I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14167

Selection and Endogenous Treatment Models with Social Interactions: An Application to the Impact of Exercise on Self-Esteem

Zhongjian Lin
Francis Vella

MARCH 2021

DISCUSSION PAPER SERIES

IZA DP No. 14167

Selection and Endogenous Treatment Models with Social Interactions: An Application to the Impact of Exercise on Self-Esteem

Zhongjian Lin
Emory University

Francis Vella
Georgetown University and IZA

MARCH 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Selection and Endogenous Treatment Models with Social Interactions: An Application to the Impact of Exercise on Self-Esteem*

We address the estimation of sample selection and endogenous treatment models with social interactions. To model the interaction between individuals in an internally consistent matter we employ a game theoretic approach based on the use of a discrete Bayesian game. We overcome the substantial computational burden this introduces through a sequential version of the nested fixed point algorithm. We describe how our methodology can be applied to a large class of commonly employed models. We employ our approach to examine the impact of an individual's frequency of exercise on her level of self esteem in a setting where an individual's exercise frequency is treated as endogenous and is potentially influenced by her belief of her friends' exercise frequency.

JEL Classification: C31, C34, C57, Z2

Keywords: endogenous treatment, sample selection, social interactions, sequential estimation, exercise, self-esteem

Corresponding author:
Francis Vella
Economics Department
Georgetown University
37th and O Streets, NW
Washington, DC 20057
USA
E-mail: fgv@georgetown.edu

* We are grateful to Xiaodong Liu, Áureo de Paula, and John Rust for helpful comments and suggestions. All errors are our own.

1 Introduction

Several methodological papers by James Heckman (see, for example Heckman, 1974, 1978, 1979) on the estimation of sample selection and endogenous treatment models have been greatly influential in both the theoretical and empirical microeconometrics literatures. The former includes extensions of the original estimators via the relaxation of distributional and parametric assumptions, incorporating alternative selection/treatment rules, employing various identification schemes, the capacity to estimate models with categorical outcome dependent variables, and the adaption of the estimators to a large range of data structures (see, for example, Manski, 1990, Das, Newey, and Vella, 2003, Honoré and Hu, 2020). A feature of these methodological extensions is their focus on statistical rather than economic considerations. One economic aspect which has been largely ignored but merits attention is the inherent equilibrium and peer effects, capturing the interaction between the individuals whose behavior is being modelled, operating in the selection or treatment decision. (see, for example, Manski, 1993, 2000, Brock and Durlauf, 2001a,b, Heckman and Vytlacil, 2007).[1]

Sample selection and endogenous treatment models feature an equation characterizing the individual's treatment decision and equation(s) describing the outcome for the respective treated and non treated groups. In selection models the outcomes are generally only observed for the treated.[2] The presence of selection and endogeneity reflects that the unobservables in the treatment equation are not independent of those in the outcome equation(s) and one cannot consistently estimate the outcome equation without an explicit consideration of the treatment decision. While the model can be estimated jointly by maximum likelihood while incorporating the correlation across equations, the more frequently employed approach in empirical work is a two step procedure. The first step estimates the treatment equation from which the appropriate control function(s) are constructed. The second step estimates the outcome equation(s) with the inclusion of the control function as a conditioning variable.

An underlying assumption of these models is that the individual's treatment decision does not explicitly incorporate his/her expectation of the other agents' choices. A well known example is the Willis and Rosen (1979) investigation of

[1] We use the terms "peer effects" and "social interactions" interchangeably.

[2] To simplify terminology across the range of models we consider we refer to the equation describing whether the individual decides to participate or be treated as the treatment equation.

whether individuals optimally invest in college education. Their approach requires the estimation of separate wage equations for those attending and not attending college while accounting for the selection bias arising from the non college/college choice. This requires the first step estimation of the college decision. However, Willis and Rosen impose that the individual's expectation of the other individuals' behavior is irrelevant. This might be problematic if a determinant of an individual's college decision is their belief of their friends' choices. Moreover, the friends' choices are likely to be influenced by their beliefs of their friends' choices. It seems necessary to account for this interaction to appropriately control for the correlation between the outcome and treatment equations.

One challenge in including "social interactions" into the Heckman methodology is in introducing the interdependence between the agents' expected behavior into the treatment decision. This reflects the need to model the interaction between individuals in an internally consistent matter. We incorporate a game-theoretic approach into these models to capture the simultaneity of the individuals' choices when there is a large number of individuals in a large social network. The associated computational burden is addressed through a Bayesian game which solves for the simultaneity of peers' choices. This is done by projecting individual's choices onto a potentially large exogenous space containing the relevant individual's characteristics, the individual's friends' characteristics, the characteristics of the friends of friends, and so on. We propose estimation via a sequential version of the fixed point algorithm (see Rust, 1987, Aguirregabiria and Mira, 2007, Lin and Xu, 2017).

We consider two types of estimators which capture agent interaction in the treatment decision. The first adapts the Heckman (1974) approach and constructs the appropriate likelihood function to jointly model the treatment and outcome equations and the correlation in the unobservables across the two equations. The second corresponds to the Heckman (1979) two step estimator which accounts for the inherent endogeneity via the appropriate control function from the treatment equation. The paper contributes to two literatures. First, we use a game theoretic approach to model how an individual's choice is affected by his/her belief of his/her friends' choices in a large class of models. While some of the issues are addressed in Lin and Xu (2017), Xu (2018), Hu and Lin (2020), this is the first paper to employ them in the endogenous treatment and sample selection context. Second, we implement our estimator to examine how an individual's level of weekly exercise influences their level of self esteem. This is an important research area given the relationship between an individual's self esteem and their mental, physical and economic well being. Our empirical investigation establishes that exercise has a positive impact on self esteem and accounting for its simultaneity increases this estimated impact. Moreover, we find that an individual's belief of the exercise frequency of their peers has a statistically significant impact on their own.

While we cover a large class of models there are four important issues we do not address. First is the nature of the information available to the agents in forming their beliefs. By assuming the relevant unobserved information is private, we construct expectations such that they are uncorrelated with the other unobservables in the model. However, in many settings there may also be publicly shared unobserved information. This introduces an additional source of endogeneity which we do not account for. A second restriction is the large social network which we assume to be exogenous. This is a strong assumption although it is frequently employed in the social interactions literature (see, for example, Manski, 1993, Brock and Durlauf, 2001a, Lee, 2007, Bramoullé, Djebbari, and Fortin, 2009, Calvó-Armengol, Patacchini, and Zenou, 2009, Lee, Liu, and Lin, 2010, Goldsmith-Pinkham and Imbens, 2013, Lee, Li, and Lin, 2014). The third issue is the parametric nature of approach. We noted above that many of the theoretical advances in the sample selection and endogenous treatment literatures are related to the relaxation of parametric assumptions. While many of these advances are likely to be applicable we do not pursue them. We introduce and apply our methodology in a setting which is commonly employed and and well understood. We delay the extensions to more general structures, in addition to the issues regarding the exogeneity of the network and the source of information, to future work. Finally, in the models we consider we only allow for interactions in the treatment decision. That is, we do not extend our analysis to incorporate social interactions in the outcome equation. For many of the models we consider the more difficult challenge is to include them in the treatment choice but we acknowledge there are likely to be many empirical settings in which they should be included in both the treatment and outcome equations. We also delay this issue to future research.

The following section outlines the large class of models we cover. Section 3 describes the explicit manner in which social interactions enter the model and discusses the employed Bayesian Nash equilibrium concept. Estimation issues are addressed in Section 4. The empirical investigation of the impact of an individual's frequency of physical exercise on their level of self esteem is examined in Section 5. Section 6 concludes. Proofs and simulation evidence are provided in the Appendix.

