Maximum likelihood estimation of social interaction effects with nonrandom group selection

Brian Krauth
Simon Fraser University

Preliminary and Incomplete

August 1, 2004

Abstract

This paper derives a maximum likelihood estimator for an econometric model of discrete choice with social interaction effects. Endogenous selection of reference group is addressed within the econometric model through the incorporation of a reduced form within-group correlation in both observed and unobserved characteristics. The estimator requires only standard survey data which provides information on a binary choice made by the respondent, a vector of the respondent's background characteristics, and the average choice made by a large reference group (for example, school or census tract). Properties of the estimator are demonstrated analytically and through Monte Carlo experiments.

1 Introduction

This paper derives a maximum likelihood estimator for an econometric model of discrete choice with social interaction effects. In the model, individuals make a binary choice and the relative benefit to each choice is affected by the prevalence of that choice in the individual's reference group. In applications, the reference group can consist of friends, neighbors, schoolmates, or any peer group. Because the other members of the reference group face the same choice and are also influenced by the prevalence within the reference group, the actual probability distribution of choices is governed by the equilibrium of some appropriately-specified game.

The problem of developing a maximum likelihood estimator of peer effects in this context was first addressed by Brock and Durlauf (2001b). However, their estimator has seen limited if any direct use in the applied literature for two reasons. First, it requires data on the choice of every group member. While this type of data is occasionally available, there are many other cases in which a set of randomly sampled respondents are each asked both his or her own choice and the average choice in his or her reference group. Second, and most importantly, the Brock/Durlauf estimator requires that there be no correlation in unobservable characteristics. Unfortunately, as noted by Manski (1993), this condition is rarely met as it would imply random (exogenous) selection of peer group. Both economic theory and evidence indicate that individuals tend to form relatively homogeneous groups; this will lead to positive correlation between peer behavior and the respondent's unobserved characteristics. As a result, both the reduced-form coefficient on peer behaviour in a simple probit or logit regression, and the coefficient on peer behavior in the Brock/Durlauf structural estimator will be biased upwards due to standard endogeneity problems.

The econometric model proposed here addresses both problems, and provides much greater direct applicability. First, the estimator does not require complete information on peer choice, only survey data of the kind described above. Second, the model is structured so as to allow for a reduced-form correlation in both observables and unobservables between group members. This correlation is a parameter to be estimated and is identified under the restriction of equal correlation in observable and unobservable characteristics. When this restriction is inappropriate for a particular application, interval estimation of social interaction effects is possible under much weaker interval restrictions.

Although the econometric model described in this paper is applicable both to peer effects (i.e., a small reference group) and neighborhood effects (i.e., a large peer group), the appropriate estimation method depends in part on the size of the reference group. The estimator based on a closed-form likelihood function described in this paper uses large-sample approximations to overall group behavior. As a result, it is appropriate only when the reference group is relatively large; for example, a neighborhood, city, or school. A related paper (Krauth 2002) derives a simulation-based estimator that is appropriate when the reference group is small; for example, close friends or classmates. The reason for the two distinct estimators is that, with nonzero social interaction effects, the respondent's choice influences the average choice in the reference group. The maximum likelihood estimator derived here uses asymptotic approximations to ignore this influence (which goes to zero as the group size goes to infinity). With smaller group sizes, this approximation is poor and simulation is needed to eliminate the simultaneous equations bias. In other words, the bias from the large-group estimator is decreasing in the group size, while the computational cost of the simulation-based estimator is increasing with the group size. Monte Carlo results reported in this paper quantify the tradeoff somewhat.

The result of this research provides a structural estimation alternative to the instrumental variables and natural experiment approaches that have dominated empirical analysis of social interactions. While such approaches have merits, the endogeneity of the instruments used in these studies is often highly controversial and the resulting estimates of social interaction effects can be erratic. As a result, alternative methods have particular value here in constructing a robust consensus on the strength of social interaction effects. In addition, the methodology described here has several advantages over IV estimation. First, it does not require a researcher has the good fortune to discover a natural experiment, but rather works on common survey data. This greatly expands the potential set of cases in which one can construct selection-robust estimates. Second, the methodology provides a more straightforward means of ascertaining the sensitivity of parameter estimates to assumptions, through the construction of interval estimates.

Having derived the estimator, I then proceed to discuss its numerical properties using a series of Monte Carlo experiments.

1.1 Related literature

The practice of incorporating average choice or average characteristics in a person's reference group as an explanatory variable for the person's choices or outcomes has a long history. Jencks and Mayer (1990) provide a good survey of this early work on social interaction effects. Although the deep identification problems associated with measuring social interaction effects are quietly discussed in some of these papers, these problems were treated much more seriously following the methodological work of Manksi (1993). Manski identifies three problems in measuring social interaction effects by simply including peer behavior as an explanatory variable. First, there is simultaneity: if there are social interaction effects, peer behavior both influences and is influenced by the respondent's behavior. As a result, there will be correlation between peer behavior and the respondent's unobservable characteristics, a classic example of simultaneous equations bias. Second, there is selection or sorting: individuals have a tendency to form relatively homogeneous groups. As a result, there will be correlation in unobserved characteristics between peers, which also implies a correlation between the peer behavior and the respondent's unobservables. Third, there is the possibility of contextual effects: a person may be influenced by the background characteristics of the peer group as well as (or instead of) by their behavior. Without strong functional form assumptions, it is difficult to include both peer characteristics and peer behavior as explanatory variables without facing serious collinearity problems. The effect of all three of these issues is that any measure of peer influence that is based on a simple regression coefficient on peer behavior is biased upwards.

Since this work, a number of empirical studies have attempted to construct unbiased measures of social interaction effects. Some of these studies (Evans, Oates and Schwab 1992, Gaviria and Raphael 2001, Norton, Lindrooth and Ennett 1998) have used instrumental variables for group average behavior. Another set of studies (Oreopoulos 2003, Kremer and Levy 2001, Sacerdote 2001) consider natural experiments in which individuals are assigned randomly to different groups by some central authority. A third (Hoxby 2000, Arcidiacono and Nicholson 2003) exploits small and seemingly random year-by-year variation in cohort composition in schools or other organizations. While these studies provide valuable information, each approach has its weaknesses. Research using group-level instrumental variables generally uses instruments whose exogeneity is highly questionable. The natural experiments approach is confined to those cases where a person's reference group is determined by a central authority. The stream of research that exploits exogenous year-to-year variation in cohort characteristics (for example, sex ratios) runs into the standard problems associated with using data on small changes to estimate the effect of larger changes. As a result of these weaknesses, there is a stronger than usual need to use multiple methods for estimating social interaction effects.

Model-based methods for estimating social interaction effects have also appeared in many applications. Several early papers adapted models from physics to econometric analysis: Glaeser, Sacerdote, and Scheinkman (1996) adapt the "voter model" to analyze the role of social interactions in explaining variations in crime across cities. Topa (2001) adapts the "contact process" model to explain local correlations in employment rates across Chicago neighborhoods. Brock and Durlauf (Brock and Durlauf 2001b) provide a general analysis of the econometrics of estimating equilibrium models of social interactions, and suggest a model that is similar in spirit to the one described here. In addition to Krauth (2002), papers by Kooreman (1994) and Kooreman and Soetevant (2002) have outlined and used simulation-based methods for estimating more tractable variants on the Brock/Durlauf model. Tamer (2002b, 2002a) analyzes the econometric consequences of multiple equilibria in this class of models.

