Binary Choice Models with Social Interactions under Heterogeneous Rational Expectations

by

Lung-fei Lee[a]*, Ji Li[b], Xu Lin[c]

[a] Ohio State University, Columbus, Ohio, USA

[b] Capital One, Richmond, Virginia, USA

[c] Tsinghua University, Beijing, CHINA

Date: September, 2007

Abstract

This paper extends Brock and Durlauf's models on binary choice with social interactions. In the extended model, an individual will form expected behaviors of peers taking into account their characteristics. The expected behaviors of peers in a group are heterogeneous. The expectations of peers are determined as rational expectations. We consider models in both group and network settings. Endogenous interaction, contextual and correlated effects are identifiable. Conditions under which an unique equilibrium exits are presented. In the group setting, multiple equilibria can be characterized by an aggregate scalar index.

As an empirical application, an adolescent's smoking behavior is studied. The results show there exists significant endogenous peer effect, and many contextual variables are significant in influencing an adolescent's smoking behavior.

JEL classification: C21, C25, I10, J13

Keywords: Discrete choice, social interaction, rational expectation, heterogeneity, smoking, teenagers

* Lee appreciates receiving financial support for his research from NSF under grant number SES-0519204.

1. Introduction

This paper extends what has been established by Brock and Durlauf (2001a, 2001b) (BD) in modeling binary choice under social interactions with rational (model consistent) expectations. Following Manski (1993), BD's model is specified in a group setting where individuals in a group may be interacted with each other. In BD, it is assumed that all members in the same reference group hold identical expectations regarding group members' behaviors. So BD's model can be viewed as one with homogeneous expectations. This paper extends their model by considering heterogeneous expectations. As the characteristics of peers relevant in their choices are observable, one may take into account such information in his/her expectations about peers' choice probabilities. This is relevant if the number of peers of an individual is not large.[1] In a network setting, this will be an appropriate specification as each of the peers of an individual has their own friends. In a network, the structure of each node may have a form of heterogeneity in itself.

In the BD's model, the group expectation is specified to be model consistent. The expectation of a group is unobservable but will be determined as a rational expectation equilibrium (RE) from the equilibrium equation. Brock and Durlauf (2001a, 2001b) provide detailed analysis on the existence and uniqueness of the RE in the logit binary choice specification. Depending on group characteristics and parameter values, the RE may be unique. But under situations where social interactions are strong, multiple RE's, which are usually three, may exist. In the binary choice BD model, the equation which characterizes the RE is a single equilibrium equation (for each group) and RE's can be determined as the points where the equilibrium equation crosses the axis. For given group characteristics, at each specific value of the parameter vector, the RE can be (numerically) evaluated. This device is valuable for empirical studies as the RE of each group needs to be evaluated as an explanatory variable in an estimation procedure. In the presence of heterogeneous expectations, a RE is no longer a scalar but a vector characterized by a system of equilibrium equations. For each group, the equilibrium equations depend on the group characteristics as well as characteristics of each of its members; and the number of the set of equations is the same as the group size. For a system of equations, the issues of equilibria and their computation will be more complicated. For both the group and network settings, we have established conditions for the uniqueness of the RE and show that the system of equations is a contraction mapping. The contraction mapping property of the system is valuable as the RE can be effectively derived within an internal equation solver by the substitution method for estimation.

For the model in a group setting, we provide also a single characterization equation which can determine

[1] If a group has a large size, whether peers' expectations are homogeneous or heterogeneous will be of less concern. As the group expectation is formed as an average of peers' expectations, the resulting group expectation will be (approximately) a constant depending on group characteristics but not distribution of peer's characteristics by the law of large numbers.

the number of multiple equilibria in the model. There exists a scalar equilibrium aggregate within the single characterization equation, which characterizes the vector of RE's; and a decentralization mechanism which distributes an equilibrium aggregate to its members satisfying the rationality condition. These features are interesting, in particular, for situations where multiple equilibria exist.[2]

As an empirical application, this paper investigates adolescents' smoking behavior under peer influence. Smoking behavior, especially youth smoking behavior, has always been an appealing topic to economists (Department of Health and Human Service 1994; Becker et al., 1994). There exists substantial literature that explores the determinants of youth smoking. Aside from many important demographic characteristics, the impact of cigarette prices and other tobacco control policies on youth smoking behavior are studied by a sizeable literature, e.g., Becker (1996), Chaloupka and Grossman (1996), Gruber (2001), Gruber and Zinman (2000), and Powell et al.,(2005). Recently, as the importance of social interactions are being recognized by economists (see, e.g., Bernheim, 1994; Glaeser et al.,1996; Akerlof, 1997; Manski 2000; Sacerdote 2001; Weinberg et al., 2004, among others), many studies of adolescents' smoking behavior also make an attempt of incorporating social interactions, particularly peer interaction, into the modeling, e.g., Gaviria and Raphael (2001), Harris and López-Valcácel (2004), Soetevent and Kooreman (2007), Krauth (2006), and Powell et al., (2005). One striking feature of these studies is that most of them focus only on endogenous peer effect by assuming no exogenous (contextual) effect. This could lead to upward bias in the estimated endogenous social effect.

In this paper, we include both endogenous and exogenous peer effects and find that both of them are significant. Another distinctive feature of our empirical research is that we consider friendship network rather than school or class, as most of the studies did. We are more interested in investigating social interactions via friends' influence – a scenario of small group interactions. Our model and/or econometric methods are also different from theirs. Gaviria and Raphael (2001) specify a linear model with social interactions, by treating a dichotomous indicator as a continuous dependent variable and estimate the linear specification by a two stage least squares method. Their model specification is motivated as a spatial autoregressive model (Cliff and Ord 1973; Anselin 1988) by Case and Katz (1991). Powell et al., (2005) use a probit specification for the dichotomous dependent variable but treating the endogenous social interaction variable as the dependent variable in a normal regression model (instead of generating from the model as a RE) in order to estimate the probit equation by a two-stage probit method. Both estimation strategies have attempted to use omitted contextual variables as instrumental variables for identification and estimation. The models in Krauth (2006) and Soetevent and Kooreman (2007) are simultaneous discrete choice models,

[2] For a network model, such a scalar equilibrium aggregate does not seem to be available. So is the multiple choices model in Brock and Durlauf (2003).

which do not involve expectations as those in the Manski, Brock and Durlauf framework. It is so also for the model in Harris and López-Valcácel (2004). Observed actual decision variables are modeled directly via a game theoretical framework without uncertainty. Such models generate multiple equilibria, which are consequences of incoherent probabilities in the limited dependent variables literature (see, e.g., Heckman 1978; Maddala 1983; Tamer 2003). In order to render the models to be coherent, an arbitrary random equilibrium selection rule is adopted in estimation.[3]

This paper is organized as follows. Section 2 details the binary choice model with heterogenous expectations and provide conditions under which a unique equilibrium exists. In the group setting, the characterization of multiple equilibria via an equilibrium aggregation is revealed. Section 3 describes how empirical estimation can be conducted for this type of model. Correlation effects due to unobserved variables can also be handled in addition to the presence of endogenous and exogenous interaction effects. Section 4 presents an empirical study of adolescents' smoking behavior under social interactions. Section 5 concludes. Proofs of propositions in Section 2 are collected in an Appendix.