2 Model

The models we consider fall into two broad categories. The first correspond to the endogenous treatment model and have the form:

$$\begin{aligned} Y_i^* &= X_i'\beta_O + \gamma D_i + u_i \\ D_i &= 1\{Z_i'\beta_D + \alpha V_i + v_i > 0\} \\ Y_i &= h(Y_i^*) \end{aligned}$$

where Y_i^* is a latent outcome of a variable of interest with corresponding observed counterpart Y_i obtained via the censoring function $h(.)$; $1\{\cdot\}$ is the indicator function and D is a binary outcome; X and Z are vectors of exogenous variables; the β, γ and α are parameters; u and v are potentially correlated error terms and V_i is a variable capturing individual $i's$ belief regarding the treatment choice of $i's$ friends. We assume there are n individuals and each has N_i direct friends which also appear in the sample. The construction of V_i is described below but note it is a known function of the expectation of the $D_j's$ where the individuals denoted with the subscript j are members of $i's$ network. The second is the sample selection model:

$$\begin{aligned} Y_i^* &= X_i'\beta_O + u_i \\ D_i &= 1\{Z_i'\beta_D + \alpha V_i + v_i > 0\} \\ Y_i &= h(Y_i^*) * D_i. \end{aligned}$$

The endogenous treatment model focuses on the estimation of the treatment parameter γ. The sample selection focuses on the estimation of β_O. The econometric challenge is accounting for the correlation between u and v while constructing V_i. We delay the explicit discussion of this issue to the following section and first highlight some popular special cases of the two general models above.

1. **Case 1:** The generalized Roy model with peer effects in treatment participation: Y_i is observed:

$$\begin{aligned} Y_i &= \begin{cases} Y_{0i}^* = X_{0i}'\beta_0 + u_{0i}, & \text{if } D_i = 0 \\ Y_{1i}^* = X_{1i}'\beta_1 + u_{1i}, & \text{if } D_i = 1 \end{cases} \\ D &= 1\{Z_i'\beta_D + \alpha V_i + v_i > 0\} \end{aligned}$$

2. **Case 2:** Endogenous treatment model: $Y_i = Y_i^*$ and outcome variable observed

for whole sample:

$$D_i = 1\{Z_i'\beta_D + \alpha V_i + v_i > 0\}$$
$$Y_i^* = X_i'\beta + \gamma D_i + u_i$$

3. **Case 3:** Multivariate binary choice with peer effects:[3] There are two binary potential outcome equations:

$$D_i = 1\{Z_i'\beta_D + \alpha V_i + v_i > 0\}$$
$$Y_i = \begin{cases} 1\{Y_{0i}^* > 0\} = 1\{X_{0i}'\beta_0 + u_{0i} > 0\}, & \text{if } D_i = 0 \\ 1\{Y_{1i}^* > 0\} = 1\{X_{1i}'\beta_1 + u_{1i} > 0\}, & \text{if } D_i = 1 \end{cases}$$

4. **Case 4:** Bivariate binary choice with peer effects: $Y_i = 1\{Y_i^* > 0\}$ is observed and covariate effects of non-intercept observable characteristics are the same for the two potential outcome equations:

$$D_i = 1\{Z_i'\beta_D + \alpha V_i + v_i > 0\}$$
$$Y_i = 1\{X_i'\beta + \gamma D_i + u_i > 0\}$$

5. **Case 5:** Sample selection model: $Y_i = Y_i^*$ and outcome variable observed for observations for which $D_i = 1$:

$$\begin{aligned} Y_i^* &= X_i'\beta_1 + u_i \\ D_i &= 1\{Z_i'\beta_D + \alpha V_i + v_i > 0\} \\ Y_i &= Y_i^* * D_i \end{aligned}$$

This is a subset of the models we cover noting that in some instances we require specific normalizations on the error variances for identification. An important implication of the assumptions that follow is that even in the presence of selection and/or treatment endogeneity the V_i can be treated as exogenous.

3 Incorporating Social Interactions in the Treatment Equation

To construct the belief variable V_i we adopt a game theoretic approach. Let $F_{ij} = 1$ denote that individual i considers j a friend. We set $F_{ii} = 0$ by convention and denote $F_i = \{j \in I : F_{ij} = 1\}$ as the set of $i's$ peers. The number of $i's$ friends is then $N_i = \#(F_i)$. We assume the u_i and v_i each represent private information only known to agent i. The

[3]Carrasco (2001) refers to this as the switching probit model.

remaining sources of information (denoted as $\mathbb{I} = \{X_i, Z_i, F_i\}_{i \in I}$) are treated as public. This private information assumption is important not only for the construction of the belief but also has implications for the nature of the endogeneity of the treatment decision.

We consider an incomplete information structure and a Bayesian game for social interactions in treatment choices. We follow Manski (2000) to conceptualize individuals as decision makers endowed with preferences, who form expectations and face constraints. Individuals make simultaneous treatment choices. With incomplete information, individuals do not observe the actions of their peers but form a belief of these choices (see, for example, Brock and Durlauf, 2001a, Lin and Xu, 2017, Xu, 2018, Hu and Lin, 2020). The belief, $E(D_j|\mathbb{I})$, rather than the observed choices, D_j, affect $i's$ decision.[4] We then construct the social interactions term, discussed above, as $V_i = \frac{1}{N_i} \sum_{i \in F_i} E(D_i|\mathbb{I})$. This represents a form of summary statistic regarding the belief of $i's$ peers which ignores the members' identities. An advantage of this approach is that it allows a tractable equilibrium characterization of the simultaneous treatment choices, while the Bayesian Nash equilibrium (contraction fixed point) facilitates the estimation of the high dimensional exogenous space.

The use of $E(D_j|\mathbb{I})$, rather than the observed choices D_j, in the treatment equation reflects that individuals are simultaneously making decisions and do not see the choices of the others in the network. The private information assumption implies that the endogeneity of the peers choices operate through these expectations. This excludes the influence of unobserved factors which influence both the individual's and their peers' choices. Including the observed choices in the treatment equation implies that there are unobserved factors which influence both the choices of i in addition to those of the other network members. This would result in the choices being endogenous even after the inclusion of V_i. We do not address this issue. Similar approaches and results can be found in Lin and Xu (2017), Xu (2018), Hu and Lin (2020) for treatments of large network games. There is a large litereature in theoretical microeconomics which studies this type of incomplete information game (see, for example, Morris and Shin, 2003, Bergemann and Morris, 2013) and it is also commonly used in the theoretical econometric literature whose focus is discrete games with incomplete information structures (see, for example, Aradillas-López, 2010, 2012, 2020, De Paula and Tang, 2012, Wan and Xu, 2014, Lewbel and Tang, 2015).

[4]The choice variables of peers are endogenous due to the simultaneity in the Bayesian game. The replacement of D_j by $E(D_j|\mathbb{I})$ mimics the IV approach of projecting the endogenous variable D_j onto the exogenous space, $\mathbb{I}$.

3.1 Bayesian Nash Equilibrium Characterization

To proceed with estimation we describe how V_i is generated in an equilibrium setting. The following assumptions characterize the employed Bayesian Nash Equilibrium (BNE).

Assumption 1. *The error terms u_i and v_i are independent across individuals and v_i follows a standard normal distribution.*

Combining Assumption 1 and the treatment rule equation gives the conditional choice probability:

$$P(D_i = 1|\mathbb{I}) = \Phi(Z_i'\beta_D + \alpha V_i), .i \in \mathcal{I}.$$

V_i contains the belief of peers' treatment decisions and this conditional choice probability characterizes the Bayesian equilibrium in n simultaneous equations of $P^* = (P(D_1 = 1|\mathbb{I}), \cdots, P(D_n = 1|\mathbb{I}))$.

Multiplicity of the game hampers the coherency of econometric models (see Tamer, 2003). When the number of players is small the total number of equilibria is also small and a partial identification approach enables interval estimation (see, for example, Tamer, 2003, Chernozhukov, Hong, and Tamer, 2007, Ciliberto and Tamer, 2009, Tamer, 2010, Balat and Han, 2020, Ciliberto, Murry, and Tamer, 2020). When the number of players is an increasing n, the number of equilibria goes exponentially to infinity and partial identification is not sufficient or possible for inference. In a large network, data from one equilibrium is observed even though there are multiple equilibria.

We make the following assumption to establish the uniqueness of the Bayesian game.