2 The model

The basic structure of the model is quite standard in the social interactions literature: an individual makes a binary choice, the relative utility from each choice is a function of the individual's observed characteristics, the average choice in the reference group, and a random utility term. This type of model dates back at least to Schelling (1978), and has been analyzed in an econometric context by Brock and Durlaf (Brock and Durlauf 2001a, Brock and Durlauf 2001b). The model here differs from that of Brock and Durlauf in that it allows for correlation in characteristics between group members. In addition, to facilitate modeling of this correlation, the random utility term is assumed to have a normal distribution rather than the logistic distribution assumed by Brock and Durlauf.

2.1 Preferences and choices

The economy features a set of large non-overlapping peer groups, each with n members. Groups are indexed by g and individuals are indexed within each group by i, so that the pair (g, i) identifies an individual. Where the group is unambiguous, I refer simply to "agent i." Each individual makes a binary choice $y_{gi} \in \{0, 1\}$, and has a utility function $u_{gi}(y_{gi}; \mathbf{y}_g)$ such that:

$$u_{gi}(1; \mathbf{y}_g) - u_{gi}(0; \mathbf{y}_g) = \beta \mathbf{x}_{gi} + \gamma \bar{y}_{gi} + \epsilon_{gi} \tag{1}$$

where $\mathbf{x}_{gi} \equiv \left(1, x_{gi}^1, x_{gi}^2, \ldots, x_{gi}^k\right)'$ is vector of exogenous characteristics which are observable in the data, $\mathbf{y}_g \equiv (y_{g1}, \ldots, y_{gn})'$ is the vector of choices made by the members of the group, $\bar{y}_{gi}$ is the average choice made by the other members of the group: $\left(\bar{y}_{gi} \equiv \frac{1}{n-1} \sum_{j \neq i} y_{gj}\right)$ and ϵ_{gi} is an exogenous variable which is not observed in the data. The parameter γ is the endogenous social effect; If $\gamma > 0$, agent i's incremental utility from choosing $y_{gi} = 1$ is increasing in the fraction of his or her peers that do so. As with much of the literature, this model assumes the absence of contextual effects.

2.2 Correlated effects

The exogenous variables are assumed to have a jointly normal probability distribution across all individuals which meets the following conditions:

1. Independence across groups
2. Exchangeability within groups: The joint probability distribution is not changed by any reordering of individuals within groups.
3. Normalization of random utility term:

$$\begin{aligned} E(\epsilon_{gi}) &= 0 \qquad \forall g, i \\ var(\epsilon_{gi}) &= 1 \qquad \forall g, i \end{aligned}$$

4. Independence of observable and unobservable characteristics:

$$cov(x_{gi}, \epsilon_{gj}) = 0 \qquad \forall i, j, g$$

The exchangeability condition follows from the fact that our ordering of individuals in the data is arbitrary. The normalization is standard in discrete choice models, as parameters in such models are always identified only to a linear transformation. The third assumption, of independence between observable and unobservable characteristics, is a stronger assumption. It is standard to assume in discrete choice models that an individual's random utility term is independent of his or her observed characteristics. We extend that standard assumption to include independence of the individual's random utility term with his or her peers' observed characteristics.

Given these basic conditions, we can derive the following useful parameterization. For any g, and any $i \neq j$:

$$\begin{bmatrix} \beta \mathbf{x}_{gi} \\ \beta \mathbf{x}_{gj} \\ \epsilon_{gi} \\ \epsilon_{gj} \end{bmatrix} \sim N\left(\begin{bmatrix} \beta\mu_x \\ \beta\mu_x \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\beta x} & \rho_x \sigma^2_{\beta x} & 0 & 0 \\ \rho_x \sigma^2_{\beta x} & \sigma^2_{\beta x} & 0 & 0 \\ 0 & 0 & 1 & \rho_\epsilon \\ 0 & 0 & \rho_\epsilon & 1 \end{bmatrix} \right) \tag{2}$$

where $(\mu_x, \rho_x, \sigma^2_{\beta x}, \rho_\epsilon)$ is a set of parameters to be estimated.

An alternative parameterization is as follows. Let each group be characterized by a "typical" value for x_{gi} and ϵ_{gi}, called $\bar{x}_g$ and $\bar{\epsilon}_g$ respectively. Each individual's deviation from the typical group value of a variable is an independent mean-zero random variable. In particular:

$$\begin{bmatrix} \bar{x}_g \\ \bar{\epsilon}_g \\ u_{g1} \\ v_{g1} \\ u_{g2} \\ v_{g2} \\ \vdots \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \begin{bmatrix} \Sigma_{\bar{x}} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{\bar{\epsilon}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \Sigma_u & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_v & 0 & 0 \\ 0 & 0 & 0 & 0 & \Sigma_u & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_v \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \right) \tag{3}$$

where μ_x is a k-vector of means, and $\Sigma_{\bar{x}}$ and Σ_u are both $k \times k$ covariance matrices. For each i, let:

$$\mathbf{x}_{gi} = \bar{x}_g + u_{gi} \tag{4}$$
$$\epsilon_{gi} = \bar{\epsilon}_g + v_{gi} \tag{5}$$

Let $\Sigma_x \equiv \Sigma_{\bar{x}} + \Sigma_u$, let $\sigma^2_{\beta\bar{x}} \equiv \beta\Sigma_{\bar{x}}\beta'$, and let $\sigma^2_{\beta u} = \beta\Sigma_u\beta'$. It can be easily demonstrated that these two parameterizations are equivalent where:

$$\sigma^2_{\beta x} = \sigma^2_{\beta\bar{x}} + \sigma^2_{\beta u} \tag{6}$$
$$\rho_x = \frac{\sigma^2_{\beta\bar{x}}}{\sigma^2_{\beta\bar{x}} + \sigma^2_{\beta u}} \tag{7}$$
$$\rho_\epsilon = \sigma^2_{\bar{\epsilon}} \tag{8}$$
$$1 = \sigma^2_{\bar{\epsilon}} + \sigma^2_v \tag{9}$$

Either parameterization will be used as convenience dictates.

2.3 Equilibrium

The set of *equilibrium average choice levels* for group g is defined as:

$$Y_g = \{\bar{y} \in [0,1] : \Pr(\beta\mathbf{x}_{gi} + \gamma\bar{y} + \epsilon_{gi} > 0 | \bar{x}_g, \bar{\epsilon}_g) = \bar{y}\} \tag{10}$$

For small to moderate values of γ, the set Y_g is a singleton, i.e., equilibrium is unique. However, for sufficiently large values, Y_g has as many as three elements. In order for there to be a well-defined likelihood function for this case, the model must be supplemented with an equilibrium selection rule. The equilibrium selection rule is not necessarily identified by the data and must be chosen by the researcher.