2. Binary Choice with Heterogenous Expectations

2.1 The Binary Choice Model

The model that we consider in this section concerns binary choice outcomes with social interaction in a group setting. Let d be a dichotomous indicator taking values 0 or 1. Conditional on exogenous variables x of all individuals of a group, the choice probability for an individual i in the group is

$$P_i = P(d_i = 1) = F\left(x_i\alpha + \lambda \frac{1}{n-1}\sum_{j\neq i}^{n} p_{ij}\right), \tag{2.1}$$

where n is the size of a group and p_{ij} is the expected probability of j's choice of $d_j = 1$ by the individual i, x_i is the vector of exogenous variables for the individual i. The x_i captures observable characteristics of the individual i as well as observed exogenous group characteristics, which may include exogenous contextual variables. This choice probability in (2.1) can be derived from the random utility maximization framework (McFadden 1974), where the utility function of an individual consists of a private component and a social component specified as a quadratic distance between an individual's choice from expected peers' choices (Brock and Durlauf, 2001b).[4] Because x_i for all i are common knowledge, they provide valuable information of which an individual might take into account in the formulation of expected choice probabilities of peers.

[3] Estimates would, in general, be inconsistent if the selection rule was not the correct one. The Monte Carlo results in Krauth (2006) have shown biased estimates but the magnitudes of biases are small or moderate.

[4] Models with social distance provide simple but rich implications on social decision, see Akerlof (1997). The utility specification with a quadratic distance between an individual's choice and those of others is a model of conformity according to Akerlof (1997).

For example, one might expect the choice probability of a male friend could be different from that of a female friend even if other characteristics are the same. Because x_j's may, in general, be different across individuals, the expected probabilities p_{ij} of different peers would also be heterogeneous. This would be intuitively appealing, in particular, for groups with small sizes as members would likely know each other very well.

Following the modeling approach of Manski (1993) and Brock and Durlauf (2001a, 2001b), we shall consider the model with rational expectations when expectations are unobservable to the econometrician.[5] While their models assume that the expected probability is constant across individuals in a group, we allow that observed x_i's are used to formulate their expectations, which can therefore be heterogeneous. Explicitly, we assume that all observed individual and group characteristics which are directly relevant to the individual's choice probabilities are in the information set.[6] Conditional on the information set, rational expectations are formed. Under the rational expectations hypothesis, $p_{ij} = P(d_j = 1)$ for all j and i because $P(d_j = 1)$ is the objective probability generated by the model. Specifically,

Rational expectation equilibrium (RE) is a vector p^*, where $p^* = (p_1^*, \cdots, p_n^*)$, such that

$$p_i^* = F\left(x_i\alpha + \lambda\frac{1}{n-1}\sum_{j\neq i}^{n} p_j^*\right), \quad i = 1, \cdots, n, \text{ given } x_1, ..., x_n. \tag{2.2}$$

In order for the model to be coherent, it is necessary that RE exists. Under the assumption that F in (2.2) is continuous, the existence of equilibrium is guaranteed by the fixed point theorem. In addition, situations where an equilibrium is unique are of particular interest. In the Manski, Brock and Durlauf's model, the RE will be a scalar $\bar{p}$ which is characterized by the equilibrium equation in the form

$$\bar{p} = E[F\left(x_i\alpha + \lambda\bar{p}\right)], \tag{2.3}$$

where the expectation operator is taken with respect to the conditional distribution of the individual specific characteristics in x, conditional on the group. Only group variables which have a direct effect on an individual decision as well as those group variables characterizing the conditional distribution of x belong to the information set in their homogenous expectation framework. Implicitly, observed individual specific characteristics do not belong to their information set. If the size of a group were large, this could be a

[5] For an empirical application of the Manski, Brock and Durlauf model for presidential voting, where subjective data on expectations are also available, Li and Lee (2006) have demonstrated that subjective data on expectations may provide valuable information in the modeling of social interactions. However, most existing data sets for social interaction studies do not have subjective expectation variables available. Manski (2004) has argued the usefulness of subjective expectations in many economic subjects.

[6] In the empirical application, we generalize the model to include the correlated effect due to an unobserved group variable. The unobserved group variable is assumed to be in the individual's information set.

reasonable assumption. In our model, we assume that observable group variables and individual specific variables are in the information set. While the homogenous expectation specification may be appropriate for groups with large sizes, our heterogeneous expectations specification will be intuitively appealing for the scenario of small group interactions.

2.2 Rational Expectation Equilibria

In this model consisting of (2.1) and (2.2), RE is a vector of individual choice probabilities of all members in a group. From its definition in (2.2), the RE is determined by a system of nonlinear equations. For the homogenous expectation model of Brock and Durlauf, as the equilibrium is a single probability for all group members, it is determined by a single nonlinear equation in (2.3). With a single nonlinear equation, equilibria can be evaluated as roots of the equation. In that way, the number of equilibria can be identified geometrically. With a system of equations, the identification of the number of equilibria becomes much more difficult. For example, there is no general result on how to determine the number of multiple equilibria in a multiple discrete choice model (Brock and Durlauf, 2003). Fortunately, for our model with a group structure, a simple characterization in terms of an aggregate quantity is available.

Proposition 1. (Characterization of Equilibrium via an Aggregate Quantity)

Let F be a continuous mapping from the real line R to $[0,1]$. Suppose that, for each i, there exists a unique continuous function $h_i(s)$ on $[0,1]$ such that, for any $s \in [0,1]$,

$$h_i(s) = F\left(x_i\alpha + \lambda(\frac{n}{n-1})s - (\frac{\lambda}{n-1})h_i(s)\right), \quad \forall i = 1, \cdots, n. \tag{2.4}$$

Consider two separate equations:

$$p_i = F\left(x_i\alpha + \lambda\frac{1}{(n-1)}\sum_{j=1,\neq i}^{n} p_j\right), \quad i = 1, \cdots, n. \tag{2.2$'$}$$

and

$$s = \frac{1}{n}\sum_{i=1}^{n} F\left(x_i\alpha + \lambda(\frac{n}{n-1})s - \frac{\lambda}{(n-1)}h_i(s)\right). \tag{2.5}$$

Then, we have the following properties:

1) There exists a vector $p^* = (p_1^*, \cdots, p_n^*)'$ which satisfies (2.2)$'$, and an s^* satisfies (2.5).

2) Let A_p be the set of solutions p^* of (2.2)$'$ and A_s be the set of solutions s^* of (2.5). Then, for each $p^* \in A_p$, there exists an $s^* \in A_s$ such that $p^* = h(s^*)$, where $h(s) = (h_1(s), \cdots h_n(s))$ from (2.4). Conversely, for each $s^* \in A_s$, there exists a $p^* \in A_p$ such that $p^* = h(s^*)$.