Assumption 2. *(i) There exists $M > 0$ such that $0 < \max N_i < M$. (ii) The social influence is moderate; i.e., $0 \leq \alpha < \sqrt{2\pi}$.*

Assumption 2 is a sufficient condition to establish the contraction mapping condition of the BNE and its uniqueness. The upper bound corresponds to the standard normal distribution. With general CDF and PDF, $F(\cdot)$ and $f(\cdot)$, the upper bound is $\frac{1}{\sup_c f(c)}$. In the large single network, combining Assumptions 1 and 2 implies another important feature. Namely, the weak dependence of the data which is required for feasible inference. For more details on the moderate social influence (MSI) and the uniqueness of the equilibrium, see Brock and Durlauf (2001a), Glaeser and Scheinkman

(2003), Horst and Scheinkman (2006), Lee, Li, and Lin (2014), Lin and Xu (2017), Xu (2018), Hu and Lin (2020), Jackson, Lin, and Yu (2020), Lin (2020).

Denote $P \equiv (P_1, \cdots, P_n)'$ as an arbitrary choice probabilities profile and define $V_i(P) \equiv \frac{1}{N_i} \sum_{j \in F_i} P_j$. Define the best response function

$$\Gamma_i(P) \equiv \Phi\Big(Z_i'\beta_D + \alpha V_i(P)\Big). \tag{1}$$

Let $\Gamma \equiv (\Gamma_1, \cdots, \Gamma_n)'$.

Proposition 1. *When Assumptions 1 and 2 hold,* Γ *is a contraction mapping in* $[0,1]^n$ *with respect to the metric* $\Delta(P, P') \equiv \max_{i \in \mathcal{I}} |P_i - P_i'|$ *and there exists a unique BNE in the Bayesian game.*

Proof. See Appendix A. □

The MSI condition further establishes the weak dependence of the conditional choice probabilities in the social network. The requirement of weak dependence is to facilitate inference. Denote N_i^h as the subnetwork centering at i and all members are within social distance of h from i. We allow h to grow to infinity, although at a slower rate than n. Let $\mathbb{I}_i^h$ denote the public information of the subnetwork N_i^h.

Proposition 2. *When Assumptions 1 and 2 hold, we have the following approximation:*

$$|P(D_i = 1|\mathbb{I}) - P(D_i = 1|\mathbb{I}_i^h)| \to 0$$

as $n \to \infty$.

Proof. See Appendix A. □

We focus on cases 1 and 4 above for illustrative purposes and note that the results for cases 2 and 3 are straightforward extensions. When the Bayesian game has a unique equilibrium the conditional choice probabilities, i.e., $P(D_i = 1|\mathbb{I})$ are identifiable from the data. Thus V_i can be treated as observable and the identification results of Case 1-5 follow directly from various treatments of these models (see, for example, Eisenhauer, Heckman, and Vytlacil, 2015, Heckman, 1974, 1978, 1979, Heckman and Vytlacil, 1999, 2001, 2005, 2007, Lewbel, 2007). We make the following assumptions for identification of structural parameters and treatment effects parameters.

Assumption 3. (u_i, v_i) *are independent of* (X_i, Z_i).

Assumption 4. *The population mean* $E|Y|$ *is finite.*

4 Estimation Strategy

We now provide estimation methods for a range of models including the five cases described above. Although we propose one strategy which involves joint estimation of the treatment and outcome equations by MLE and another which involves the sequential estimation of the two equations, the complication for both is the presence of V_i in the treatment equation. Thus for joint MLE estimation the likelihood function $L_n(\theta; P)$ is for both the treatment and outcome equations while for the sequential estimator the likelihood function is for the treatment equation. Accordingly we denote the parameter $\theta \equiv (\theta_S, \theta_O)$ where θ_S is the treatment equation parameter and θ_O is the outcome equation parameter. To characterize the correlation between u and v we assume:

Assumption 5. *The error terms u and v are normally distributed with variances σ_u^2, $\sigma_v^2 = 1$ and covariance σ_{uv}.*

Write the pseudo likelihood as $L_n(\theta; P) \equiv \frac{1}{n}\sum_{i=1}^{n} l_i(\theta, P)$ where P denotes an arbitrary choice of the probabilities profile. The associated pseudo social interaction term is defined as $V_i(P) = \frac{1}{N_i}\sum_{j \in F_i} P_j$. Before we introduce the nested pseudo joint likelihood method consider the challenges in estimating the model by MLE. The estimator is given as:

$$\hat{\theta}_{MLE} = \arg\max_{\theta \in \Theta} L_n(\theta; P) \text{ s.t. } P = \Gamma(\theta, P). \tag{2}$$

To implement MLE one can employ the nested fixed point algorithm of Rust (1987) which requires; i) repeated solution of the model for each trial value of the parameters (for example, for $\tilde{\theta}$ in the Newton-Raphson method); ii) solution of fixed points $\tilde{P}(\tilde{\theta})$ from the equilibrium, and iii) computation of the likelihood value. The cost of estimating the model via this algorithm is high or possibly infeasible due to the large dimension of the equilibrium characterization. Even in the single agent dynamic discrete choice model (see Aguirregabiria and Mira, 2002) and the dynamic discrete game with a fixed number of players (see Aguirregabiria and Mira, 2007), there is a high computational burden of solving fixed points in the nested fixed point algorithm. It is more computationally intensive here as the dimension of the Bayesian game is the network size, n, and this is allowed to go to infinity in our asymptotic analysis. To overcome these issues we define a nested pseudo joint likelihood (NPJL) estimator which only solves the model once, with the cost of several outer iterations:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L_n(\theta; P) \text{ and } P = \Gamma(\theta_S, P). \tag{3}$$

Implementation of the NPJL method is as follows:

1. **Initiation:** Conjecture the conditional choice probabilities in the Bayesian Nash equilibrium and denote these $P^{(0)}$.[5] Construct $V_i^{(0)}(P^{(0)}) = \frac{1}{N_i} \sum_{j \in F_i} P_j^{(0)}$.
2. **Iteration:** Given $V^{(K)}(P^{(K)})$, obtain estimate $\hat{\theta}^{(K+1)} = \left(\hat{\theta}_S^{(K+1)}, \hat{\theta}_O^{(K+1)}\right)$ via MLE using the likelihood $L_n(\theta, P^{(K)})$. Update the choice probabilities in each iteration as:
$$P^{(K+1)} = \Gamma\left(\hat{\theta}_S^{(K+1)}, P^{(K)}\right).$$
3. **Convergence:** Iterate until the specified convergent criterion based on $||P^{(K)} - P^{(K+1)}||$ is satisfied. Denote the K-th estimate as $\hat{\theta}$ and $\hat{P}$.

We now use a subscript in the choice probabilities profile to denote its dimension, i.e., $P_{[n]}$. Denote θ_0 as the true parameter. Let:

$$\begin{aligned}
L_0(\theta, P_{[n]}) &\equiv \mathbb{E}\left[l_i(\theta, P_{[n]})\right], \\
\tilde{\theta}_0(P_{[n]}) &\equiv \arg\max_{\theta \in \Theta} \; L_0(\theta, P_{[n]}), \\
\psi_0(P_{[n]}) &\equiv \Gamma(\tilde{\theta}_0(P_{[n]}), P_{[n]}), \\
\tilde{\theta}_n(P_{[n]}) &\equiv \arg\max_{\theta \in \Theta} \; L_n(\theta, P_{[n]}), \\
\psi_n(P_{[n]}) &\equiv \Gamma(\tilde{\theta}_n(P_{[n]}), P_{[n]}).
\end{aligned}$$

Define the population NPJL fixed points set as $\Lambda_{0n} \equiv \{(\theta, P_{[n]}) \in (\Theta, \mathcal{P}_n) : \theta = \tilde{\theta}(P_{[n]}), P_{[n]} = \psi_0(P_{[n]})\}$ and the NPJL fixed points set of sample size n as $\Lambda_n \equiv \{(\theta, P_{[n]}) \in (\Theta, \mathcal{P}_{[n]}) : \theta = \tilde{\theta}_n(P_{[n]}), P_{[n]} = \psi_n(P_{[n]})\}$. Let $\mathcal{N}$ denote a closed neighborhood of $(\theta_0, P^*_{[n]})$. Let $\widetilde{Z}_i \equiv (Z_i', V_i)$.

To establish the $\sqrt{n}$ consistency and asymptotic normality of $\hat{\theta}$ in the treatment equation we impose the following rank and regularity conditions.