An *equilibrium selection rule* for this model is defined as a triplet $\Lambda \equiv (\lambda_L, \lambda_M, \lambda_H)$ in the unit simplex. The selection rule should be interpreted as giving the probability that the lowest, highest, or middle equilibrium is selected. In order for the likelihood function to be finite valued for all parameter values and data sets, the selection rule should also give strictly positive probability to each equilibrium.

2.4 Selection on observables and unobservables

In order to obtain point estimates of model parameters, it is necessary to impose an additional restriction on ρ_ϵ, the between-peer correlation in unobservables. The primary restriction used in this paper is that the correlation is the same as the correlation in observables, i.e.

$$\rho_\epsilon = \rho_x \tag{11}$$

The idea of using the degree of selection on observables as a proxy for the degree of selection on unobservables was first proposed by Altonji, Elder, and Taber (2000) to correct for selection effects in measuring the effect of attending a Catholic school. These authors demonstrate that equality in these two correlations will hold (in expectation) if the observables are a random subset of a large set of relevant variables. Alternatively, if the observed variables are more highly correlated between peers than the unobserved variables, the equal-correlation point estimate of the peer effect will be biased downwards. This is a distinct possibility, as personal information that is particularly easily gathered in surveys (race, sex, age) may also be more easily observed by potential friends. In any case, the model can also be estimated under alternative restrictions on ρ_ϵ, including interval restrictions. As a result, one can report results that allow readers with different beliefs about the selection process to construct their own range of estimates consistent with both the data and their prior beliefs.

3 Estimation from a random sample

This section describes how the model can be estimated from a random sample on individuals in which average peer choice is reported by the respondent. Let $\theta \equiv (\beta, \gamma, \rho_x, \rho_\epsilon, \mu_x, \Sigma_x)$ be the vector of parameters to be estimated and let θ^* be the true value of θ. Suppose that we have a random sample of size N on $(x_{gi}, y_{gi}, \bar{y}_{gi})$. Without loss of generality assign the group index g to the person described in the gth observation. This section derives the likelihood function:

$$\begin{aligned} L(\theta) &= \prod_{g=1}^{N} \Pr(y_{gi}, \bar{y}_{gi}, x_{gi}) & (12) \\ &= \prod_{g=1}^{N} \Pr(y_{gi}|x_{gi}, \bar{y}_{gi}) \Pr(\bar{y}_{gi}|x_{gi}) \Pr(x_{gi}) & (13) \end{aligned}$$

Consistent and efficient estimates of θ^* can then be derived by maximizing $L(\theta)$ or $\ln L(\theta)$.

3.1 Preliminaries

Note that the parameters in the k-vector of means μ_x and the k by k covariance matrix Σ_x can be estimated directly from the matrix of x variables using the sample averages and sample covariance matrix. Although these parameters could be incorporated directly into the maximum likelihood problem, little efficiency will be lost by simply using those standard estimates and maximizing the likelihood conditional on the x variables.s The respondent's choice y_{gi} is discrete, and the average choice in the group $\bar{y}_{gi}$ is continuous. So we can write the log-likelihood function as:

$$\begin{aligned}
\ell(\theta) &= \ln L(\theta) \qquad (14)\\
&= \sum_{g=1}^{N} y_{gi} \ln \Pr(y_{gi}=1|x_{gi},\bar{y}_{gi})\\
&+ \sum_{g=1}^{N} (1-y_{gi}) \ln(1-\Pr(y_{gi}=1|x_{gi},\bar{y}_{gi}))\\
&+ \sum_{g=1}^{N} \ln\left(\left.\frac{d\Pr(\bar{y}_{gi} \le \bar{y}|x_{gi})}{d\bar{y}}\right|_{\bar{y}=\bar{y}_{gi}} \right)
\end{aligned}$$

Now we need to derive formulas for $\Pr(y_{gi} = 1|x_{gi}, \bar{y}_{gi})$ and $\frac{d\Pr(\bar{y}_{gi} \le \bar{y}|x_{gi})}{d\bar{y}}$.

First, apply standard rules for the distributions of linear functions of multivariate normal random vectors to get:

$$\begin{bmatrix} \beta\bar{x}_g + \bar{\epsilon}_g \\ \beta x_{gi} \\ \beta x_{gi} + \epsilon_{gi} \end{bmatrix} \sim N\left(\begin{bmatrix} \beta\mu_x \\ \beta\mu_x \\ \beta\mu_x \end{bmatrix}, \begin{bmatrix} \sigma^2_{\beta\bar{x}} + \sigma^2_{\bar{\epsilon}} & \sigma^2_{\beta\bar{x}} & \sigma^2_{\beta\bar{x}} + \sigma^2_{\bar{\epsilon}} \\ \sigma^2_{\beta\bar{x}} & \sigma^2_{\beta\bar{x}} + \sigma^2_{\beta u} & \sigma^2_{\beta\bar{x}} + \sigma^2_{\beta u} \\ \sigma^2_{\beta\bar{x}} + \sigma^2_{\bar{\epsilon}} & \sigma^2_{\beta\bar{x}} + \sigma^2_{\beta u} & \sigma^2_{\beta\bar{x}} + \sigma^2_{\bar{\epsilon}} + \sigma^2_{\beta u} + \sigma^2_v \end{bmatrix} \right) \qquad (15)$$

Next, apply the standard rules for conditional distributions under the multivariate normal to derive:

$$\begin{aligned}
(\beta x_{gi} + \epsilon_{gi})|(\beta\bar{x}_g + \bar{\epsilon}_g) &\sim N\left(\beta\bar{x}_g + \bar{\epsilon}_g, \sigma^2_{\beta u} + \sigma^2_v\right) \qquad (16)\\
\beta\bar{x}_g + \bar{\epsilon}_g|x_{gi} &\sim N\left(\frac{\sigma^2_{\beta u}\beta\mu_x + \sigma^2_{\beta\bar{x}}\beta x_{gi}}{\sigma^2_{\beta u} + \sigma^2_{\beta\bar{x}}}, \frac{\sigma^2_{\bar{\epsilon}}\sigma^2_{\beta u} + \sigma^2_{\bar{\epsilon}}\sigma^2_{\beta\bar{x}} + \sigma^2_{\beta\bar{x}}\sigma^2_{\beta u}}{\sigma^2_{\beta u} + \sigma^2_{\beta\bar{x}}} \right) \qquad (17)\\
\beta x_{gi} + \epsilon_{gi}|(\beta\bar{x}_g + \bar{\epsilon}_g), x_{gi} &\sim N(M_{gi}, S) \qquad (18)
\end{aligned}$$