3) A_p is a singleton if and only if A_s is a singleton.

4) If F is a strictly increasing function, the mapping $h : A_s \to A_p$ defined implicitly in (2.4) on the restricted domain A_s is one-one and onto the range A_p.

This proposition reduces the problem of finding a solution vector $(p_1^*, \cdots, p_n^*)$ from (2.2) into finding a single scalar value s^* from (2.5). This will not only simplify the investigation of the existence of unique equilibrium but also provides a geometric way to investigate the possible number of equilibria when multiple equilibria occur.

This proposition depends on the validity of (2.4) in that a unique solution $h_i(s)$ of the equation $h_i(s) = F(x_i\alpha + \lambda(\frac{n}{n-1})s - (\frac{\lambda}{n-1})h_i(s))$ exists. The condition (2.4) can be satisfied as in the following proposition.

Proposition 2. (Allocation Mechanism)

Suppose that F is continuously differentiable with its density function f. If $\lambda = 0$ or $\lambda > -(\frac{n-1}{\max_u f(u)})$, then, for each i, there exists a unique continuous function $h_i(s)$ such that, for any s,

$$h_i(s) = F\left(x_i\alpha + \lambda(\frac{n}{n-1})s - (\frac{\lambda}{n-1})h_i(s)\right), \quad \forall\ i = 1, \cdots, n.$$

Thus, in particular, (2.4) is satisfied whenever λ is nonnegative. With a bounded $f(.)$, a range of negative values of λ will also satisfy the condition. In applications, models with nonnegative endogenous social effects (nonnegative λ) are of special interest (see, e.g., Glaeser et al., 1996).

An equilibrium value s^* satisfying (2.5) is an aggregate quantity, which is the average of rational expectations choice probabilities of individuals in a group. Eq.(2.4) can be interpreted as a disaggregate allocation mechanism which allocates the group aggregate into equilibrium choice probabilities of which one for each individual is based on the individual characteristic x_i. For any aggregate equilibrium s^*, the allocated equilibrium satisfies the individual rationality in (2.2). The allocation mechanism is valid for some situations including cases for models with any positive endogenous social interaction.

2.3 Unique Equilibrium

Now let us consider conditions on λ that may provide a unique RE for the system (2.2).

Proposition 3. (Equilibrium Uniqueness)

Under the setting (2.4) in Proposition 1, the RE solution $(p_1^*, \cdots, p_n^*)$ of

$$p_i^* = F\left(x_i\alpha + \lambda\frac{1}{n-1}\sum_{j\neq i}^{n} p_j^*\right), \quad i = 1, \cdots, n,$$

exists and is unique when $-\frac{n-1}{\max_u f(u)} < \lambda < \frac{1}{\max_u f(u)}$, where f is the corresponding density function of F.

This proposition indicates that the equilibrium can be unique when the social interaction (in terms of absolute value of λ) is weak or moderate but multiple equilibria may occur if the social interaction is rather strong. This feature on the number of equilibria is similar to that of the BD model with homogeneous expectations.

When the setting (2.4) holds, the number of equilibria of (2.2) with p^* is the number of solutions s^* of (2.5). The characterization of an equilibrium in terms of an aggregate scalar value s^* provides a useful (numerical) procedure to evaluate the number of multiple equilibria when those occur. In order to count the number of equilibria, it is sufficient to investigate the number of roots of the function $q(s)$, where

$$q(s) = \frac{1}{n}\sum_{i=1}^{n} F(u_i(s)) - s \quad \text{with} \quad u_i(s) = x_i\alpha + \lambda(\frac{n}{n-1})s - (\frac{\lambda}{n-1})h_i(s). \tag{2.6}$$

As shown in the proof of Proposition 3, $\frac{\partial q(s)}{\partial s} = (\frac{n-1}{n})\sum_{i=1}^{n} \frac{\lambda f(u_i(s))-1}{n+\lambda f(u_i(s))-1}$. The uniqueness of equilibrium occurs when $\frac{\partial q(s)}{\partial s}$ is a negative function. However, when $\lambda > \frac{1}{\max_u f(u)}$, the uniqueness of equilibrium might not be guaranteed and $q(s)$ might not be a strictly decreasing function of s. The number of equilibria will depend on the number of times that $q(s)$ crosses the horizontal axis in the interval $[0,1]$. If $q(s)$ has the shape that is initially strictly convex and then strictly concave on $[0,1]$, the number of equilibria will be either one or three. However, for our model, because

$$\frac{\partial u_i(s)}{\partial s} = \lambda(\frac{n}{n-1})[1 - \frac{1}{n}\frac{\partial h_i(s)}{\partial s}] = \frac{\lambda n}{[n-1+\lambda f(u_i(s))]}, \tag{2.7}$$

the second order derivative of $q(s)$ is

$$\frac{\partial^2 q(s)}{\partial s^2} = \lambda(n-1)\sum_{i=1}^{n} \frac{1}{[n-1+\lambda f(u_i(s))]^2}\frac{\partial f(u_i(s))}{\partial u}\frac{\partial u_i(s)}{\partial s}. \tag{2.8}$$

When λ is positive, $\frac{\partial u_i(s)}{\partial s} > 0$ for all s from (2.7) and, hence, $u_i(s)$ is a strictly increasing function of s. If $\frac{\partial f(u_i(s))}{\partial u}$ is nonnegative for all $u_i(s)$, $q(s)$ will be convex on $[0,1]$ from (2.8). Or if $\frac{\partial f(u_i(s))}{\partial u}$ is negative for all $u_i(s)$, $q(s)$ will be concave on $[0,1]$. For both cases, the equilibrium will be unique. For the probit or logit model, $\frac{\partial f(u_i(s))}{\partial u}$ is positive if $u_i(s) < 0$ but negative when $u_i(s) > 0$. When $\frac{\partial f(u_i(s))}{\partial u}$ can be positive and negative at different values of $u_i(s)$, the sign of $\frac{\partial^2 q(s)}{\partial s^2}$ is difficult to analytically determine as it may involve the sum of positive and negative terms. However, the number of roots of $q(s)$ can be numerically evaluated via $q(s)$ in (2.6), given parameter values of α and data $(x_1, \cdots, x_n)$.[7]

In the presence of multiple equilibria $p(s_1^*)$ and $p(s_2^*)$, their order is determined by the magnitudes of s_1^* and s_2^*. As $p_i(s) = F(u_i(s))$, $\frac{\partial p_i(s)}{\partial s} = f(u_i(s))\frac{\partial u_i(s)}{\partial s}$ is positive when $\lambda > 0$. In this case, with two distinct roots s_1^* and s_2^* such that $s_1^* < s_2^*$, one has $p_i(s_1^*) < p_i(s_2^*)$ for all $i = 1, \cdots, n$, i.e., the larger root s^* corresponds to high expected equilibrium probabilities $p_i(s^*)$ for all i.