Assumption 6. *(i)$E(\widetilde{Z}_i\widetilde{Z}_i')$ has full rank and $E(X_iX_i')$ has full rank.*

Assumption 7. *(i) Θ is compact, θ_0 is an interior point of Θ, and $\mathcal{P}_n$ is a compact and convex subset of $(0,1)^n$; (ii) $(\theta_0, P^*_{[n]})$ is an isolated population NPJL fixed point; i.e., it is either unique*

[5] These can be obtained, for example, via probit estimation for the selection equation ignoring the social interactions term.

or there is an open ball around it that does not contain any other element of Λ_0; (iii) $\frac{\partial^2 \mathcal{L}_0(\theta, P^*_{[n]})}{\partial\theta\partial\theta'}$ is a nonsingular matrix; (iv) the operator $\psi_0(P_{[n]}) - P_{[n]}$ has a nonsingular Jacobian matrix at $P^*_{[n]}$; (v) $I - \left(\frac{\partial\Gamma(\theta_0;P^*)}{\partial P}\right)'$ is invertibale for n sufficiently large; (vi) there exist non-singular matrices $V_1(\theta_0)$ and $V_2(\theta_0)$ such that:

$$\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta_0, P^*)}{\partial\theta\partial\theta'} + \frac{\partial^2 \mathcal{L}(\theta_0, P^*)}{\partial\theta\partial P'} \cdot \left[I - \left(\frac{\partial\Gamma(\theta_0; P^*)}{\partial P}\right)'\right]^{-1} \cdot \frac{\partial\Gamma(\theta_0; P^*)}{\partial\theta'}\right] \to \Omega_1(\theta_0),$$

$$\mathbb{E}\left[n \frac{\partial \mathcal{L}(\theta_0, P^*)}{\partial\theta} \frac{\partial \mathcal{L}(\theta_0, P^*)}{\partial\theta'}\right] \to \Omega_2(\theta_0).$$

Moreover, $\Omega_1(\theta_0)$ is negative definite.

Theorem 1. *When Assumptions 1 to 6 and Assumption 7(i)-(iv) hold, $\hat{\theta} \xrightarrow{p} \theta_0$. When additionally, Assumption 7 (v) also holds, we have:*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, \Omega_1(\theta_0)\Omega_2^{-1}(\theta_0)\Omega_1'(\theta_0)\right).$$

The proof is similar to that of Hu and Lin (2020) and omitted here.

Remark 1. *For cases 1, 2 and 5 we can also implement the estimation by the two-step method. For brevity we illustrate the Nested pseudo two-step method for case 2.*

1. ***Initiation:*** *Conjecture the conditional choice probabilities $P^{(0)}$ in the Bayesian Nash equilibrium and construct $V^{(0)}(P^{(0)}) = \frac{1}{N_i}\sum_{j\in F_i} P_j^{(0)}$..*

2. ***Iteration:*** *Probit MLE of D_i on Z_i and $V_i^{(K)}(P^{(K)})$ to get $\hat{\beta}_D^{(K+1)}$ and $\hat{\alpha}^{(K+1)}$. Update the choice probabilities by:*

$$P^{(K+1)} = \Gamma\left(Z_i'\hat{\beta}_D^{(K+1)} + \hat{\alpha}^{(K+1)} V_i(P^{(0)})\right),$$

3. ***Convergence:*** *Iterate until the specified convergent criterion based on $||P^{(K)} - P^{(K+1)}||$ is satisfied. Denote the K-th estimate as $\hat{\beta}_D$, $\hat{\alpha}$ and $\hat{P}$. Construct the generalized residual:*

$$GR_i = D_i \times \left[\frac{\phi\left[Z_i'\hat{\beta}_D + \hat{\alpha}V_i(\hat{P})\right]}{\Phi\left[Z_i'\hat{\beta}_D + \hat{\alpha}V_i(\hat{P})\right]}\right] + (1 - D_i) \times \left[\frac{-\phi\left[Z_i'\hat{\beta}_D + \hat{\alpha}V_i(\hat{P})\right]}{1 - \Phi\left[Z_i'\hat{\beta}_D + \hat{\alpha}V_i(\hat{P})\right]}\right]$$

4. ***Outcome Equation:*** *Least squares regression of Y on X, D and GR to get estimates $\hat{\beta}_O$, $\hat{\gamma}$ and $\hat{\rho}$.*

Appendix C provides simulation evidence that the proposed estimators perform well.[6]

[6]The simulations also examine the capacity of the proposed procedures to identify the presence of

5 Exercise and Self-Esteem

We now investigate the impact of an individual's frequency of exercise on their level of self-esteem. Self-esteem is considered an important aspect of quality of life and mental well-being. Campbell (1984) refers to self-esteem as the "First Law of Human Nature" and improvement in self esteem is frequently the primary objective in studies of health intervention. While acknowledging that the process determining an individual's self esteem is extremely complicated, empirical evidence suggests that an individual's level of physical exercise is considered to be one of its more important determinants (see, for example, Sonstroem, 1984, Sonstroem and Morgan, 1989, Sonstroem, Harlow, and Josephs, 1994). This is based on evidence from existing studies utilizing randomized controlled trials and/or experiments (see, for example, Ekeland, Heian, and Hagen, 2005, Fox, 2000b, Tiggemann and Williamson, 2000). One proposed mechanism is that exercise affects an individual's sense of autonomy and personal control over one's physical appearance and functioning (Fox, 2000a). A substantial empirical literature is devoted to exploring the relationship (see, for example, Fox, 2000a, Spence, McGannon, and Poon, 2005). A challenge in estimating the impact of exercise is that it is likely to be endogenous to self-esteem. That is, the same unobservable factors are likely to determine both and it is possible that causation runs in both directions (see Furnham, Badmin, and Sneade, 2002, Strelan, Mehaffey, and Tiggemann, 2003).

To examine this relationship we estimate the following model:

$$\begin{aligned} Self\text{-}Esteem_i &= X_i'\beta_O + \gamma Exercise_i + u_i \\ Exercise_i &= 1\{X_i'\beta_S + \alpha V_i + v_i > 0\} \end{aligned}$$

where *Exercise* is measured by an indicator function denoting that the individual's level of weekly exercise is above some specified frequency threshold; $Self\text{-}Esteem$ is a measure of self reported self-esteem described below; the $X's$ denote a vector of exogenous explanatory variables and the V denotes the network effect. The primary parameter of interest is the treatment effect γ.

An important aspect of this relationship which has been ignored in this literature is the role of peer effects. An individual's exercise frequency is likely to be influenced by their expectation of that of their friends'. This may capture peer pressure, other forms of motivation, or simply the capacity to participate in exercise which requires multiple participants. The impact of these peer effects is captured by the parameter α. As the

selection and/or endogeneity and the bias which arises if the observed value of $D_j's$, are mistakenly employed.

level of *Exercise* reflects an individual choice it is likely that it jointly determined with *Self-Esteem*. Thus it is necessary to accommodate both this simultaneity and the joint determination of the network variable in estimating the treatment effect. Following a description of the data we estimate the model in two ways. As the *Exercise* variable is an endogenous binary treatment we first estimate a model where the self esteem variable is treated as a continuous outcome. We then consider a empirical model where *Self-Esteem* is treated as a binary outcome denoting the individual is above or below some level on the continuous measure. Although each of the models are reliant on the same distributional assumptions we estimate both to illustrate the applicability of our approach.

5.1 Data

We examine data from the Wave I of *National Longitudinal Study of Adolescent Health* (*Add Health)* dataset which was conducted in 1994-1995 and surveys students in grades 7-12 from sample of representative schools. The survey collects information on a number of student characteristics including academic performance, health conditions and socio-economic demographic variables. The survey asks each student to list as many as five best female and five best male friends. We construct the friendship network using these responses.

Our measure of self-esteem measure is constructed using six items in the *Add Health* dataset based on the Rosenberg (2015) self-esteem scale. These questions ask students to state their level of agreement with the following six statements regarding their self-perception or self-worth: i) "You have a lot of good qualities"; ii) "You have a lot to be proud of"; iii) "You like yourself just the way you are"; iv) "You feel like you are doing everything just about right"; v) "You feel socially accepted"; and vi) "You feel loved and wanted". The permitted responses are "strongly agree", "agree", "neither agree nor disagree", "disagree", and "strongly disagree" and each answer is coded 0- 4 with the higher score indicating greater self-esteem. This produces a self-esteem scale ranges from 0 to 24.