where

$$\begin{aligned}
S &\equiv \frac{\sigma^2_{\bar{\epsilon}}\sigma^2_{\beta u}\sigma^2_v + \sigma^2_{\bar{\epsilon}}\sigma^2_{\beta u}\sigma^2_{\beta\bar{x}} + \sigma^2_{\bar{\epsilon}}\sigma^2_{\beta\bar{x}}\sigma^2_v + \sigma^2_{\beta\bar{x}}\sigma^2_{\beta u}\sigma^2_v}{\sigma^2_{\bar{\epsilon}}\sigma^2_{\beta u} + \sigma^2_{\bar{\epsilon}}\sigma^2_{\beta\bar{x}} + \sigma^2_{\beta\bar{x}}\sigma^2_{\beta u}} \qquad (19)\\
M_{gi} &\equiv \frac{\left(\sigma^2_{\bar{\epsilon}}\sigma^2_{\beta u} + \sigma^2_{\bar{\epsilon}}\sigma^2_{\beta\bar{x}}\right)(\beta\bar{x}_g + \bar{\epsilon}_g) + \left(\sigma^2_{\bar{\epsilon}}\sigma^2_{\beta u} + \sigma^2_{\beta\bar{x}}\sigma^2_{\beta u}\right)\beta x_{gi} - \sigma^2_{\bar{\epsilon}}\sigma^2_{\beta u}\beta\mu_x}{\sigma^2_{\bar{\epsilon}}\sigma^2_{\beta u} + \sigma^2_{\bar{\epsilon}}\sigma^2_{\beta\bar{x}} + \sigma^2_{\beta\bar{x}}\sigma^2_{\beta u}}
\end{aligned}$$

These three distributions can be used to derive the likelihood function.

3.2 Calculating $\Pr(\bar{y}_g|x_{gi})$

First, we calculate $\Pr(\bar{y}_{gi} \leq \bar{y}|x_{gi})$. Now,

$$\Pr(y_{gi} = 1|\beta\bar{x}_g + \bar{\epsilon}_g) = \Pr(\beta x_{gi} + \epsilon_{gi} + \gamma\bar{y}_g > 0|\bar{x}_g + \bar{\epsilon}_g) \quad (20)$$

Since this equation describes the CDF of a linear function of a vector of normally distributed random variables, we can normalize to get:

$$\Pr(y_{gi} = 1|\beta\bar{x}_g + \bar{\epsilon}_g) = \Phi\left(\frac{\gamma\bar{y}_g + \beta\bar{x}_g + \bar{\epsilon}_g}{\sqrt{\sigma^2_{\beta u} + \sigma^2_v}}\right) \quad (21)$$

where Φ is the CDF of the standard normal distribution.

Equation (21) holds for all i, so $E(\bar{y}_g|\beta\bar{x}_g + \bar{\epsilon}_g) = E(y_{gi}|\bar{x}_g + \bar{\epsilon}_g)$. Applying the law of large numbers, we have:

$$\bar{y}_g = \Phi\left(\frac{\gamma\bar{y}_g + \beta\bar{x}_g + \bar{\epsilon}_g}{\sqrt{\sigma^2_{\beta u} + \sigma^2_v}}\right) \quad (22)$$

Equation (22) describes the relationship between group characteristics and equilibrium group choice.

For a given set of parameter values, equation (22) defines for each value of $\bar{y}_g$, a uniquely defined value of $(\beta\bar{x}_g + \bar{\epsilon}_g)$ such that $\bar{y}_g$ is an equilibrium.

$$(\beta\bar{x}_g + \bar{\epsilon}_g) = \Phi^{-1}(\bar{y}_g)\sqrt{\sigma^2_{\beta u} + \sigma^2_v} - \gamma\bar{y}_g \quad (23)$$

However, the converse is not necessarily true: there may be multiple equilibria. Note that equation (22) defines a function mapping each value of $\bar{y}_g$ into a value of $(\beta\bar{x}_g + \bar{\epsilon}_g)$. There is a corresponding function mapping $(\beta\bar{x}_g + \bar{\epsilon}_g)$ into $\bar{y}_g$ if and only if the function defined in equation (23) is invertible (i.e., monotonic). Taking derivatives, we get:

$$\frac{\partial(\beta\bar{x}_g + \bar{\epsilon}_g)}{\partial\bar{y}_g} = \frac{\sqrt{\sigma^2_{\beta u} + \sigma^2_v}}{\phi(\Phi^{-1}(\bar{y}_g))} - \gamma \quad (24)$$

where ϕ is the PDF of the standard normal distribution. The parameter values imply a unique equilibrium if and only if this derivative is positive for all values of $\bar{y}_g$. The minimal value of this derivative occurs for $\bar{y}_g = 0.5$, as $\Phi^{-1}(0.5) = 0$. Substituting and solving, the equilibrium is always unique if

$$\gamma < \frac{\sqrt{\sigma^2_{\beta u} + \sigma^2_v}}{\phi(0)} \quad (25)$$

and potentially multiple otherwise. Notice that multiplicity is a function of parameter values, not of any feature in the data. The form of the likelihood depends on whether or not equilibrium is unique.

3.2.1 Case 1: Unique equilibrium

First, suppose that equation (25) holds and equilibrium is unique. Then the equilibrium $\bar{y}_g$ is a strictly increasing function of $\beta\bar{x}_g + \bar{\epsilon}_g$. An example is depicted in Figure 1. We can derive:

$$\Pr(\bar{y}_g \leq \bar{y}|x_{gi}) = \Pr(\beta\bar{x}_g + \bar{\epsilon}_g \leq \Phi^{-1}(\bar{y})\sqrt{\sigma^2_{\beta u} + \sigma^2_v} - \gamma\bar{y}|x_{gi}) \tag{26}$$

We already showed that:

$$\beta\bar{x}_g + \bar{\epsilon}_g|x_{gi} \sim N\left(\frac{\sigma^2_{\beta u}\beta\mu_x + \sigma^2_{\beta\bar{x}}\beta x_{gi}}{\sigma^2_{\beta u} + \sigma^2_{\beta\bar{x}}}, \frac{\sigma^2_{\epsilon}\sigma^2_{\beta u} + \sigma^2_{\epsilon}\sigma^2_{\beta\bar{x}} + \sigma^2_{\beta\bar{x}}\sigma^2_{\beta u}}{\sigma^2_{\beta u} + \sigma^2_{\beta\bar{x}}}\right) \tag{27}$$

We use the usual trick of normalizing to get

$$\Pr(\bar{y}_g \leq \bar{y}|x_{gi}) = \Phi\left(\frac{\Phi^{-1}(\bar{y})\sqrt{\sigma^2_{\beta u} + \sigma^2_v} - \gamma\bar{y} - \frac{\sigma^2_{\beta u}\beta\mu_x + \sigma^2_{\beta\bar{x}}\beta x_{gi}}{\sigma^2_{\beta u} + \sigma^2_{\beta\bar{x}}}}{\sqrt{\frac{\sigma^2_{\epsilon}\sigma^2_{\beta u} + \sigma^2_{\epsilon}\sigma^2_{\beta\bar{x}} + \sigma^2_{\beta\bar{x}}\sigma^2_{\beta u}}{\sigma^2_{\beta u} + \sigma^2_{\beta\bar{x}}}}}\right) \tag{28}$$