2.4 Contraction of the Probabilities System

[7] For the probit model, we have done some numerical simulations. We found cases where $q(s)$ have many convex and concave segments. However, we have not yet found cases with multiple equilibria more than three. This observation seems interesting but we do not have a good interpretation for it.

Let $p = (p_1, \cdots, p_n)'$. The choice probabilities in $(2.2)'$ for all members in a group form a system, $p = H(p)$, where

$$H(p) = (F(u_1(p)), \cdots, F(u_n(p)))', \text{ with } u_i(p) = x_i\alpha + \frac{\lambda}{(n-1)}\sum_{j\neq i}^{n} p_j, \quad i = 1, \cdots, n. \tag{2.9}$$

The $H(p)$ has a unique RE vector p^* when λ is restricted to the range specified in Proposition 3. With a slightly more restricted range on possible negative values of λ, $H(p)$ turns out to be a contraction mapping. The contraction property of $H(p)$ provides a simple algorithm to compute the solution p^*. Starting with an arbitrary value $p_{(1)}$, the iterative algorithm $p_{(m+1)} = H(p_{(m)})$ will provide the solution p^* as its limiting value, i.e., $p_{(m)} \to p^*$ as $m \to \infty$.

Proposition 4. (Contraction Mapping)

When λ satisfies the condition that $|\lambda| < \frac{1}{\max_u f(u)}$, the mapping $H(p)$ in (2.9) is a contraction mapping.

For the popular model with logit probability, $F(u) = \frac{e^u}{1+e^u}$, its density function is $f(u) = F(u)[1-F(u)]$, which has its mode at $u = 0$ with its maximum value being $\frac{1}{4}$. Thus, a sufficient condition for a unique RE in the logit framework is $|\lambda| < 4$. If the model is specified with the probit framework, because the standard normal density has its mode at 0 and its maximum value being $\frac{1}{\sqrt{2\pi}}$, a sufficient condition for a unique RE in the probit model is when $|\lambda| < \sqrt{2\pi}(\approx 2.5)$.

2.5 Alternative Coding

The binary choice model considered above has adopted the conventional coding with a dichotomous indicator d having values 0 and 1. For some authors, the values -1 and 1 are coded for the dichotomous indicator, e.g., Brock and Durlauf (2001a, 2001b). For contingency table analysis, 1 and -1 are standard codes for a dichotomous indicator. With the coded indicator being y with values of $\{-1, 1\}$, Brock and Durlauf express their binary choice social interactions model in term of expected value m of y, i.e., $m = E(y)$. Their model has the choice probability for $y = 1$ as

$$P(y_i = 1) = F\left(x_i\beta + \rho\frac{1}{n-1}\sum_{j\neq i}^{n} m_{ij}\right), \tag{2.10}$$

where m_{ij} is the individual i's expected value of friend j's decision. Because $m_i = (1)p_i + (-1)(1-p_i) = 2p_i - 1$, the corresponding RE is defined as a vector $p^* = (p_1^*, \cdots, p_n^*)$ such that $p_i^* = F(x_i\beta + \rho\frac{1}{n-1}\sum_{j\neq i}^{n}(2p_j^* - 1))$, or, equivalently, a vector $m^* = (m_1^*, \cdots, m_n^*)$ such that $m_i^* = 2F\left(x_i\beta + \rho\frac{1}{n-1}\sum_{j\neq i}^{n} m_j^*\right) - 1, \quad i = 1, \cdots, n$. With the alternative coding, the characterization in Proposition 1 shall be reformulated. The setting in (2.4) will be replaced by

For each i, there exists a unique continuous function $m_i(s)$ such that, for any $s \in [-1, 1]$,

$$m_i(s) = 2F\left(x_i\beta + \rho(\frac{n}{n-1})s - \frac{\rho}{(n-1)}m_i(s)\right) - 1, \ \forall\ i = 1, ..., n. \tag*{(2.4)$'$}$$

Eq.(2.2)$'$ will be replaced by

$$m_i = 2F\left(x_i\beta + \frac{\rho}{(n-1)}\sum_{j\neq i}^{n} m_j\right) - 1, \quad \forall\ i = 1, \cdots, n. \tag{2.2$''$}$$

and (2.5) will be replaced by

$$s = \frac{2}{n}\sum_{i=1}^{n} F\left(x_i\beta + \rho(\frac{n}{n-1})s - \frac{\rho}{(n-1)}m_i(s)\right) - 1. \tag{2.5$'$}$$

As the choice probability in (2.10) is a reformulation of (2.1) and vice versa, the parameters in the two formulations are related: $x\alpha = x\beta - \rho$ and $\lambda = 2\rho$ because $m = 2p - 1$. So the sufficient conditions for the existence of unique RE and the contraction mapping property will be accordingly adjusted in terms of ρ. For the logit with

$$P(y_i = 1) = F\left(x_i\beta + \rho\frac{1}{n-1}\sum_{j\neq i}^{n} m_{ij}\right) = \frac{\exp(x_i\beta + \rho\frac{1}{n-1}\sum_{j\neq i}^{n} m_{ij})}{1 + \exp(x_i\beta + \rho\frac{1}{n-1}\sum_{j\neq i}^{n} m_{ij})}, \tag{2.11}$$

the RE is unique when $|\rho| < 2$. In Brock and Durlauf (2001b), the logit has the form

$$P(y_i = 1) = \frac{\exp(x_i\delta + \gamma\frac{1}{n-1}\sum_{j\neq i}^{n} m_{ij})}{\exp(x_i\delta + \gamma\frac{1}{n-1}\sum_{j\neq i}^{n} m_{ij}) + \exp[-(x_i\delta + \gamma\frac{1}{n-1}\sum_{j\neq i}^{n} m_{ij})]}. \tag{2.12}$$

The formulation (2.12) is related to (2.11) with $\beta = 2\delta$ and $\rho = 2\gamma$. So for the specification in (2.12), the RE is unique when $|\gamma| < 1$, which is the same sufficient condition in Brock and Durlauf (2001b) for their homogeneous expectation model. The formulation of heterogeneous expectations in the logit social interaction model does not impose a more restrictive condition than that of the homogeneous model.

2.6 Extension: The Network Model and Contraction

Consider the binary choice model in which individuals may or may not be in a group setting. An individual i may have his/her friends who have their own ones. In this case, friends do not necessarily form a close group and the friendship network has a tree (or spatial) structure. We call it a network model.

Let n be the number of connected nodes of a network. For the individual i in the network, let $w_i = (w_{i1}, \cdots, w_{in})$ be the weights on other individuals. The weights will be positive for friends and zero otherwise. A typical specification may assume that each friend of an individual receives equal weight and the weights may be row-normalized so that the row sum is a unit. In its generality, the weights may be different if the influences of different friends are not the same. The use of weights is motivated by the literature of spatial models.[8] Then, the choice probability of alternative 1 by individual i has

$$p_i = F(x_i\alpha + \lambda w_i p), \quad i = 1, \cdots, n, \tag{2.13}$$

[8] In Case (1991), Moffitt (2001) and Lee (2007), spatial autoregressive models are suggested to formulate linear-in-mean social interaction models. Contrary to Manski's linear-in-mean model, when the group sizes vary across groups, identification of endogenous and exogenous effects are possible as shown in Lee (2007).

where $p = (p_1, \cdots, p_n)'$ is a column vector of dimension n. As a system, $p = H(p)$, where $H(p) = (F(u_1(p)), \cdots, F(u_n(p)))'$ with $u_i(p) = x_i\alpha + \lambda w_i p$. The RE is a solution of this system under the assumption that x_i's are in individuals' information sets. For the network model, heterogeneous rational expectations seem natural as each friend of an individual can be special.

