
| KB centrality                   | 0   | 3.14| 0.82 | 0.60          |
| Average KB centrality of friends| 0   | 2.68| 0.84 | 0.49          |
| Number of friends               | 0   | 10  | 4.86 | 2.86          |

4.2. Empirical results

In our application, we choose $\Sigma^{[0]} = (0, \ldots, 0)$ as the initial value for the NPLE algorithm.\(^4\) Thus, the first iterative of the algorithm is a standard Logit estimation. Table 2 reports our estimation results. Clearly, male students are ‘dared’ to do dangerous things more than female students, and students of higher social-influence status are less likely to engage in dangerous behaviour. The effects of age, however, are not significant. Moreover, we find significant peer pressure on players to choose the same action when their friends choose dangerous behaviour.

Table 2. Estimation results.

| Variable | Our model | CPE model | Logit model |
|----------|-----------|-----------|-------------|
| Age      | $-0.020$  | $-0.006$  | $-0.018$    |
|          | (0.026)   | (0.025)   | (0.026)     |
| Female   | $-0.778^{**}$ | $-0.783^{**}$ | $-0.796^{**}$ |
|          | (0.084)   | (0.083)   | (0.083)     |
| Own $S_i$| $-0.404^{**}$ | $-0.162^{**}$ | $-0.309^{**}$ |
|          | (0.095)   | (0.075)   | (0.086)     |
| Avg. friends’ $\bar{S}_{ji}$ | –         | –         | $-0.510^{**}$ |
| Constant | 0.551     | 0.258     | 0.718*      |
|          | (0.423)   | (0.416)   | (0.413)     |
| $\phi_1$ | 1.694**   | –         | –           |
|          | (0.603)   | –         | –           |
| $\psi_0$ | 0.707**   | 0.369     | –           |
|          | (0.315)   | (0.310)   | –           |
| $\psi_1$ | 0.774     | –         | –           |
|          | (0.764)   | –         | –           |
| $\phi_1 + \psi_1$ | 2.468** | –         | –           |
|          | (1.335)   | –         | –           |

Note: ** and * denote significance at the 5% and 10% levels, respectively. Significance of peer effects obtained from the one-sided test.

\(^4\) To verify uniqueness of the NPL fixed point, we also tried several other initial values (e.g. $\Sigma^{[0]} = (\omega_1, \ldots, \omega_n) \in [0, 1]^n$, where $\{\omega_i : i \leq n\}$ are independently drawn from $U[0, 1]$) for our estimation procedure. The NPLE algorithm converges to the same fixed point as is reported in Table 2.
However, such peer effects are insignificantly affected by friends’ relative social-influence status. However, when friends choose to avoid dangerous behaviour, we find significant effects on peer pressure from friends’ social-influence status. That is, given that friends make the decision to refrain from dangerous behaviour, students benefit more from their own conformity or pay more for their disobedience, if their friends are of high social influence rather than if their friends have low social-influence status.

We also compare the results from our model with two other models: the constant peer effects (CPE) model in Xu (2017) and the standard Logit model. Note that the CPE model is nested in our model. Thus, we estimate the CPE model by using our NPLE.\(^5\) Moreover, our model and the CPE model are structural approaches while the Logit model is a reduced-form model. The coefficient estimates for age and gender are quite similar across the three models. The effects from one’s own social-influence status are similar in our model and the Logit model. Both \(\phi_1\) and \(\psi_0\) are statistically significant at the 5% level. In contrast, peer effects are insignificant in the CPE model. This result suggests that the nested CPE model might be misspecified and might give peer-effect estimates that are a mixture of effects from both friends with high social-influence status and many uninfluential friends. Furthermore, we also incorporate the average relative social influence of friends into the Logit model, even though an economics interpretation for the coefficient is implausible. The estimate of its coefficient is negative and statistically significant at the 5% level. Such a result suggests a negative correlation between players’ decisions and their friends’ social-influence status.