Now we take the derivative and evaluate at $\bar{y} = \bar{y}_g$ to get:

$$\begin{aligned}\left.\frac{\partial \Pr(\bar{y}_g \leq \bar{y}|x_{gi})}{\partial \bar{y}}\right|_{\bar{y}=\bar{y}_g} &= \left(\frac{\frac{\sqrt{\sigma^2_{\beta u} + \sigma^2_v}}{\phi(\Phi^{-1}(\bar{y}_g))} - \gamma}{\sqrt{\frac{\sigma^2_{\epsilon}\sigma^2_{\beta u} + \sigma^2_{\epsilon}\sigma^2_{\beta\bar{x}} + \sigma^2_{\beta\bar{x}}\sigma^2_{\beta u}}{\sigma^2_{\beta u} + \sigma^2_{\beta\bar{x}}}}}\right) \\ &\times \phi\left(\frac{\Phi^{-1}(\bar{y}_g)\sqrt{\sigma^2_{\beta u} + \sigma^2_v} - \gamma\bar{y}_g - \frac{\sigma^2_{\beta u}\beta\mu_x + \sigma^2_{\beta\bar{x}}\beta x_{gi}}{\sigma^2_{\beta u} + \sigma^2_{\beta\bar{x}}}}{\sqrt{\frac{\sigma^2_{\epsilon}\sigma^2_{\beta u} + \sigma^2_{\epsilon}\sigma^2_{\beta\bar{x}} + \sigma^2_{\beta\bar{x}}\sigma^2_{\beta u}}{\sigma^2_{\beta u} + \sigma^2_{\beta\bar{x}}}}}\right)\end{aligned} \tag{29}$$

3.2.2 Case 2: Multiple equilibria

When equation (25) is not satisfied, equilibrium is not always unique. Figure 2 depicts the equilibrium correspondence graphically for the case where there are multiple equilibria. For $(\beta\bar{x}_g + \bar{\epsilon}_g)$ between x^L and x^H, there are three equilibria.

Next we need to solve for the four quantities $(y^L, y^{ML}, y^{MH}, y^H)$ depicted in Figure 2. First we find y^{MH}, which is the highest possible unstable equilibrium:

$$y^{MH} \equiv \max y \text{ such that } \left(\phi(\Phi^{-1}(y)) = \frac{\sqrt{\sigma^2_{\beta u} + \sigma^2_v}}{\gamma}\right) \tag{30}$$

The lowest possible unstable equilibrium y^{ML} is:

$$y^{ML} \equiv \min y \text{ such that } \left(\phi(\Phi^{-1}(y)) = \frac{\sqrt{\sigma_{\beta u}^2 + \sigma_v^2}}{\gamma}\right) \quad (31)$$

Now x^L is the value of $(\beta\bar{x}_g + \bar{\epsilon}_g)$ such that y^{MH} is an equilibrium, and x^H is the value such that y^{ML} is an equilibrium:

$$x^L \equiv \Phi^{-1}\left(y^{MH}\right)\sqrt{\sigma_{\beta u}^2 + \sigma_v^2} - \gamma y^{MH} \quad (32)$$

$$x^H \equiv \Phi^{-1}\left(y^{ML}\right)\sqrt{\sigma_{\beta u}^2 + \sigma_v^2} - \gamma y^{ML} \quad (33)$$

The low equilibrium for x^L is denoted by y^L:

$$y^L \equiv \lim_{i\to\infty} y_i^L \text{ where } y_0^L = 0, y_{i+1}^L = \Phi\left(\frac{\gamma y_i^L + x^L}{\sqrt{\sigma_{\beta u}^2 + \sigma_v^2}}\right) \quad (34)$$

starting with $y_0^L = 0$. The high equilibrium for x^H is denoted by y^H:

$$y^H \equiv \lim_{i\to\infty} y_i^H \text{ where } y_0^H = 1, y_{i+1}^H = \Phi\left(\frac{\gamma y_i^H + x^H}{\sqrt{\sigma_{\beta u}^2 + \sigma_v^2}}\right) \quad (35)$$

Having determined these quantities, let $f(.)$ be the conditional PDF for the random variable $(\beta\bar{x}_g + \bar{\epsilon}_g)$, conditional on βx_{gi}, and let $h(.)$ be the right side of equation (23). Then:

$$\Pr(\bar{y}_g \leq \bar{y}) = \begin{cases} \int_{-\infty}^{h(\bar{y})} f(x)dx & \bar{y} < y^L \\ \int_{-\infty}^{x^L} f(x)dx + \lambda_L \int_{x^L}^{h(\bar{y})} f(x)dx & \bar{y} \in [y^L, y^{ML}] \\ \int_{-\infty}^{x^L} f(x)dx + \lambda_L \int_{x^L}^{x^H} f(x)dx + \lambda_M \int_{h(\bar{y})}^{x^H} f(x)dx & \bar{y} \in [y^{ML}, y^{MH}] \\ \int_{-\infty}^{x^L} f(x)dx + (\lambda_L + \lambda_M)\int_{x^L}^{x^H} f(x)dx + \lambda_H \int_{x^L}^{h(\bar{y})} f(x)dx & \bar{y} \in [y^{MH}, y^H] \\ \int_{-\infty}^{h(\bar{y})} f(x)dx & \bar{y} > y^H \end{cases} \quad (36)$$

Taking derivatives, we get:

$$\begin{aligned} \frac{\partial \Pr(\bar{y} \leq \bar{y}_i)}{\partial \bar{y}_i} &= \phi\left(\frac{\Phi^{-1}(\bar{y})\sqrt{\sigma_{\beta u}^2 + \sigma_v^2} - \gamma\bar{y} - \frac{\sigma_{\beta u}^2 \beta\mu_x + \sigma_{\beta\bar{x}}^2 x_i}{\sigma_{\beta u}^2 + \sigma_{\beta\bar{x}}^2}}{\sqrt{\frac{\sigma_{\bar{\epsilon}}^2\sigma_{\beta u}^2 + \sigma_{\bar{\epsilon}}^2\sigma_{\beta\bar{x}}^2 + \sigma_{\beta\bar{x}}^2\sigma_{\beta u}^2}{\sigma_{\beta u}^2 + \sigma_{\beta\bar{x}}^2}}}\right) \\ &* \left(\frac{\frac{\sqrt{\sigma_{\beta u}^2 + \sigma_v^2}}{\phi(\Phi^{-1}(\bar{y}))} - \gamma}{\sqrt{\frac{\sigma_{\bar{\epsilon}}^2\sigma_{\beta u}^2 + \sigma_{\bar{\epsilon}}^2\sigma_{\beta\bar{x}}^2 + \sigma_{\beta\bar{x}}^2\sigma_{\beta u}^2}{\sigma_{\beta u}^2 + \sigma_{\beta\bar{x}}^2}}}\right) \\ &* \Lambda_i \end{aligned} \quad (37)$$

where:

$$\Lambda_i \equiv \begin{cases} 1 & \text{if } \bar{y} \in (-\infty, y^L) \cup (y^H, \infty) \\ \lambda_L & \text{if } \bar{y} \in [y^L, y^{ML}] \\ -\lambda_M & \text{if } \bar{y} \in [y^{ML}, y^{MH}] \\ \lambda_H & \text{if } \bar{y} \in [y^{MH}, y^H] \end{cases} \tag{38}$$