For the network model, the existence of RE vector of choice probabilities will not be summarized by an aggregate scalar quantity s^*. Although the existence of RE can still be guaranteed by the fixed point theorem, simple ways to count the number of equilibria remain difficult to find. Situations with a unique RE and the contraction property will be of special interest. The following paragraph establishes sufficient conditions for the contraction mapping property of the network model in (2.13).

The gradient of $H(p)$ is

$$\frac{\partial H(p)}{\partial p'} = \begin{pmatrix} \lambda w_{11} f(u_1(p)) & \cdots & \lambda w_{1n} f(u_1(p)) \\ \vdots & \vdots & \vdots \\ \lambda w_{n1} f(u_n(p)) & \cdots & \lambda w_{nn} f(u_n(p)) \end{pmatrix}.$$

Let $||\cdot||_\infty$ be the row sum matrix norm. It follows that

$$|| \frac{\partial H(p)}{\partial p'} ||_\infty = |\lambda| \max_{i=1,\cdots,n} \sum_{j=1}^{n} |w_{ij}| \cdot f(u_i(p)) \leq \lambda ||W||_\infty \max_u f(u).$$

In the case where $w_{ij} \geq 0$ and the matrix W is row normalized, i.e., $\sum_{j=1}^n w_{ij} = 1$ when $w_{ij} \neq 0$ for some j, $||W||_\infty = 1$. For the row normalized case, $H(p)$ is a contraction mapping under the condition $|\lambda| < \frac{1}{\max_u f(u)}$. This generalizes the result in Proposition 4 from the group setting to the network setting.

3. Estimation Method

In our subsequent empirical application, F is specified to be the logistic distribution. The model can be estimated by the method of maximum likelihood (ML) taking into account the endogenous formation of rational expectations.

Suppose that the number of individuals in a network is n. The sample of the network consists of their actual choices $Y = (y_1, ...y_j, ..., y_n)'$, their exogenous characteristics $X = (x_1', ..., x_j', ..., x_n')'$ with x_j being a $1 \times k$ vector, and an $n \times n$ row-normalized weights matrix $W = (w_1', \ldots, w_j', \cdots w_n')'$ with zeros along the diagonal. The individual i's choice probability incorporating both endogenous and exogenous social interactions is

$$\begin{aligned} P(y_i = 1) &= \frac{1}{1 + \exp\left[-2(x_i\alpha + w_i X\delta + \lambda w_i M)\right]} \\ P(y_i = -1) &= \frac{1}{1 + \exp\left[2(x_i\alpha + w_i X\delta + \lambda w_i M)\right]}, \end{aligned} \tag{3.1}$$

where the n-dimensional column vector M is the solution to the nonlinear equation system:

$$M = (m_1, \cdots, m_j, \cdots, m_n)' = \tanh(X\alpha + WX\delta + \lambda WM). \tag{3.2}$$

Eq.(3.2) is built on the rational expectation assumption with X and W in an individual's information set. The logit model in (3.1) has adopted the parameterization as in (2.12). We can clearly see the heterogeneity of rational expectations from Eq.(3.2).

The parameter vector to be estimated is $\eta = (\alpha', \delta', \lambda)'$, in which δ measures the so-called "contextual social effects" as $w_i X$ can be regarded as "contextual" variables, and λ measures "endogenous social effect". The log likelihood function is

$$\ln L(\eta; Y|X, W) = \sum_{i=1}^{n} \left\{ \frac{1+y_i}{2} \ln[P(y_i = 1)] + \frac{1-y_i}{2} \ln[P(y_i = -1)] \right\}. \tag{3.3}$$

The parameters can be estimated by the ML method with an internal subroutine that solves the RE via the system of equations in (3.2).[9] The estimation method is iterative. We start with an initial guess of the parameter values and solve M based on (3.2), we then conduct the ML interaction by substituting M into the log likelihood function. The ML interaction gives an updated estimate of the parameter vector, which is used to update M, and conduct ML iteration again. This process continues until it converges.

The asymptotic variance of the MLE can be estimated by

$$J^{-1} = \left(\frac{1}{N} \sum_{i=1}^{N} \frac{\partial f_i}{\partial \eta} \frac{\partial f_i}{\partial \eta'} \right)^{-1}$$

where, due to the logit form in (3.1),

$$\frac{\partial f_i}{\partial \eta'} = 2 \left(\frac{1+y_i}{2} - P(y_i = 1) \right) \left(\begin{pmatrix} x_i & w_i X & w_i M \end{pmatrix} + \lambda w_i \frac{\partial M}{\partial \eta'} \right). \tag{3.4}$$

The last part in (3.4) appears because of the dependence of M on the parameters. Based on (3.2), the implicit function theorem implies that

$$\frac{\partial M}{\partial \eta'} = \left(I_n - \lambda \frac{\partial \tanh(T)}{\partial T'} W \right)^{-1} \frac{\partial \tanh(T)}{\partial T'} \begin{pmatrix} X & WX & WM \end{pmatrix}$$

where I_n is the $n \times n$ identity matrix,

$$T = X\alpha + WX\delta + \lambda WM = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix} = \begin{pmatrix} x_1\alpha + w_1 X\delta + \lambda w_1 M \\ x_2\alpha + w_2 X\delta + \lambda w_2 M \\ \vdots \\ x_n\alpha + w_n X\delta + \lambda w_n M \end{pmatrix},$$

[9] In the event that a large network can be partitioned into sub-networks such that nodes within a sub-network are linked together but not across another sub-network, then each sub-network will formulate a separate equation system. One may separately solve each of the equation systems instead of pooling them together for computational simplicity. This is what has been done in our empirical application where each sub-network corresponds to a grade for students in a school.

and

$$\frac{\partial \tanh(T)}{\partial T'} = \begin{pmatrix} \frac{\partial \tanh(t_1)}{\partial t_1} & 0 & \cdots & 0 \\ 0 & \frac{\partial \tanh(t_2)}{\partial t_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial \tanh(t_n)}{\partial t_n} \end{pmatrix}_{n \times n}.$$

4. An Empirical Application to Adolescents' Smoking Behaviors

Many studies in the sociology literature, including Duncan et al., (2001) and Haynie (2002), analyze the influence of friends on various adolescents' behaviors and outcomes. As an important context for providing information flow, shaping social norms, social acceptances, social expectations, and so on, friendship networks are believed to play a key role in an adolescent's development. These studies show that adolescents may be significantly influenced by their friends. Economists have also recognized the importance of friends. Becker (1996) mentions that adolescents are influenced by their friends' bad behaviors, which illustrates his social capital theory (Becker, 1974). As another example, in a recent study, Lin (2007) finds that an adolescent's academic achievement is significantly affected by their friends, both endogenously and exogenously, even after controlling for the unobserved correlated effects.