Our estimates are related to those empirical results in the literature of peer effects on risky behaviour. Using the same dataset, Haynie (2001) investigated the delinquent behaviour of adolescents (i.e. ‘Smoked cigarettes’, ‘Drank alcohol’, ‘Got drunk’, ‘Raced on bike, boat or car’, ‘Been in danger due to dare’ and ‘Skipped school without an excuse’). Using a negative binomial regression model, she found that the CPE model yields smaller estimates of peer effects than the estimates from the model incorporating network centrality in peer effects (0.04 with a standard error 0.00 versus 0.05 with a standard error 0.00). In her benchmark model, however, the empirical results of Haynie (2001) also indicate that the network centrality coefficient estimate is not statistically significant (−0.01 with a standard error 0.02), indicating that the centrality is not associated with delinquency involvement. Moreover, Calvó-Armengol et al. (2009) also use the Add Health data to study (constant) peer effects on school performance index. They found statistically significant peer effects of smaller magnitude (0.55 with a standard error 0.12). Furthermore, Gaviria and Raphael (2001) and Kawaguchi (2004) use the National Education Longitudinal Study (NELS) dataset and the National Longitudinal Survey Youth 97 (NLSY97) dataset, respectively, to study a variety of youth behaviour ranging from drug consumption to church attendance. Their estimates of (constant) peer effects have smaller magnitude than ours. In Gaviria and Raphael (2001), for instance, the largest peer effect found is peer effects for drug use (0.25 with a standard error 7.17). In Kawaguchi (2004), the probability of a teenager using a substance (e.g. marijuana, alcohol or tobacco) increases from 1.4 to 2.6%, if the percentage of their peers’ usage increases by 10%.

\(^5\) To investigate computational performance, we also compute the NPLE and the MLE-type estimator of Xu (2017) in the CPE setting, which took 1.88 and 10,900.33 seconds, respectively. In particular, we choose the tuning parameter \(h = 4\) for the estimator of Xu (2017). Not surprisingly, both estimates are qualitatively similar.
ACKNOWLEDGEMENTS

A previous version of this paper was circulated under the title ‘Identification and estimation of hierarchy effects in social interactions’. We thank Jason Abrevaya, Anton Badev, Mark Hoekstra, Cheng Hsiao, Hae-shin Hwang, Qi Li, Jason Lindo, Xiaodong Liu, Essie Maasoumi, Steven Puller, Robin Sickles, Ken Wolpin, Mo Xiao, Keli Xu and participants at the Econometric Society Summer Meeting 2014, the Texas Econometrics Camp 2014, Emory, Monash, UT Austin and Texas A&M for helpful feedback. We also thank Vivian Goldman-Leffler and Stephanie White for proofreading this paper. Support from the Inter-university Consortium for Political and Social Research (ICPSR) and the Carolina Population Center of the National Longitudinal Study of Adolescent Health dataset are gratefully acknowledged.

REFERENCES

Aguirregabiria, V. and P. Mira (2007). Sequential estimation of dynamic discrete games. *Econometrica* 75, 1–53.

Allison, P. D. (1992). The cultural evolution of beneficent norms. *Social Forces* 71, 279–301.

Badev, A. (2013). Discrete games in endogenous networks: theory and policy. Technical Report, University of Pennsylvania.

Bajari, P., H. Hong, J. Krainer and D. Nekipelov (2012). Estimating static models of strategic interactions. *Journal of Business and Economic Statistics* 28, 469–82.

Bala, V. and S. Goyal (2000). A noncooperative model of network formation. *Econometrica* 68, 1181–229.

Bjorn, P. A and Q. H. Vuong (1984). Simultaneous equations models for dummy endogenous variables: a game theoretic formulation with an application to labor force participation. Working Paper 537, California Institute of Technology.

Blume, L. E., W. A. Brock, S. N. Durlauf and R. Jayaraman (2015). Linear social interactions models. *Journal of Political Economy* 123, 444–96.

Bonacich, P. (1987). Power and centrality: a family of measures. *American Journal of Sociology* 92, 1170–182.

Borgatti, S. P. and M. G. Everett (2006). A graph-theoretic perspective on centrality. *Social Networks* 28, 466–84.