The exercise variable is constructed using responses to the question "How many times in a normal week to you work, play, or exercise hard enough to make you sweat and breathe heavily?". The possible responses are: i) "never"; ii) "1 or 2 times"; iii) "3 to 5 times"; iv) "6 or 7 times"; and v) "more than 7 times". We initially define the exercise variable as an indicator function taking the value 1 if the individual responds either "6 or 7 times" or "more than 7 times" and zero otherwise. We employ this definition to create a binary treatment variable.[7]

Table 1 provides the summary statistics of variables we employ in our empirical investigation. A small majority of the sample is female and over sixty percent are white. The self esteem variable has a mean of nearly 19 indicating an average response of "agree" to the various statements. The mean is reassuring and the first quartile of the distribution of responses is 17 indicating a large fraction in this sample reporting levels of self-esteem above the mean. However there are many individuals reporting a very low value of self esteem. The exercise variable has an average value of .4 indicating that 40 percent of the sample exercises at least 6 times a week. This may appear high but as the sample comprises adolescents at high school it is not surprising.

5.2 Main Results

We begin by regressing the continuous measure of $Self$-*Esteem* on the binary exercise variable and a number of background variables such as the individual's GPA, the individual's age and gender, family income, and a self reported intelligence measure. These results are reported in Table 2 and show that a number of the regressors have a statistically significant impact on the level of $Self$-*Esteem*. We delay a discussion of the impact of these other regressors until we determine the preferred specification. The estimate of the impact of exercise in this model is .733. The estimate is statistically significant but surprisingly small in magnitude. Given the construction of $Self$-*Esteem* this corresponds to less than the difference in the level of response to one sub-questions.

We now estimate the model accounting for the possible endogeneity and the presence of peer effects. We estimate the model by joint MLE and the two step approach which requires the construction and use of the control function. We include the same variables in the exercise equation noting the model is identified by the inclusion of the peer effects variable which is a function of the $X's$ of the other members in the network. We estimate the *Exercise* equation both separately and jointly with the $Self$-*Esteem* equation. We only discuss the results for the joint MLE estimation of the *Exercise* equation, in Table 5, noting that the estimates for the separate estimation, reported in Table 3, are almost identical.

A number of features of the estimates from the *Exercise* equation are of interest as several of the explanatory variables are statistically significant. Exercise decreases

[7] We acknowledge that this choice is arbitrary. As the variable is also employed in constructing the network variable it is also of interest to examine the impact of choosing other threshold values. This would then estimate the impact of different peer effects and alternative exercise treatments on self esteem. Although we do not report the results here an initial investigation revealed that at low levels of frequency there is neither a peer effect nor an exercise effect.

with age although there is not a great deal of variation in age across individuals in our sample. Females exercise less and students who respond that they are white exercise more. The estimated female effect seems large. Higher GPA and family incomes increase the probability that the individual exercises more than 5 times a week. The most interesting coefficient is that associated with the expected behavior of the individual's peers. As the coefficient value is .172, and mean and the standard deviation of this variable are 0.205 and 0.225 respectively, the impact on the probability of exercising more than 5 times a week is not small given that each individual nominates 10 friends. The marginal impact of going from the lowest to highest value of V is to increase the probability of exercise from .397 to .438. This is not surprising given the various different ways such an effect is likely to occur. Moreover, the coefficient is statistically different from zero at conventional testing levels. Note that the magnitude of the marginal effect is consistent with our assumption 2 regarding the level of social influence.

We now focus on the estimation of the $Self$-*Esteem* equation. Table 4 reports the results from two step estimation of the model in which the generalized residual from the *Exercise* equation is included as an additional regressor. A number of coefficients are of interest but we limit attention to those associated with the exercise variable and the generalized residual. The *Exercise* coefficient has now increased to 4.686 and continues to be statistically significant. This not only represents a large increase in comparison to the unadjusted OLS estimate but also represents a very large impact of frequency of exercise on an individual's self esteem. The cause of this large increase is the large coefficient on the generalized residual. This coefficient is also statistically significant providing clear evidence of endogeneity. The coefficient value of -2.472 indicates that the unobservables which increase the probability of exercising more than 5 times weekly are negatively correlated with one's level of self esteem. While the coefficient on the *Exercise* variable appears high it does seem consistent with previous findings (see, for example Furnham, Badmin, and Sneade, 2002). The negative correlation between the unobservables appears reasonable although the direction of this relationship appears to be an empirical matter given the variety of mechanisms and factors which may be generating it.

We also estimated the treatment and outcome equations jointly by MLE and the results for the outcome equation from this procedure are in Table 6. They are generally similar to the two step procedure although there are some notable exceptions.[8] The exercise coefficient is 2.357 and is statistically significant at the 5 percent level. Although this is half the value of that for the two step procedure the impact continues to be large and indicates that exercise has a substantial impact on an individual's level of self esteem. The statistically significant negative estimate of the correlation coefficient supports the presence of endogeneity and suggests that the unobservables are negatively correlated across equations. This is consistent with the two step estimates.

In addition to the magnitude of the *Exercise* coefficient there are other notable differences across the two sets of estimates. The most remarkable difference is the larger and statistically significant effect for the female variable. The estimates from the joint MLE procedure are generally more similar to the unadjusted OLS estimates although the exercise effect is much larger for the joint MLE estimates. While we do not test it formally it is possible that the two adjusted estimates are not statistically different. Although it is beyond the scope of this paper to investigate this issue, it is possible that the joint MLE procedure's heavier reliance on normality is responsible for the difference across the estimates.

Given the somewhat arbitrary metric imposed on the self esteem variable, and to also illustrate the applicability of our proposed procedure, we conclude the empirical investigation by re-estimating the model by bivariate probit after constructing a binary self esteem outcome variable which indicates that the individual is above the mean sample value of the continuous measure. The estimates from this procedure support those from based on the continuous measure. That is, exercising more than 5 times a week has a large and statistically significant impact on self esteem and there is clear evidence of endogeneity with a negative relationship between the unobservables. This conclusion is based on the coefficient on the exercise variable of 1.275 and an estimate of the correlation measure of -.679. The estimated average treatment effect is .467.

6 Conclusion

We provide a methodology to incorporate the presence of social interactions in a large class of models which estimate endogenous treatment effects or parameters in the presence of sample selection bias. We do so by adapting several of the estimators inspired by the seminal work of Heckman to incorporate a role for the individual's peers' treatment choices on their own and vice versa. This is done by incorporating

[8]The mean and standard deviation of V from this model are 0.202 and 0.222 respectively. Thus they are very similar to those from estimating the model jointly.

a game-theoretic approach via a discrete Bayesian game into these models to capture the simultaneity of individuals choices in the treatment equation. Accounting for the implicit interaction between choices introduces a computational burden in estimation which we address via a nested pseudo joint likelihood (NPJL) estimator. We describe how the procedure can be applied to a large number of models which are frequently employed in empirical work. Our estimator represents the first procedure to tackle these issues in the presence of social interactions and is likely to be applicable to a large number of settings. We provide simulation evidence that the estimators perform well in terms of the capacity to accurately estimate the parameters of interest and identify selection bias and endogeneity. We also apply our procedure to estimate the impact of an individual's level of exercise on their self esteem. We find that an individual's level of exercise is influenced by their expectation of their peer's exercise activity. Moreover, our procedure provides statistical evidence that exercise is endogenous to self esteem and that accounting for this endogeneity increases the impact of exercise on self esteem.

References

Aguirregabiria, V. and P. Mira (2002). Swapping the nested fixed point algorithm: A class of estimators for discrete markov decision models. *Econometrica 70*(4), 1519–1543.

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

Aradillas-López, A. (2010). Semiparametric estimation of a simultaneous game with incomplete information. *Journal of Econometrics 157*(2), 409–431.

Aradillas-López, A. (2012). Pairwise-difference estimation of incomplete information games. *Journal of Econometrics 168*(1), 120–140.

Aradillas-López, A. (2020). The econometrics of static games. *Annual Review of Economics 12*, 135–165.

Balat, J. and S. Han (2020). Multiple treatments with strategic substitutes. *Forthcoming in Journal of Econometrics*.

Bergemann, D. and S. Morris (2013). Robust predictions in games with incomplete information. *Econometrica 81*(4), 1251–1308.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of Econometrics 150*(1), 41–55.