In this paper, we utilize the National Longitudinal Study of Adolescent Health (Add Health) survey (Bearman et al., 1997) to conduct our analysis of adolescents' smoking choices under peer influence. The Add Health data set allows us to build relatively informal peer groups, friendship networks, and study how such informal but more intimate peer groups influence adolescents' behaviors.

4.1 Data

4.1.1 Add Health Survey

The Add Health survey is a nationally representative study that explores the health-related behaviors of adolescents in grades 7 through 12. Particularly, the survey Health seeks to examine how social contexts (families, friends, peers, schools, neighborhoods, and communities) influence adolescents' health and risk behaviors. With such an objective, the Add Health survey covers detailed information about respondents' demographics, family background, academic outcomes, health related behaviors, as well as their social networks.

The Add Health Wave I in-school survey was conducted from September 1994 to April 1995, where an in-school questionnaire was given to all the students attending the sampled schools, resulting in a total sample of over 90,000 students.[10] The questions covered the respondent's demographics, family background, academic

[10] Here is the sample design for the Add Health survey: "A sample of 80 high schools and 52 middle schools from the US was selected with unequal probability of selection. Incorporating systematic sampling methods and implicit stratification into the Add Health study design ensured this sample is representative of US schools with respect to region of country, urbanicity, school size, school type, and ethnicity".

outcomes, health related behaviors including smoking, drinking and drug use, and teenage pregnancy, etc. The respondents were asked to nominate up to five male and five female friends, along with linkable friend identification numbers, which makes it possible to link a respondent's information to her friends'. This survey covered all the students in the sampled schools, meaning that a respondent's friends are also likely to be in the sample.[11] We construct a complete friendship network for our study from the samples in this in-school survey.

4.1.2 Reference Group

Most of the existing literature that studies adolescents' behaviors define the sphere of interaction formally, largely due to data availability. For example, in studying adolescents' smoking and other health-related behaviors, both Gaviria et al.,(2001), and Powell et al.,(2005) use school as the relevant sphere of social interactions. In this study, we specify the sphere of social interactions at the level of friendship networks within a school-grade. We define a "group" as a school-grade, which provides the universe of a network. Respondents are in the same group if they are in the same grade level of the same school. A respondent's *friends* are identified as her "*peers*" only when they are in the same group. The peers of an individual are given equal weight, while all the others in the same group are assigned zero weight. The jth entry, w_{ij}, of the weights vector, w_i, for the individual i will be $\frac{1}{n_i}$ if respondent i identifies person j as one of her n_i friends in her grade and zero otherwise.[12]

4.1.3 Sample Summary Statistics

For the set of individual characteristics, given data availability, we use age, years in school, gender, race, family structure, participation in sports, mother's education and mother's occupation as explanatory variables. The sample consists of 75,586 observations. Table 1 provides the summaries and definitions of the variables used in this study. Contextual characteristics are the same variables averaging over named friends. This is the standard practice in the social interaction studies (e.g., Manski, 1993; Brock and

[11] This is classified as Wave I in the Add Health. Three waves of data are available in the Add Health. About 14,000 adolescents who participated in the in-home survey were interviewed again one year later in 1996, which consists of the Wave II in-home survey. From August 2001 to April 2002, all the Wave I respondents who could be located were interviewed for the Wave III in-home survey, which contains a sample size of over 15,000. For our purpose, the desired data set is the Wave I in-school survey.

[12] This is a simple but popular specification in social interaction models. It could be that not every friend would be the same. There are preliminary evidences in our investigation with our colleague, Bruce Weinberg, that popularity of a friend, and even the order of friendship nomination, etc., might be informative about the importance of a friend. People interact (talk over phone, visit each other's house) more with people who are listed earlier in the friendship nominations, and as a person's friend becomes more popular, they do more with that person than they do of a less popular friend. However, a systematic way to incorporate this information into the construction of weights remains to be seen.

Durlauf, 2001a). The last column in Table 1 reports the correlations between own characteristics and their corresponding contextual characteristics.

The dependent variable is a dichotomous indicator that takes 1 if the respondent is a smoker, and -1 if the respondent is not.[13] By such a specification, we are adopting the alternative coding following Brock and Durlauf as explained earlier. On average, the respondents are about 15 years old, with about 2.5 years in the current school. There are slightly more females than males in the sample and more than 55% are white. Hispanic, Black, Asian account for 13.8%, 18.2% and 6.6%, respectively. The rest 5.6% are of other race. 73.1% of the sample live in an intact family with both parents. 52.6% of the respondents participate in at least one sports club. As to the respondent's mother's highest education level, about 38% of them achieved high school, 41.2% beyond high school, and 10% less than high school. 10.7% of the respondents failed to report their mothers' highest education. We put in a dummy variable for this.[14] As for the mother's occupation, we group the occupations into 4 general categories, along with a missing indicator as in Lin (2007). In the sample, 26.2% of the respondents' mothers work at professional occupations, including teachers, doctors, lawyers and executives. 28.2% of the mothers are homemakers or do not work. 35.8% of the mothers hold other jobs. Only a few (less than 1%) mothers are reported to be welfare recipients. As we can see from Table 1, the means of the friends' mean characteristics are somewhat similar to those of own characteristics. And the correlations between own characteristics and the corresponding contextual characteristics vary from small to moderate, with that for the variable "black" being the largest (0.665).[15]

Finally, we have 584 school-grade groups. The mean size of group is 129.43 and the median is 112 with the minimum being 1 and the maximum being 560. For number of named friends, its mean is 2.86, median is 2, maximum is 10 and minimum is 0.[16] Note that among 75,586 respondents in the sample, 19,546 named no friends.

4.2 Model Specifications and Parameter Estimation

In this section, we shall present our models for the empirical study, with detailed presentation of the

[13] An individual is coded as a non-smoker if she never smoked or just smoked once or twice during the past 12 months.

[14] This is effectively an imputation method for handling missing data. For the justification, see Lin (2007).

[15] Such correlations might reflect the sorting of friends. With the inclusion of such contextual variables (both observed and unobserved ones), it might ease the endogeneity issue of friendship network in estimation.

[16] There are about 30% of respondents who have named the maximum five female or male friends in a school-grade. This would be a censoring on the number of friends in the survey design. Some preliminary simulation study on spatial autoregressive models has indicated that there could be some downward biases on the estimate of the endogenous interaction effect. We do not expect that the biases would be large. There are also empirical evidences on small to moderate biases in estimates due to censoring of group size in a group setting (Li and Lee 2006). In the group setting, one can restrict the sample to groups with complete group sizes. However, in a network model, this can not be done.

model with an unobserved group variable. The unobserved group variable may capture possible unobserved correlation effect of the group. Other specifications can be viewed as special cases with constrained parameters. We present the estimation results in Tables 2 and 3.