Brock, W. A. and S. N. Durlauf (2001). Discrete choice with social interactions. *Review of Economic Studies* 68, 235–60.

Calvó-Armengol, A., E. Patacchini and Y. Zenou (2009). Peer effects and social networks in education. *Review of Economic Studies* 76, 1239–267.

Christakis, N. A., J. H. Fowler, G. W. Imbens and K. Kalyanaraman (2010). An empirical model for strategic network formation. Working Paper 16039, National Bureau of Economic Research.

Cialdini, R. B. and M. R. Trost (1998). Social influence: social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske and G. Lindzey (Eds.), *Handbook of Social Psychology, Volumes 1–2*, 151–92. New York, NY: Oxford University Press.

de Paula, A. (2013). Econometric analysis of games with multiple equilibria. *Annual Review of Economics* 5, 107–31.

de Paula, A. (2017). Econometrics of network models. In B. Honoré, A. Pakes, M. Piazzesi and L. Samuelson (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Eleventh World Congress*. Cambridge: Cambridge University Press.

© 2017 Royal Economic Society.
de Paula, A., S. Richards-Shubik and E. Tamer (2017). Identifying preferences in networks with bounded degree. Forthcoming in *Econometrica*.

Easley, D. and J. Kleinberg (2010). *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge: Cambridge University Press.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks* 1, 215–39.

Friedkin, N. E. and E. C. Johnsen (1990). Social influence and opinions. *Journal of Mathematical Sociology* 15, 193–206.

Gaviria, A. and S. Raphael (2001). School-based peer effects and juvenile behavior. *Review of Economics and Statistics* 83, 257–68.

Haynie, D. L. (2001). Delinquent peers revisited: does network structure matter? *American Journal of Sociology* 106, 1013–57.

Horst, U. and J. A. Scheinkman (2006). Equilibria in systems of social interactions. *Journal of Economic Theory* 130, 44–77.

Hurwicz, L. (1950). Generalization of the concept of identification. *Statistical Inference in Dynamic Economic Models* 10, 245–57.

Jackson, M. O. (2010). *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

Jackson, M. O. and A. Wolinsky (1996). A strategic model of social and economic networks. *Journal of Economic Theory* 71, 44–74.

Jonas, K. (1992). Modelling and suicide: a test of the werther effect. *British Journal of Social Psychology* 31, 295–306.

Kasahara, H. and K. Shimotsu (2012). Sequential estimation of structural models with a fixed point constraint. *Econometrica* 80, 2303–19.

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika* 18, 39–43.

Kawaguchi, D. (2004). Peer effects on substance use among american teenagers. *Journal of Population Economics* 17, 351–67.

Knoke, D. and R. S. Burt (1983). Prominence. In R. S. Burt and M. J. Minor (Eds.), *Applied Network Analysis*, 195–222. New York, NY: Sage.

Liu, X. and L-f. Lee (2010). GMM estimation of social interaction models with centrality. *Journal of Econometrics* 159, 99–115.

Mele, A. (2010). A structural model of segregation in social networks. CWP 32/10, Centre for Microdata Methods and Practice, Institute for Fiscal Studies and University College London.

Menzel, K. (2016). Strategic network formation with many agents. Working Paper CRATE-2016-1, Center for Research in Applied and Theoretical Econometrics (CRATE), New York University.

Nakajima, R. (2007). Measuring peer effects on youth smoking behaviour. *Review of Economic Studies* 74, 897–935.

Pesendorfer, M. and P. Schmidt-Dengler (2010). Sequential estimation of dynamic discrete games: a comment. *Econometrica* 80, 833–42.

Shi, Z. and A. B. Whinston (2013). Network structure and observational learning: evidence from a location-based social network. *Journal of Management Information Systems* 30, 185–212.

Stack, S. (1990). A reanalysis of the impact of non celebrity suicides. *Social Psychiatry and Psychiatric Epidemiology* 25, 269–73.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies* 70, 147–65.

Tamer, E. (2010). Partial identification in econometrics. *Annual Review of Economics* 2, 167–95.

Van der Vaart, A. W. (2000). *Asymptotic Statistics, Volume 3*. Cambridge: Cambridge University Press.