3.3 Calculating $\Pr(y_{gi}|\bar{y}_g, x_{gi})$

Finally, we look at y_{gi}. We note that

$$\begin{aligned} \Pr(y_{gi} = 1|x_{gi}, \bar{y}_g) &= \Pr(x_{gi} + \epsilon_{gi} + \gamma\bar{y}_g > 0|x_{gi}, \bar{y}_g) \qquad (39) \\ &= \Pr\left(\frac{(x_{gi} + \epsilon_{gi}) - M_{gi}}{\sqrt{S}} > \frac{-\gamma\bar{y}_g - M_{gi}}{\sqrt{S}}\middle| x_{gi}, \bar{y}_g\right) \\ &= \Phi\left(\frac{\gamma\bar{y}_g + M_{gi}}{\sqrt{S}}\right) \end{aligned}$$

where

$$\begin{aligned} S &\equiv \frac{\sigma_{\bar{\epsilon}}^2\sigma_{\beta u}^2\sigma_v^2 + \sigma_{\bar{\epsilon}}^2\sigma_{\beta u}^2\sigma_{\beta\bar{x}}^2 + \sigma_{\bar{\epsilon}}^2\sigma_{\beta\bar{x}}^2\sigma_v^2 + \sigma_{\beta\bar{x}}^2\sigma_{\beta u}^2\sigma_v^2}{\sigma_{\bar{\epsilon}}^2\sigma_{\beta u}^2 + \sigma_{\bar{\epsilon}}^2\sigma_{\beta\bar{x}}^2 + \sigma_{\beta\bar{x}}^2\sigma_{\beta u}^2} \qquad (40) \\ M_{gi} &\equiv \frac{\left(\sigma_{\bar{\epsilon}}^2\sigma_{\beta u}^2 + \sigma_{\bar{\epsilon}}^2\sigma_{\beta\bar{x}}^2\right)\left(\Phi^{-1}(\bar{y}_g)\sqrt{\sigma_{\beta u}^2 + \sigma_v^2} - \gamma\bar{y}_g\right) + \left(\sigma_{\bar{\epsilon}}^2\sigma_{\beta u}^2 + \sigma_{\beta\bar{x}}^2\sigma_{\beta u}^2\right)\beta x_{gi} - \sigma_{\bar{\epsilon}}^2\sigma_{\beta u}^2\beta\mu_x}{\sigma_{\bar{\epsilon}}^2\sigma_{\beta u}^2 + \sigma_{\bar{\epsilon}}^2\sigma_{\beta\bar{x}}^2 + \sigma_{\beta\bar{x}}^2\sigma_{\beta u}^2} \end{aligned}$$

3.4 Calculating the likelihood function

To summarize, the log-likelihood function is given by:

$$\begin{aligned} \ell(\theta) &= \sum_{g=1}^{n} y_{gi}\ln\left(\Phi\left(\frac{\gamma\bar{y}_g + M_{gi}}{\sqrt{S}}\right)\right) \\ &+ \sum_{g=1}^{n}(1 - y_{gi})\ln\left(1 - \Phi\left(\frac{\gamma\bar{y}_g + M_{gi}}{\sqrt{S}}\right)\right) \\ &+ \sum_{g=1}^{n}\ln\left|\frac{\frac{\sqrt{\sigma_{\beta u}^2+\sigma_v^2}}{\phi(\Phi^{-1}(\bar{y}_g))} - \gamma}{\sqrt{\frac{\sigma_{\bar{\epsilon}}^2\sigma_{\beta u}^2+\sigma_{\bar{\epsilon}}^2\sigma_{\beta\bar{x}}^2+\sigma_{\beta\bar{x}}^2\sigma_{\beta u}^2}{\sigma_{\beta u}^2+\sigma_{\beta\bar{x}}^2}}}\right| \\ &+ \sum_{g=1}^{n}\ln\phi\left(\frac{\Phi^{-1}(\bar{y}_g)\sqrt{\sigma_{\beta u}^2 + \sigma_v^2} - \gamma\bar{y}_g - \frac{\sigma_{\beta u}^2\beta\mu_x+\sigma_{\beta\bar{x}}^2\beta x_{gi}}{\sigma_{\beta u}^2+\sigma_{\beta\bar{x}}^2}}{\sqrt{\frac{\sigma_{\bar{\epsilon}}^2\sigma_{\beta u}^2+\sigma_{\bar{\epsilon}}^2\sigma_{\beta\bar{x}}^2+\sigma_{\beta\bar{x}}^2\sigma_{\beta u}^2}{\sigma_{\beta u}^2+\sigma_{\beta\bar{x}}^2}}}\right) \\ &+ \sum_{g=1}^{n}\ln|\Lambda_{gi}| \qquad (41) \end{aligned}$$

where Λ_{gi}, M_{gi}, and S are as defined in equations (38) and (40),

$$\mu_x = \hat{\mu}_x \tag{42}$$

$$\sigma^2_{\beta\bar{x}} + \sigma^2_{\beta u} = \beta\hat{\Sigma}_x\beta' \tag{43}$$

and $(\hat{\mu}_x, \hat{\Sigma}_x)$ are just the sample mean and covariance matrix of the x_{gi}'s.

It may also be useful to define the log-likelihood function in terms of the alternative parameterization. This can be done by simply substituting into equation (41):

$$\begin{aligned}
\ell(\theta) &= \sum_{g=1}^{n} y_{gi} \ln\left(\Phi\left(\frac{\gamma\bar{y}_g + M_{gi}}{\sqrt{S}}\right)\right) \\
&+ \sum_{g=1}^{n} (1 - y_{gi}) \ln\left(1 - \Phi\left(\frac{\gamma\bar{y}_g + M_{gi}}{\sqrt{S}}\right)\right) \\
&+ \sum_{g=1}^{n} \ln\left|\frac{\frac{\sqrt{(1-\rho_x)\sigma^2_{\beta x} + 1 - \rho_\epsilon}}{\phi(\Phi^{-1}(\bar{y}_g))} - \gamma}{\sqrt{\rho_\epsilon + \rho_x(1-\rho_x)\sigma^2_{\beta x}}}\right| \\
&+ \sum_{g=1}^{n} \ln\phi\left(\frac{\Phi^{-1}(\bar{y}_g)\sqrt{(1-\rho_x)\sigma^2_{\beta x} + 1 - \rho_\epsilon} - \gamma\bar{y}_g - (1-\rho_x)\beta\mu_x - \rho_x\beta x_{gi}}{\sqrt{\rho_\epsilon + \rho_x(1-\rho_x)\sigma^2_{\beta x}}}\right) \\
&+ \sum_{g=1}^{n} \ln|\Lambda_{gi}|
\end{aligned} \tag{44}$$

where:

$$S \equiv \frac{\rho_\epsilon(1-\rho_\epsilon) + \rho_x(1-\rho_x)\sigma^2_{\beta x}}{\rho_\epsilon + \rho_x(1-\rho_x)\sigma^2_{\beta x}}$$

$$M_{gi} \equiv \frac{\rho_\epsilon\left(\Phi^{-1}(\bar{y}_g)\sqrt{(1-\rho_x)\sigma^2_{\beta x} + 1 - \rho_\epsilon} - \gamma\bar{y}_g\right) + \left(\rho_\epsilon * (1-\rho_x) + \rho_x(1-\rho_x)\sigma^2_{\beta x}\right)\beta x_{gi} - \rho_\epsilon(1 -}{\rho_\epsilon + \rho_x(1-\rho_x)\sigma^2_{\beta x}}$$