Brock, W. and S. Durlauf (2001a). Discrete choice with social interactions. *The Review of Economic Studies 68*(2), 235–260.

Brock, W. A. and S. N. Durlauf (2001b). Interactions-based models. In *Handbook of*

Econometrics, Volume 5, pp. 3297–3380. Elsevier.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. *The Review of Economic Studies 76*(4), 1239–1267.

Campbell, R. N. (1984). *The new science: Self esteem psychology*. University Press of America.

Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica 75*(5), 1243–1284.

Ciliberto, F., C. Murry, and E. T. Tamer (2020). Market structure and competition in airline markets. *Available at SSRN 2777820*.

Ciliberto, F. and E. Tamer (2009). Market structure and multiple equilibria in airline markets. *Econometrica 77*(6), 1791–1828.

Das, M., W. K. Newey, and F. Vella (2003). Nonparametric estimation of sample selection models. *The Review of Economic Studies 70*(1), 33–58.

De Paula, A. and X. Tang (2012). Inference of signs of interaction effects in simultaneous games with incomplete information. *Econometrica 80*(1), 143–172.

Eisenhauer, P., J. Heckman, and E. Vytlacil (2015). The generalized roy model and the cost-benefit analysis of social programs. *Journal of Political Economy 123*(2), 413–443.

Ekeland, E., F. Heian, and K. B. Hagen (2005). Can exercise improve self esteem in children and young people? a systematic review of randomised controlled trials. *British Journal of Sports Medicine 39*(11), 792–798.

Fox, K. R. (2000a). The effects of exercise on self-perceptions and self-esteem. In *Physical activity and psychological well-being*, pp. 81–118.

Fox, K. R. (2000b). Self-esteem, self-perceptions and exercise. *International Journal of Sport Psychology*.

Furnham, A., N. Badmin, and I. Sneade (2002). Body image dissatisfaction: Gender differences in eating attitudes, self-esteem, and reasons for exercise. *The Journal of Psychology 136*(6), 581–596.

Glaeser, E. and J. Scheinkman (2003). Nonmarket interactions. *Advances in Economics and Econometrics 1*, 339–370.

Goldsmith-Pinkham, P. and G. W. Imbens (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics 31*(3), 253–264.

Heckman, J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. and E. Vytlacil (2007). Econometric evaluation of social programs, part i: Causal models, structural models and econometric policy evaluation. *Handbook of Econometrics 6*, 4779–4874.

Heckman, J. J. (1974). Shadow prices, market wages, and labor supply.

Econometrica 42(4), 679–694.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica 46*(4), 931–959.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Heckman, J. J. and E. Vytlacil (2001). Policy-relevant treatment effects. *American Economic Review 91*(2), 107–111.

Heckman, J. J. and E. J. Vytlacil (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences 96*(8), 4730–4734.

Honoré, B. E. and L. Hu (2020). Selection without exclusion. *Econometrica 88*(3), 1007–1029.

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

Hu, Y. and Z. Lin (2020). Binary choice with misclassification and social interactions, with an application to peer effects in attitude. *Working paper*.

Jackson, M. O., Z. Lin, and N. Yu (2020). Adjusting for peer-influence in propensity scoring when estimating treatment effects. *Working paper*.

Lee, L.-f. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics 140*(2), 333–374.

Lee, L.-f., J. Li, and X. Lin (2014). Binary choice models with social network under heterogeneous rational expectations. *Review of Economics and Statistics 96*(3), 402–417.

Lee, L.-f., X. Liu, and X. Lin (2010). Specification and estimation of social interaction models with network structures. *The Econometrics Journal 13*(2), 145–176.

Lewbel, A. (2007). Endogenous selection or treatment model estimation. *Journal of Econometrics 141*(2), 777–806.

Lewbel, A. and X. Tang (2015). Identification and estimation of games with incomplete information using excluded regressors. *Journal of Econometrics 189*(1), 229–244.

Lin, Z. (2020). Social interactions quantiles. *Working paper*.

Lin, Z. and H. Xu (2017). Estimation of social–influence–dependent peer pressures in a large network game. *The Econometrics Journal 20*(3), 86–102.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review 80*(2), 319–323.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*(3), 531–542.

Manski, C. F. (2000). Economic analysis of social interactions. *Journal of Economic Perspectives 14*(3), 115–136.

Morris, S. and H. S. Shin (2003). Global games: theory and applications. *Advances in Economics and Econometrics 1*, 56–114.

Rosenberg, M. (2015). *Society and the adolescent self-image*. Princeton university press.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica 55*(5), 999–1033.

Sonstroem, R. J. (1984). Exercise and self-esteem. *Exercise and Sport Sciences Reviews 12*, 123–155.

Sonstroem, R. J., L. L. Harlow, and L. Josephs (1994). Exercise and self-esteem: Validity of model expansion and exercise associations. *Journal of Sport and Exercise psychology 16*(1), 29–42.

Sonstroem, R. J. and W. P. Morgan (1989). Exercise and self-esteem: rationale and model. *Medicine & Science in Sports & Exercise*.

Spence, J. C., K. R. McGannon, and P. Poon (2005). The effect of exercise on global self-esteem: A quantitative review. *Journal of Sport and Exercise Psychology 27*(3), 311–334.

Strelan, P., S. J. Mehaffey, and M. Tiggemann (2003). Brief report: Self-objectification and esteem in young women: The mediating role of reasons for exercise. *Sex roles 48*(1-2), 89–95.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies 70*(1), 147–165.

Tamer, E. (2010). Partial identification in econometrics. *Annual Review of Economics 2*(1), 167–195.

Tiggemann, M. and S. Williamson (2000). The effect of exercise on body satisfaction and self-esteem as a function of gender and age. *Sex roles 43*(1-2), 119–127.

Wan, Y. and H. Xu (2014). Semiparametric identification of binary decision games of incomplete information with correlated private signals. *Journal of Econometrics 182*(2), 235–246.

Willis, R. J. and S. Rosen (1979). Education and self-selection. *Journal of Political Economy 87*(5, Part 2), S7–S36.

Xu, H. (2018). Social interactions in large networks: A game theoretic approach. *International Economic Review 59*(1), 257–284.

Appendix A Proofs

Proof of Proposition 1. Start with any $P, P' \in [0,1]^n$ and let $\tilde{P} = \Gamma(P)$ and $\tilde{P}' = \Gamma(P')$. For every $i \in \mathcal{I}$, it follows that

$$\begin{aligned}
|\tilde{P}_i - \tilde{P}'_i| &= |\Phi[Z'_i\beta_D + \alpha V_i(P)] - \Phi[Z'_i\beta_D + \alpha V_i(P')]| \\
&= \phi[Z'_i\beta_D + \alpha V_i(\bar{P})]| \cdot \left|\alpha \frac{1}{N_i} \sum_{j \in F_i} (P_j - P'_j)\right| \\
&\leq \frac{1}{\sqrt{2\pi}} \alpha \cdot \max_{j \in F_i} |P_j - P'_j| \\
&< \max_{j \in \mathcal{I}} |P_j - P'_j|
\end{aligned}$$

where $\bar{P} = (\bar{P}_1, \cdots, \bar{P}_n)$ is between P and P'. Taking maximization over $\mathcal{I}$ on the left hand side, we obtain:

$$\max_{i \in \mathcal{I}} |\tilde{P}_i - \tilde{P}'_i| < \max_{j \in \mathcal{I}} |P_j - P'_j|$$

which is a contraction. By the contraction mapping theorem, there exists a unique fixed point for Γ in $[0,1]^n$, say P^*. So $P^* = \Gamma(P^*)$. Since $P = \Gamma(P)$ if and only if P describes an equilibrium, P^* is the unique equilibrium of the Bayesian game. The proof takes a very conservative expansion of terms and the violation of MSI assumption does not necessarily cause multiple equilibria. □