Xu, H. (2017). Social interactions: a game theoretic approach. Forthcoming in *International Economic Review*.

© 2017 Royal Economic Society.
APPENDIX: PROOFS OF RESULTS

Proof of Theorem 2.1: We show the result by contradiction. Let $\Sigma^*(\mathbb{W}) = (\sigma^*_1(\mathbb{W}), \ldots, \sigma^*_n(\mathbb{W}))'$ and $\Sigma^1(\mathbb{W}) = (\sigma^1_1(\mathbb{W}), \ldots, \sigma^1_n(\mathbb{W}))'$ be two different profiles of equilibrium choice probabilities. Suppose $\Sigma^*(\mathbb{W}) \neq \Sigma^1(\mathbb{W})$. For $\Sigma \in [0, 1]^n$, let

$$
\Gamma_i(\Sigma, \mathbb{W}) \equiv \frac{\exp(\beta(X_i) + \sum_{j \in F_i} \tilde{a}(S_j - S_i) \times \sigma_j(\mathbb{W}))}{1 + \exp(\beta(X_i) + \sum_{j \in F_i} \tilde{a}(S_j - S_i) \times \sigma_j(\mathbb{W}))}.
$$

By definition, $\Sigma^*(\mathbb{W})$ and $\Sigma^1(\mathbb{W})$ are two different solutions to the following equation:

$$
\sigma_i = \Gamma_i(\Sigma, \mathbb{W}), \quad \forall \ i = 1, \ldots, n.
$$

For player $i$, note that

$$
\sigma^*_i(\mathbb{W}) - \sigma^1_i(\mathbb{W}) = \sum_{j \in F_i} \frac{\partial \Gamma_i(\Sigma, \mathbb{W})}{\partial \sigma_j} \times (\sigma^*_j(\mathbb{W}) - \sigma^1_j(\mathbb{W}))
$$

$$
= \sum_{j \in F_i} \Gamma_i(\Sigma, \mathbb{W}) \times (1 - \Gamma_i(\Sigma, \mathbb{W})) \times \tilde{a}(S_j - S_i) \times (\sigma^*_j(\mathbb{W}) - \sigma^1_j(\mathbb{W}))
$$

where $\tilde{a}$ is a choice probability between $\Sigma^*(\mathbb{W})$ and $\Sigma^1(\mathbb{W})$. Because $\Gamma_i(\mathbb{W}) \in (0, 1)$, then $\Gamma_i \times (1 - \Gamma_i) \leq 1/4$. Therefore,

$$
|\sigma^*_i(\mathbb{W}) - \sigma^1_i(\mathbb{W})| \leq \frac{1}{4} \sum_{j \in F_i} |\tilde{a}(S_j - S_i) \times (\sigma^*_j(\mathbb{W}) - \sigma^1_j(\mathbb{W}))|
$$

$$
\leq \frac{1}{4} \max_{j \in F_i} |\sigma^*_j(\mathbb{W}) - \sigma^1_j(\mathbb{W})| \times \sup_{s \in S_0} |\alpha(s)|.
$$

By Assumption 2.2, we have

$$
|\sigma^*_i(\mathbb{W}) - \sigma^1_i(\mathbb{W})| < \max_{j \in F_i} |\sigma^*_j(\mathbb{W}) - \sigma^1_j(\mathbb{W})|.
$$

Therefore,

$$
\max_{i, i = 1, \ldots, n} |\sigma^*_i(\mathbb{W}) - \sigma^1_i(\mathbb{W})| < \max_{j \in F_i} |\sigma^*_j(\mathbb{W}) - \sigma^1_j(\mathbb{W})|
$$

leads to a contradiction. \qed

Proof of Theorem 3.1: Let

$$
L_n(\theta, \Sigma) = \frac{1}{n} \sum_{i=1}^{n} E[Y_i \ln \Gamma_i(\Sigma, \theta; \mathbb{W}) + (1 - Y_i) \ln(1 - \Gamma_i(\Sigma, \theta; \mathbb{W}))],
$$

and $\Lambda_n = \{\theta \in \Theta : \theta = \arg \max_{c \in \Theta} L_n(c, \Sigma(\theta; \mathbb{W}))\}$.