Recall that we define a group (a sub-network) as a school-grade, so a scalar random variable u_g is introduced in order to capture school-grade specific unobserved characteristics such as teacher's emphasis on discipline. Suppose we have a sample of G groups and n_g individuals in group g. For group g, we observe members' actual choices $Y_g = (y_{g1}, ...y_{gi}, ..., y_{gn_g})'$ where $y_{gi} = 1$ if adolescent i from group g smokes and $y_{gi} = -1$ otherwise, their exogenous characteristics $X_g = (x'_{g1}, ..., x'_{gi}, ..., x'_{gn_g})'$ with x_{gi} being a $1 \times k$ vector, and a $n_g \times n_g$ row-normalized weights matrix $W_g = (w'_{g1}, \ldots, w'_{gi}, \cdots w'_{gn_g})'$ that defines the network structure. We assume the unobserved group variable follows the normal distribution with zero mean and standard deviation σ. Denote $u_g \sim N(0,1)$ and σu_g would capture the unobserved (group) correlation effect. The choice probabilities of individual i in group g are modeled as:

$$\begin{aligned} P(y_{gi} = 1|u_g) &= \frac{1}{1 + \exp\left[-2(x_{gi}\alpha + w_{gi}X_g\delta + \lambda w_{gi}M_g + \sigma u_g)\right]}, \\ P(y_{gi} = -1|u_g) &= \frac{1}{1 + \exp\left[2(x_{gi}\alpha + w_{gi}X_g\delta + \lambda w_{gi}M_g + \sigma u_g)\right]}. \end{aligned} \tag{4.1}$$

In addition to the observed variables X_g and the weights W_g, we assume that u_g is also in the information set of an individual in forming expectations. The $n_g \times 1$ vector M_g is, thus, the solution to the following nonlinear equation system under our rational expectation assumption:

$$M_g = \tanh(X_g\alpha + W_gX_g\delta + \lambda W_gM_g + \sigma u_g l_g) \tag{4.2}$$

with l_g a $n_g \times 1$ vector with 1 in each component.

The parameters to be estimated are $(\alpha, \delta, \lambda, \sigma)$. In the presence of u_g, this model (specification 5) will be estimated by a simulated ML method (see, e.g., Gouriéroux and Monfort 1996). We simulate S (in the empirical estimation, $S = 100$) independent draws $u_g^{(s)}$ from the standard normal distribution.[17] The simulated probabilities are:

$$\begin{aligned} \widehat{P}(y_{gi} = 1) &= \sum_{s=1}^{S} \frac{1}{1 + \exp\left[-2(x_{gi}\alpha + w_{gi}X_g\delta + \lambda w_{gi}M_g^{(s)} + \sigma u_g^{(s)})\right]}, \\ \widehat{P}(y_{gi} = -1) &= \sum_{s=1}^{S} \frac{1}{1 + \exp\left[2(x_{gi}\alpha + w_{gi}X_g\delta + \lambda w_{gi}M_g^{(s)} + \sigma u_g^{(s)})\right]}, \end{aligned} \tag{4.3}$$

where $M_g^{(s)}$ is the corresponding solution from (4.2) with u_g^s, and the simulated likelihood function is

$$\begin{aligned} &\ln L\left(\alpha, \delta, \lambda, \sigma; \{Y_g\}_{g=1,\ldots,G} \,|\, \{X_g\}_{g=1,\ldots,G}, \{W_g\}_{g=1,\ldots,G}\right) \\ &= \sum_{g=1}^{G}\sum_{i=1}^{n_g}\left\{\frac{1+y_{gi}}{2}\ln\left[\widehat{P}(y_{gi}=1)\right] + \frac{1-y_{gi}}{2}\ln\left[\widehat{P}(y_{gi}=-1)\right]\right\}. \end{aligned} \tag{4.4}$$

[17] In our experience, simulation draws of 100 are sufficient for most practical purposes.

Note, in estimation, we reparameterize σ as $\exp(\varpi)$ where ϖ is unconstrained to ensure that σ is nonnegative.

Setting one or several parameters to be zeros generates other model specifications. Setting $[\delta', \lambda, \sigma] = 0$ gives the model with no social effects (spec1 hereafter); setting $[\delta', \sigma] = 0$ gives the model with only an endogenous social effect (spec2 hereafter); setting $[\lambda, \sigma] = 0$ gives the model with contextual effects (spec3 hereafter); and finally setting $\sigma = 0$ gives the model with both endogenous and exogenous social effects but no unobserved correlated effect (spec4 hereafter). These constraints change the probability functions, the nonlinear implicit function of M_g and the log-likelihood function straightforwardly, so we will not present them here. With $\sigma = 0$, all four models are estimated by the ML method as simulation is no longer needed.

Parameter estimates for these five model specifications are reported in Tables 2 and 3. Table 2 reports parameters estimation for the models with no social effects (spec1) and endogenous social effect only (spec2). Table 3 reports parameter estimates for the models with the contextual social effect only (spec3), with both social effects (spec4), and with additional unobserved correlated effect (spec5). Since the models are nonlinear, explanations of estimated parameter values are not straightforward as in the linear model, and marginal effects have been emphasized instead. In Section 4.3, we detail how marginal effects are calculated incorporating social interactions and, in Section 4.4, we digest the estimation results, combining parameter estimates and marginal effects.

4.3 Calculation of Marginal Effects

With the presence of social interactions, any change of a particular individual's characteristic has multiple impacts. First, it naturally alters the individual's own behavior, as traditional economic models postulate. Second, it also changes the contextual characteristics of others who have this individual as a friend, and thus changes their behavior (through the so-called "contextual effect"). Last, it generally changes the group's RE because it depends on all members' exogenous characteristics in the network, and thus influences the behavior of all (through the so-called "endogenous effect").

These lead us to consider marginal effects in the network scenario rather than just looking at one representative individual. We start with choosing a "representative network". We choose among 584 groups in our data a group whose mean characteristics are close to the mean characteristics of the whole sample of 75,586 observations. The mean characteristics of the group is reported in Table 4. The group has 121 members. It would be uninformative and even misleading to base the measurement of marginal effects on the characteristics of one single member, as the calculation depends on this individual's characteristics and the network structure. It does not help much even if we assume this member's characteristics are all at their population means, because we are still left unsure about the network structure of this "typical" individual, and it would be hard to designate a "representative" network structure (how many friends and who, friends of how many and whom) to him or her. To circumvent such inconveniences, we calculate the marginal effects multiple times with different sets of characteristics and use the *average* of these calculations. We choose 10% of the group size (12 individuals) whose characteristics are diverse enough to alleviate the dependence of the calculation of marginal effects on any single set of characteristics.