Λ_i is as defined in equation (38), with:

$$y^{MH} \equiv \max y \text{ such that } \left(\phi(\Phi^{-1}(y)) = \frac{\sqrt{(1-\rho_x)\sigma^2_{\beta x} + 1 - \rho_\epsilon}}{\gamma}\right)$$

$$y^{ML} \equiv \min y \text{ such that } \left(\phi(\Phi^{-1}(y)) = \frac{\sqrt{(1-\rho_x)\sigma^2_{\beta x} + 1 - \rho_\epsilon}}{\gamma}\right)$$

$$x^L \equiv \Phi^{-1}\left(y^{MH}\right)\sqrt{(1-\rho_x)\sigma^2_{\beta x} + 1 - \rho_\epsilon} - \gamma y^{MH}$$

$$
\begin{aligned}
x^H &\equiv \Phi^{-1}\left(y^{ML}\right)\sqrt{(1-\rho_x)\sigma^2_{\beta x}+1-\rho_\epsilon}-\gamma y^{ML} \\
y^L &\equiv \lim_{i\to\infty} y_i^L \text{ where } y_0^L=0, y_{i+1}^L=\Phi\left(\frac{\gamma y_i^L+x^L}{\sqrt{(1-\rho_x)\sigma^2_{\beta x}+1-\rho_\epsilon}}\right) \\
y^H &\equiv \lim_{i\to\infty} y_i^H \text{ where } y_0^H=1, y_{i+1}^H=\Phi\left(\frac{\gamma y_i^H+x^H}{\sqrt{(1-\rho_x)\sigma^2_{\beta x}+1-\rho_\epsilon}}\right)s
\end{aligned}
$$

and

$$
\begin{aligned}
\mu_x &= \hat{\mu}_x && (46) \\
\sigma^2_{\beta x} &= \beta\hat{\Sigma}_x\beta' && (47)
\end{aligned}
$$

where $(\hat{\mu}_x, \hat{\Sigma}_x)$ are just the sample mean and covariance matrix of the x_{gi}'s.

4 Monte Carlo results

This section describes the results of a series of Monte Carlo experiments aimed at characterizing the performance of the model in different environments.

The baseline experiment has the following features:

1. The vector of explanatory variables x_{gi} consists of an intercept and a single random variable with a $N(0,1)$ distribution across g.
2. The number of group members is varied across experiments, as are the values of the parameters ρ, γ, β_0 (the intercept), and β_1 (the coefficient on the explanatory variable). In most cases the intercept is set at $\beta_0 = \gamma/2$, in order to avoid having an extremely large fraction of observations at the same value of y_{gi}.
3. A single member from each group is sampled in order to generate a random sample from a large population.
4. For a finite number of group members, instead of the group choice outcome being determined by the "equilibrium average choice", the outcome is chosen from the set of Nash equilibria. The set of equilibrium average choice levels is simply the limit of the set of Nash equilibria as the group size goes to infinity.
5. The equilibrium selection rule in the simulated data is as follows: the lowest-average-choice Nash equilibrium is selected.

6. The equilibrium selection rule assumed in estimation is $(1/3, 1/3, 1/3)$.

Given these features, the Monte Carlo experiment is standard. For each value of θ and n_g, we draw a sample of size $N = 1,000$ and calculate the maximum likelihood parameter estimate $\hat{\theta}$. We repeat this experiment 100 times in order to get the approximate probability distribution of $\hat{\theta}$. In particular, we are primarily concerned with $E(\hat{\rho} - \rho)$ and $E(\hat{\gamma} - \gamma)$, the bias in the estimates of the peer effect and selection effect.

Table 1 reports the results from the Monte Carlo experiments. As one might expect, the estimator performs very poorly for small reference groups, with a large upward bias in $\hat{\gamma}$. This upward bias is not surprising because the ML estimator derived in this paper uses approximations which are motivated by a large reference group. In those cases, the simulation-based estimator in Krauth (2002) will be more appropriate. As one might also expect, the performance of the estimator improves significantly as the size of the reference group increases. In most of the experiments, the bias of the ML estimator is negligible for reference groups with more than 100 to 200 members.

Next, we consider the variance of the estimator. Two questions arise with respect to the variance. First, how accurate is the asymptotic covariance matrix of the estimator in moderately sized samples? Second, are the social interaction effect parameters estimated with reasonable precision in moderately sized samples? In order to answer this question, Table 2 reports estimates of the standard error and correlation coefficient of $\hat{\gamma}$ and $\hat{\rho}$ using both the asymptotic covariance matrix (averaged across the trials), and using the sample standard deviation and correlation coefficient across the trials. In the interests of space, Table 2 only reports these results for those values of the reference group size that do not imply large bias in the coefficient estimates. The results suggest that the asymptotic covariance matrix estimator performs fairly well on sample sizes of 1,000 observations or more, and that the estimator itself has moderate variance for such sample sizes.

5 Conclusion

The model and estimation method outlined in this paper enable a researcher to consistently estimate social interaction effects in binary choice from randomly sampled survey data, while allowing for at least some types of selection effects. The method can be used in a wide variety of applications, including estimating neighborhood effects and school effects on teen pregnancy, alcohol, tobacco, and drug use, school dropout, and many other

ρ	γ	β_0	β_1	Stat.	Size of reference group 5	10	20	50	100	200	500	∞
0.1	0.0	0.0	1.0	$Avg(\hat{\rho})$								
				$Avg(\hat{\gamma})$								
0.1	0.5	-0.25	1.0	$Avg(\hat{\rho})$								
				$Avg(\hat{\gamma})$								
0.1	1.0	-0.5	1.0	$Avg(\hat{\rho})$	0.06	0.17	0.20	0.17	0.13	0.12	0.11	0.10
				$Avg(\hat{\gamma})$	2.72	1.55	0.83	0.62	0.86	0.79	0.92	1.00
0.1	3.0	-1.5	1.0	$Avg(\hat{\rho})$								
				$Avg(\hat{\gamma})$								
0.5	0.0	0.0	1.0	$Avg(\hat{\rho})$	0.10	0.31	0.41	0.49	0.50	0.50	0.50	0.49
				$Avg(\hat{\gamma})$	2.42	1.26	0.64	0.20	0.10	0.04	0.01	0.01
0.5	0.5	-0.25	1.0	$Avg(\hat{\rho})$	0.07	0.23	0.36	0.46	0.49	0.50	0.51	0.50
				$Avg(\hat{\gamma})$	2.85	2.16	1.38	0.81	0.64	0.58	0.45	0.49
0.5	1.0	-0.5	1.0	$Avg(\hat{\rho})$								
				$Avg(\hat{\gamma})$								
0.5	3.0	-1.5	1.0	$Avg(\hat{\rho})$								
				$Avg(\hat{\gamma})$								
0.1	0.5	-0.25	0.1	$Avg(\hat{\rho})$								
				$Avg(\hat{\gamma})$								
0.5	0.5	-0.25	0.1	$Avg(\hat{\rho})$								
				$Avg(\hat{\gamma})$								
0.5	0.0	-0.25	0.1	$Avg(\hat{\rho})$								
				$Avg(\hat{\gamma})$								
0.1	3.0	-1.5	1.0	$Avg(\hat{\rho})$								
		(L)		$Avg(\hat{\gamma})$								
0.5	3.0	-1.5	1.0	$Avg(\hat{\rho})$								
		(L)		$Avg(\hat{\gamma})$								
0.1	3.0	-1.5	1.0	$Avg(\hat{\rho})$								
		(H)		$Avg(\hat{\gamma})$								
0.5	3.0	-1.5	1.0	$Avg(\hat{\rho})$								
		(H)		$Avg(\hat{\gamma})$								