Note that $L_n(\cdot, \Sigma(\cdot; \mathbb{W}))$ is a continuously differentiable function in $\theta \in \Theta$. Because

$$
E[Y_i \ln \Gamma_i(\Sigma, \theta; \mathbb{W})] = E[\sigma^*_i(\mathbb{W}) \ln \Gamma_i(\Sigma, \theta; \mathbb{W})],
$$

which is a continuously differentiable function of $\theta \in \Theta$ with bounded derivatives (uniformly over $n$), then $L_n(\cdot, \Sigma(\cdot; \mathbb{W}))$ uniformly converges to $L(\cdot, \Sigma(\cdot; \mathbb{W}))$ under Assumption 3.5. It follows that $\Lambda_n \rightarrow \{\theta_0\}$ as $n \rightarrow \infty$.

Moreover, following Xu (2017), we have

$$
\sup_{\theta \in \Theta} |\hat{L}_n(\theta, \Sigma) - L_n(\theta, \Sigma)| \xrightarrow{p} 0.
$$

© 2017 Royal Economic Society.
Furthermore, by the argument in Aguirregabiria and Mira (2007),

$$d_H(\hat{\theta}_{\text{NPLE}}, \Lambda_n) \xrightarrow{p} 0,$$

where $d_H$ is the Hausdorff distance measure. Therefore, $\hat{\theta}_{\text{NPLE}} \xrightarrow{p} \theta_0$. □

**Proof of Theorem 3.2:** From the first-order condition, we have

$$\frac{\partial \hat{L}_n(\hat{\theta}_{\text{NPLE}}, \Sigma(\hat{\theta}_{\text{NPLE}}; \mathbb{W}))}{\partial \theta} = 0.$$

By a Taylor expansion on the above equation around the true parameter $\theta_0$, we have

$$\begin{align*}
\frac{\partial \hat{L}_n(\theta, \Sigma)}{\partial \theta} & = \frac{1}{n} \sum_{i=1}^{n} Z_i(Y_i - \Gamma_i(\Sigma, \theta; \mathbb{W})), \\
\frac{\partial^2 \hat{L}_n(\theta, \Sigma)}{\partial \theta \partial \theta'} & = -\frac{1}{n} \sum_{i=1}^{n} Z_i Z_i' \sigma_i(1 - \sigma_i) \\
\frac{\partial^2 \hat{L}_n(\theta, \Sigma; \mathbb{W})}{\partial \theta \partial \Sigma} & = -\frac{1}{n} \sum_{i=1}^{n} \sum_{j \in F_i} Z_i Z_j' \sigma_j^*(1 - \sigma_j^*) \times \frac{\psi_0 + (\phi_1 + \psi_1) \bar{S}_{ij}}{Q_i}.
\end{align*}$$

This gives us

$$\begin{align*}
\left(\frac{\partial^2 \hat{L}_n(\theta, \Sigma; \mathbb{W})}{\partial \theta \partial \theta'} + \frac{\partial^2 \hat{L}_n(\theta, \Sigma; \mathbb{W})}{\partial \theta \partial \Sigma} \frac{\partial \Sigma(\theta; \mathbb{W})}{\partial \theta}\right) \times \sqrt{n}(\hat{\theta}_{\text{NPLE}} - \theta_0) \\
= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i' (Y_i - \sigma_i^*(\mathbb{W})) + o_p(1).
\end{align*}$$

Because $Y_i$ is conditionally independent (conditional on $W_n$), by a conditional central limit theorem – see, e.g. Van der Vaart (2000) – and Assumption 3.7, we have

$$\sqrt{n}(\hat{\theta}_{\text{NPLE}} - \theta_0) \xrightarrow{d} N(0, \Omega(\theta_0)),$$

where $\Omega(\theta_0)$ is given by Theorem 3.2. □

**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article at the publisher’s web site:

Online Appendix
Replication files

© 2017 Royal Economic Society.