Proof of Proposition 2. Denote $P^*_i = P(D_i = 1|\mathbb{I})$ and $\overline{P}_i = P(D_i = 1|\mathbb{I}^h_i)$. For every $j \in N^h_i$, by the mean value theorem, it follows that:

$$\begin{aligned}
|P^*_j - \overline{P}_j| &= \phi[Z'_j\beta_D + \alpha V_j(\widetilde{P})] \cdot \alpha \cdot \left|\frac{1}{N_j} \sum_{j' \in F_j} (P^*_{j'} - \overline{P}_{j'})\right| \\
&\leq \phi[Z'_j\beta_D + \alpha V_j(\widetilde{P})] \cdot \alpha \cdot \left|\max_{j' \in F_j} (P^*_{j'} - \overline{P}_{j'})\right|
\end{aligned}$$

where $\widetilde{P}$ is some real number between P^* and $\overline{P}$. Because for every $j \in N^{h-1}_i$, any influencer j' of j belongs to N^h_i, it follows that:

$$|P^*_{j'} - \overline{P}_{j'}| \leq \phi[Z'_{j'}\beta_D + \alpha V_{j'}(\widetilde{P})] \cdot \alpha \cdot \max_{j'' \in F_{j'}, j' \in F_j} |(P^*_{j''} - \overline{P}_{j''})|$$

By induction, for any $q \leq h$, it follows that:

$$\max_{j \in F_{(i,h-q)}} |P^*_j - \overline{P}_j| \leq \phi[Z_j\beta_D + \alpha V_j(\widetilde{P})] \cdot \alpha^{q+1} \cdot 1.$$

Since $i \in N^0_i$, we have:

$$|P_i^* - \overline{P}_i| \leq \frac{\alpha^{h+1}}{\sqrt{2\pi}} \to 0$$

as $h \to \infty$ by Assumption 2 (ii).

□

Appendix B Tables

Table 1: Statistic Summary of Variables

Variable	Mean	Standard Deviation
Age	15.638	1.635
Female	0.525	0.518
GPA	2.834	0.797
Intelligence (1&2)	0.048	0.214
Intelligence (3&4)	0.589	0.492
Intelligence (5&6)	0.363	0.481
White	0.619	0.486
Family Income	0.049	0.055
Exercise	0.408	0.491
Self-Esteem	18.763	3.559

Table 2: OLS Estimation of Self-Esteem (Continuous Self-Esteem)

Variable	Estimate	Standard Error
Age	-0.106**	0.022
Female	-0.832**	0.073
GPA	0.133**	0.049
Intelligence (3&4)	0.910	0.172
Intelligence (5&6)	2.042	0.179
White	-0.224**	0.76
Family Income	-0.287	0.667
Exercise	0.733**	0.077
Intercept	19.064**	0.421

**: 5% Significance; *: 10% Significance

Table 3: Exercise Equation with Peer Effects

Variable	Estimate	Standard Error
Age	-0.109**	0.009
Female	-0.665**	0.028
GPA	0.049**	0.019
Intelligence (3&4)	0.030	0.066
Intelligence (5&6)	0.092	0.068
White	0.228**	0.029
Family Income	0.671**	0.253
Intercept	1.375**	0.159
Peer Effects	0.153**	0.062

Table 4: Two-Step Adjusted Self-Esteem Equation (Continuous Self-Esteem)

Variable	Estimate	Standard Error
Age	0.051	0.057
Female	0.148	0.331
GPA	0.061	0.054
Intelligence (3&4)	0.848**	0.173
Intelligence (5&6)	1.890**	0.186
White	-0.559**	0.134
Family Income	-1.255*	0.739
Exercise	4.686**	1.306
GR	-2.417**	0.797
Intercept	15.030**	1.396

Table 5: Joint MLE Estimation of Exercise Equation with Peer Effects (Continuous Self-Esteem)

Variable	Estimate	Standard Error
Age	-0.107**	0.009
Female	-0.667**	0.027
GPA	0.050**	0.018
Intelligence (3&4)	0.028	0.065
Intelligence (5&6)	0.090	0.068
White	0.228**	0.029
Family Income	0.663**	0.250
Intercept	1.366**	0.159
Peer Effects	0.172**	0.060

Table 6: Joint MLE Estimation of Self-Esteem Equation (Continuous Self-Esteem)

Variable	Estimate	Standard Error
Age	-0.042	0.031
Female	-0.430**	0.153
GPA	0.104**	0.051
Intelligence (3&4)	0.885**	0.177
Intelligence (5&6)	1.980**	0.185
White	-0.362**	0.090
Family Income	-0.684	0.695
Exercise	2.357**	0.543
Intercept	17.407**	0.697
σ	3.519**	0.061
ρ	-0.282**	0.089

Table 7: Joint MLE Estimation of Exercise Equation with Peer Effects (Binary Self-Esteem)

Variable	Estimate	Standard Error
Age	-0.106**	0.009
Female	-0.660**	0.023
GPA	0.049**	0.018
Intelligence (3&4)	0.024	0.064
Intelligence (5&6)	0.085	0.067
White	0.225**	0.029
Family Income	0.660**	0.250
Intercept	1.373**	0.157
Peer Effects	0.149**	0.057

Table 8: Joint MLE Estimation of Self-Esteem Equation (Binary Self-Esteem)

Variable	Estimate	Standard Error
Age	-0.003	0.015
Female	0.076	0.080
GPA	-0.014	0.019
Intelligence (3&4)	0.144**	0.063
Intelligence (5&6)	0.480**	0.082
White	-0.170**	0.030
Family Income	-0.228	0.244
Exercise	1.275**	0.208
Intercept	-0.738**	0.307
Dependence	-0.679**	(-0.877,-0.291)

For the dependence parameter, confidence interval is reported.

Appendix C Simulation Evidence

We now examine the performance of the proposed both joint MLE and two step estimators via simulation exercises.

C.1 Endogenous Treatment Model in the Presence of Social Interactions

Consider the following model:

$$S : D_i = 1\Big\{\beta_{S0} + \beta_{S1}X_{1i} + \beta_{S2}X_{2i} + \alpha_S V_i + \varepsilon_{Si} > 0\Big\}, i \in \mathcal{I},$$
$$O : Y_i = \beta_{O0} + \beta_{O1}X_{1i} + \beta_{O2}X_{2i} + \gamma D_i + \varepsilon_{Oi}, i \in \mathcal{I},$$

where $V_i \equiv \frac{1}{N_i}\sum_{j\neq i} F_{ij}\mathbb{E}(D_j|\mathbb{I})$, X_{1i} is drawn from the standard normal distribution and X_{2i} is drawn from uniform distribution over $[-1, 1]$. The $X_i's$ are independent of the error terms generated as:

$$\begin{pmatrix} \varepsilon_{Si} \\ \varepsilon_{Oi} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

The parameters are set as $\beta_O = (-1, 1, -1)$, $\beta_S = (-0.5, 0.5, -0.5)$, $\rho = 0.6$ and $\gamma = 1$. We consider the case that $\alpha = 1$.[9] We generate a random social network of n individuals as follows. Each individual i has a degree independently drawn from $N_i \in \{0, 1, \cdots, 10\}$ with equal likelihood on each degree. We randomly choose N_i of the other $n - 1$ individuals as individual's friends. The network is directed since j can influence i without requiring (but not precluding) that i influences j. We could simulate alternative network structures but the one adopted simplifies the data generating process.

The model has a binary outcome for the treatment equation and a continuous outcome with a constant treatment effect γ. The selection model arises if the outcomes were only observed for one of the D values. In addition to the non-linearity induced in the mapping from the conditioning values to the control function the model is also identified from the inclusion of V_i which is constructed via a non linear mapping from the X_j. As V_i is only a function of exogenous variables there is no issue of endogeneity of the explanatory variables in the treatment equation. V_i is generated by the summing over the expected choices (conditional choice probabilities) of individual $i's$ friends where these probabilities are solved in the fixed point algorithm with known parameters in the treatment equation.

[9] We also investigate the performance of the model with $\alpha = 2$. The results are qualitatively similar to those with $\alpha = 1$.