Table 1: Monte Carlo results, based on 100 trials per experiment. Cases marked with (L) were estimated using $(\lambda_L, \lambda_M, \lambda_H) = (0.9, 0.05, 0.05)$, others were estimated using $(\lambda_L, \lambda_M, \lambda_H) = (1/3, 1/3, 1/3)$.

$(\rho, \gamma, \beta_0, \beta_1)$	Estimator	Size of reference group								
		100			200			500		
		$\hat{\sigma}_{\hat{\gamma}}$	$\hat{\sigma}_{\hat{\rho}}$	$\hat{\rho}_{\hat{\gamma}\hat{\rho}}$	$\hat{\sigma}_{\hat{\gamma}}$	$\hat{\sigma}_{\hat{\rho}}$	$\hat{\rho}_{\hat{\gamma}\hat{\rho}}$	$\hat{\sigma}_{\hat{\gamma}}$	$\hat{\sigma}_{\hat{\rho}}$	$\hat{\rho}_{\hat{\gamma}\hat{\rho}}$
$(0.1, 0.0, 0.0, 1.0)$	Asymptotic	n	n	n	n	n	n	n	n	n
	MC Sample	n	n	n	n	n	n	n	n	n
$(0.1, 0.0, 0.0, 1.0)$	Asymptotic	n	n	n	n	n	n	n	n	n
	MC Sample	n	n	n	n	n	n	n	n	n
$(0.1, 0.0, 0.0, 1.0)$	Asymptotic	n	n	n	n	n	n	n	n	n
	MC Sample	n	n	n	n	n	n	n	n	n
$(0.1, 0.0, 0.0, 1.0)$	Asymptotic	n	n	n	n	n	n	n	n	n
	MC Sample	n	n	n	n	n	n	n	n	n

Table 2: Monte Carlo results, based on 100 trials per experiment (calculations in progress).

choices of interest to policymakers.

There are several avenues of future research along these lines. First, the estimator derived in Section 3 is based on a random sampling design. However, studies of social interaction effects often are based on what could be called a group-stratified sample. In other words, there is a random sampling of groups, then within each group all or a portion of individuals are surveyed. For example, Gaviria and Raphael (2001) use a school-based sample in which all students in a particular set of schools were surveyed. Although the estimator derived in this paper can be used to make consistent parameter estimates for this type of data (by simply randomly dropping all but one observation from each group), such an estimator is clearly inefficient. As a result, an extension of the methodology to efficiently use group-stratified samples would be useful.

References

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," Working Paper 7831, NBER 2000.

Arcidiacono, Peter and Sean Nicholson, "Peer effects in medical school," *Journal of Public Economics*, 2003, *Forthcoming.*

Brock, William A. and Steven N. Durlauf, "Discrete choice with social interactions," *Review of Economic Studies*, 2001, *68* (2), 235–260.

___ **and** ___ , "Interactions-based models," in James J. Heckman and Edward Leamer, eds., *Handbook of Econometrics, Volume 5*, North-Holland, 2001.

Evans, William N., Wallace E. Oates, and Robert M. Schwab, "Measuring peer group effects: A study of teenage behavior," *Journal of Political Economy*, 1992, *100* (5), 966–991.

Gaviria, Alejandro and Steven Raphael, "School-based peer effects and juvenile behavior," *Review of Economics and Statistics*, 2001, *83* (2), 257–268.

Glaeser, Edward L., Bruce I. Sacerdote, and José A. Scheinkman, "Crime and Social Interactions," *Quarterly Journal of Economics*, 1996, *111* (2), 507–548.

Hoxby, Caroline M., "Peer effects in the classroom: Learning from gender and race variation," Working Paper 7867, NBER 2000.

Jencks, Christopher and Susan E. Mayer, "Residential Segregation, Job Proximity, and Black Job Opportunities," in Lawrence E. Lynn, Jr and Michael G.H. McGeary, eds., *Inner-City Poverty in the United States*, National Academy Press, 1990, pp. 187–222.

Kooreman, Peter, "Estimation of econometric models of some discrete games," *Journal of Applied Econometrics*, 1994, *9*, 255–268.

___ **and Adriaan Soetevent**, "A discrete choice model with social interactions: An analysis of high school teen behavior," Working Paper, University of Groningen 2002.

Krauth, Brian V., "Simulation-based estimation of peer effects," Working Paper, Simon Fraser University 2002.

Kremer, Michael and Dan M. Levy, "Peer effects from alcohol use among college students," Working Paper 9876, NBER 2001.

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

Norton, Edward C., Richard C. Lindrooth, and Susan T. Ennett, "Controlling for the endogeneity of peer substance use on adolescent alcohol and tobacco use," *Health Economics*, 1998, *7*, 439–453.

Oreopoulos, Philip, "The long-run consequences of growing up in a poor neighborhood," *Quarterly Journal of Economics*, November 2003, *118* (4), 1533–1575.

Sacerdote, Bruce, "Peer effects with random assignment: Results for Dartmouth roommates," *Quarterly Journal of*

Economics, May 2001, *116* (2), 681–704.
Schelling, Thomas, *Micromotives and Macrobehavior*, W.W. Norton, 1978.
Tamer, Elie T., "Empirical strategies for estimating discrete games with multiple equilibria," Working Paper, Princeton University 2002.
___ , "Incomplete simultaneous discrete response model with multiple equilibria," *Review of Economic Studies*, 2002. Forthcoming.
Topa, Giorgio, "Social Interactions, Local Spillovers, and Unemployment," *Review of Economic Studies*, 2001, *68*, 261–295.

Figure 1: Equilibrium group average choice ($\bar{y}$) as a function of group characteristics ($\bar{x} + \bar{\epsilon}$).

Figure 2: Equilibrium group average choice ($\bar{y}$) as a function of group characteristics ($\bar{x} + \bar{\epsilon}$).