We estimate the parameters via joint MLE and the two step estimator. The results for the two step estimator are reported under the heading *Endogenous Treatment 2 step (ET2S)* and we assess estimator performance by an examination of average bias and mean squared error (MSE). The joint MLE estimator is presented under the heading *Endogenous Treatment MLE (ETML).* We also provide the unadjusted OLS estimates of the outcome equation and the results for the two step estimator which replaces V_i with $AD_i \equiv \frac{1}{N_i} \sum_{j \neq i} F_{ij} D_j$, (i.e. average behavior of i's peers) in the treatment equation under the heading *Endogenous Treatment AD (ETAD).*

We begin with the unadjusted OLS estimates for the outcome equation. Table 9 indicates that the bias for all of the coefficients is large with the bias for the treatment coefficient almost 100 percent. This indicates that the design is generating substantial treatment endogeneity. Table 10 reports the bias and MSE for the binary selection/treatment equation. The ET2S and ETML each perform very well and are almost identical with respect to the degree of bias. The ETML appears slightly more efficient. The performance of ETAD is poor and this reflects the endogeneity bias resulting from the inappropriate replacement of V with AD. Table 11 reports the performance of the estimation of outcome equations associated with these procedures. The ET2S estimator shows remarkably little bias even at sample size 200. The ETML does poorly at the smaller sizes in estimating the treatment effect parameter. At sample size 800 each of these estimators perform well and again there are signs, suggested by the MSE, that the ETML is more efficient. Each of the estimators does well in estimating the parameter associated with the cross equation correlation although there is some bias in the ETML estimate of ρ at the smaller samples. Somewhat surprisingly the performance of the ETAD estimator is not as poor as the first step suggested. There are signs of bias at the smaller samples and while the performance is inferior to the ET2S and ETML estimators there is relatively little bias at sample size 800. As with all simulation exercises this reflects the design and does not suggest that V can be replaced by AD without implications for the accuracy of the second step. Although we do not explore it here it is likely that an alternative design could be constructed to produce a worse performance for ETAD.

We also explore the ability of the procedures to identify the presence of selection of bias. This is conducted via the test of statistical significance that the coefficients for GR and ρ are different from zero. This test is particularly important as an examination of Table 9 and Table 11 reveals that the adjusted and unadjusted estimates are very different. This performance is reported in Table 12. The ET2S and ETML estimators both do reasonably well in correctly identifying the presence of selection. The ETML does particularly well and by sample size 800 it is correctly identifying the presence of selection in 86 percent of the replications. This compares with 53 percent for the ET2S. The corresponding test for the ETAD estimator is only identifying selection in 25 percent of the cases at sample size 800. Thus, while the inappropriate use of AD as a conditioning variable does not result in substantial bias for the outcome equation the inability to identify selection would result in the appropriate use of the unadjusted OLS estimates. In addition to incorrect inference in the first step the use of AD may lead to incorrect inference in the second step.

Table 9: OLS Estimate

n	OLS Average Bias			
	β_O			γ
200	-0.469	-0.167	0.167	0.978
400	-0.461	-0.162	0.168	0.966
800	-0.465	-0.166	0.168	0.973
	OLS MSE			
	β_O			γ
200	0.228	0.033	0.041	0.974
400	0.217	0.028	0.034	0.942
800	0.218	0.029	0.031	0.951

Table 10: Selection Equation

n	ET2S				ETAD				ETML			
	Average Bias											
	β_S			α_S	β_S			α_S	β_S			α_S
200	-0.008	0.014	-0.005	0.015	0.289	0.009	0.000	-0.667	-0.016	0.016	-0.007	0.035
400	0.004	0.009	-0.013	-0.007	0.292	0.004	-0.007	-0.667	-0.008	0.01	-0.013	0.02
800	0.003	0.002	0.001	-0.01	0.292	-0.003	0.006	-0.672	-0.002	0.002	0.001	0.002
	Mean Squared Error											
	β_S			α_S	β_S			α_S	β_S			α_S
200	0.07	0.013	0.031	0.331	0.114	0.013	0.031	0.559	0.069	0.013	0.031	0.329
400	0.035	0.006	0.015	0.162	0.099	0.006	0.015	0.499	0.032	0.006	0.015	0.145
800	0.015	0.003	0.007	0.072	0.092	0.003	0.007	0.478	0.013	0.002	0.007	0.064

Table 11: Outcome Equation

n	ET2S					ETAD					ETML				
	Average Bias														
	β_O			γ	GR	β_O			γ	GR	β_O			γ	ρ
200	-0.025	-0.011	0.012	0.053	-0.029	-0.098	-0.037	0.037	0.206	-0.127	-0.112	-0.042	0.043	0.235	-0.194
400	0.022	0.011	-0.007	-0.046	0.023	-0.009	-0.001	0.006	0.021	-0.023	-0.059	-0.017	0.021	0.124	-0.093
800	-0.005	-0.002	0.004	0.010	-0.006	-0.021	-0.008	0.01	0.044	-0.032	-0.037	-0.013	0.015	0.077	-0.050
	Mean Squared Error														
	β_O			γ	GR	β_O			γ	GR	β_O			γ	ρ
200	0.393	0.05	0.066	1.683	0.629	0.602	0.075	0.09	2.609	0.976	0.235	0.032	0.046	1.003	0.227
400	0.161	0.023	0.027	0.697	0.258	0.317	0.040	0.048	1.388	0.515	0.080	0.012	0.017	0.345	0.094
800	0.068	0.010	0.011	0.293	0.109	0.147	0.019	0.021	0.635	0.236	0.030	0.005	0.007	0.124	0.039

Table 12: Detection of Selection Bias

n	ET2S	ETAD	ETML
200	0.175	0.102	0.549
400	0.287	0.166	0.711
800	0.504	0.256	0.862

C.2 Bivariate Probit Model in the Presence of Peer Effects

We now consider estimation of a bivariate probit model with peer effects using a similar data generating process. The model has the form:

$$Y_{1i} = 1\{\beta_{10} + \beta_{11}X_{1i} + \beta_{21}X_{2i} + \alpha V_i + \varepsilon_{Si} > 0\}$$
$$Y_{2i} = 1\{\beta_{20} + \beta_{21}X_{2i} + \gamma Y_{1i} + \varepsilon_{Oi} > 0\}$$

where $X_1, X_2 \sim \mathbb{N}(0, 1)$ and:

$$\begin{pmatrix} \varepsilon_S \\ \varepsilon_O \end{pmatrix} \sim \mathbb{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

We set $\beta_1 = (-1, 1, 1), \beta_2 = (-1, 1)$, $\rho = 0.6$ and $\gamma = 1$ and again consider $\alpha = 1$. We also continue to present the performance of the estimator when V is incorrectly replaced with AD.

Table 13 reports the results in terms of the average bias and the MSE. Our procedure, using V, is denoted BPV while that employing AD is denoted BPAD. With respect to the estimation of the parameters in the selection/treatment equation there is very little difference across procedures. The coefficient α is estimated reasonably accurately with both procedures. This clearly reflects the high level of correlation between V and AD. The biggest difference across the procedures is the estimation of the constant term in the selection equation. The estimation of the outcome equation's parameters is very similar across the two procedures although the inaccurate estimation of the selection equation constant is also manifested in a biased estimate of the outcome constant. Table 6 illustrates that both procedures identify endogeneity.

Table 13: Model Estimates

n	BPV							BPAD						
	Outcome			Selection				Outcome			Selection			
	Average Bias													
	β_2		γ	β_1			α	β_2		γ	β_1			α
200	-0.04	0.05	0.05	0.03	-0.02	0.03	0.03	0.13	0.03	0.04	-0.46	-0.02	0.03	0.02
400	-0.01	0.02	0.02	0.00	0.00	0.02	0.00	0.14	0.01	0.01	-0.48	0.00	0.02	0.00
800	0.00	0.01	0.01	-0.02	0.00	0.01	0.00	0.15	0.00	0.00	-0.48	0.00	0.01	0.00
	Mean Squared Error													
	β_2		γ	β_1			α	β_2		γ	β_1			α
200	0.06	0.03	0.03	0.37	0.05	0.03	0.19	0.06	0.03	0.03	0.41	0.05	0.03	0.19
400	0.03	0.01	0.01	0.18	0.02	0.02	0.09	0.04	0.01	0.01	0.32	0.02	0.02	0.09
800	0.01	0.01	0.01	0.08	0.01	0.01	0.04	0.03	0.01	0.01	0.28	0.01	0.01	0.04

Table 14: Detection of Selection Bias

n	BPV	BPAD
200	0.547	0.519
400	0.885	0.873
800	0.995	0.991

The evidence suggests that the estimators work well in the settings we examined. It also appears that ignoring the process by which the peer effects operate and employing the observed behavior will frequently lead to incorrect inference.