We consider two ways of calculating marginal effects. Naive estimation calculates the marginal effects in the usual way (see Eq.(4.5) below, or for example, Greene (2000), 19.4.2), assuming changes of contextual variables and friends' expected average choices are exogenous. For the marginal effects of changes in an individual's characteristics, the proper (sophisticated) way may take into account the equilibrium nature of the system. The changes in one's *personal* characteristics will change his or her choice probability and probabilities of the rest of the group members due to social interactions. Both methods bear some meanings. The native estimation considers the changes of the (smoking) environment, or the changes of personal characteristics without an environment change. The sophisticated way is to consider the equilibrium effect of a marginal change in the personal characteristics on the system. To make it more clear, for the more sophisticated calculation, we report separately the marginal effects on those whose characteristics change ("changer"), and on those whose own characteristics do not change but whose choice probabilities do change due to social interactions ("affected").

Consider a general model with both contextual and endogenous social effects.[18] For naive calculation, we simply take the first derivative of all regressors as in (4.5), regardless of their exogeneity. Denote $\eta = (\alpha', \delta', \lambda)'$ and $r_{gi} \equiv \begin{pmatrix} x_{gi} & w_{gi}X_g & w_{gi}M_g \end{pmatrix}$, we have

$$\frac{\partial P(y_{gi}=1)}{\partial r'_{gi}} = 2P(y_{gi}=1)\left(1-P(y_{gi}=1)\right)\eta. \tag{4.5}$$

In the more sophisticated way, if the κ^{th} exogenous explanatory variable is continuous (for example, "age" or "years in school"), its marginal effect is

$$\begin{aligned} \frac{\partial P(y_{gi}=1)}{\partial x_{gi\kappa}} &= 2P(y_{gi}=1)\left(1-P(y_{gi}=1)\right)\left(\alpha_k + \lambda w_{gi}\frac{\partial M}{\partial x_{gi\kappa}}\right) \\ \frac{\partial P(y_{gj}=1)}{\partial x_{gi\kappa}} &= 2P(y_{gj}=1)\left(1-P(y_{gj}=1)\right)\left(w_{gj}\frac{\partial X_g}{\partial x_{gi\kappa}}\delta + \lambda w_{gj}\frac{\partial M_g}{\partial x_{gi\kappa}}\right) \end{aligned} \tag{4.6}$$

where

$$\frac{\partial M_g}{\partial x_{gi\kappa}} = \left(I_{N_g} - \lambda\frac{\partial \tanh(T)}{\partial T'}W_g\right)^{-1}\frac{\partial \tanh(T)}{\partial T'}\left(\frac{\partial X_g}{\partial x_{gi\kappa}}\alpha + W\frac{\partial X_g}{\partial x_{gi\kappa}}\delta\right)$$

based on Eq.(4.2) (with constraint $\sigma = 0$). $\frac{\partial X_g}{\partial x_{gik}}$ is the $N \times k$ matrix with 1 in (i, κ) entry and zeros elsewhere. If the κ^{th} exogenous explanatory variable is discrete, its marginal effect is the difference between

[18] Note because spec.5 shows no significant unobserved group effect, we do not calculate marginal effects from it. Note that some studies, such as Lin (2007), and Soetevent and Kooreman (2007), introduce a group fixed effect term (instead of a random effect term as ours) to capture the unobserved group effects and find that it changes the estimation results significantly.

choice probabilities with $x_{gi\kappa} = 1$ vs $x_{gi\kappa} = 0$; i.e., $P(y_l = 1|x_{gi\kappa} = 1, M(x_{gi\kappa} = 1)) - P(y_l = 1|x_{gi\kappa} = 0, M(x_{gi\kappa} = 0))$, $l = i$ or j. Note again this change of x_{gik} may change contextual characteristics of some and in general changes rational expectation equilibrium of all.[19]

The calculated marginal effects are reported in Table 5 (for spec1 and spec2) and Table 6 (for spec3 and spec4). We have several general observations. First, for the spec1 model (no social effects), one's choice probability is solely the function of his or her own characteristics, so the naive estimation is identical to the sophisticated one, and because there is no social interaction, there are no so-called "affected". For the spec2 model (endogenous social effect only), marginal effects for "own characteristics" are different between naive and sophisticated calculations because additional effects are channeled through the *endogenous* social interaction. For the spec3 model (contextual social effect), marginal effects of "own characteristics" are identical in naive and sophisticated calculations for the individual 'changer', because an individual's "context" does not include himself/herself. Last, for the spec4 model, both contextual and endogenous effects are in the model so again we have different calculated marginal effects. One last note is that Tables 5 and 6 report marginal effects in percentage points.

4.4 Summary of Results

For the spec1 and spec2 models in Table 2, we find that almost all explanatory variables are significant at 5% or 1% significance level, except "Male" in spec2. Also, for most explanatory variables, parameter estimates in spec2 are smaller than those in spec1, indicating upward biases in spec1's estimates. Age appears nonlinearly in the models, its effect on the probability of smoking increases first and then decreases at age=17 or so. Races also have significant effects. Adolescents from minority racial groups such as African American, Asian and Hispanic have lower probabilities of smoking than White adolescents, although those from other racial groups such as American Indian appear to be more likely to smoke. Living with both parents lowers an adolescent's probability of smoking. Participating in sports club also has a positive impact in lowering the probability. The mother's educational level has a significant impact on an adolescent's smoking behavior. Those with less educated mothers tend to be more likely to smoke. Comparing with those who have stay-at-home moms, adolescents who have working moms are more likely to smoke. These results on own characteristics are largely consistent with existing literature, including Case and Katz (1991), Gaviria and Rapheal (2001), Gruber (2001), among others, who also find significant difference across race, gender, age and family background on youth smoking and/or other related behaviors. The spec2 model identifies a highly significant endogenous effect, indicating that an adolescent's smoking behavior is significantly influenced by

[19] The formulas of both naive and sophisticated calculations of marginal effects are derived based on the logistic distribution we have assumed.

their friends' expected smoking choices.[20] Many previous studies also find significant endogenous peer effects on a youth's cigarette smoking behavior (e.g., Gaviria and Rapheal, 2001, and Powell et al., 2005).

Table 5, which reports marginal effects based on parameter estimates from Table 2, provides a more straightforward quantification. The first column based on the spec1 model shows that, with other factors being the same, being Black or Asian or Hispanic reduces the probabilities of smoking by at least 8.2 percentage points and as high as 20.5 percentage points. Living with both parents or participating in sports club also generate positive impacts by reducing the probabilities of smoking, respectively, by 5.8 percentage points and 4.9 percentage points. On the other hand, being 1 year older means having a 3.8 percentage points higher probability of smoking. Having mom on welfare increases this probability by 4.0 percentage points.

For the spec2 model, the naive calculation shows that an adolescent's smoking probability increases by 12.5 percentage points as the average of his/her friends' expected smoking choices increases by 1 (for example, moving an adolescent from a group where all friends are expected to be non-smokers to a group where half of friends are expected to smoke.). This value is larger than Gaviria and Raphael's finding (Garvia and Raphael, 2001) of 7.9 percentage points, but much smaller than Powell et al.'s (2005) finding of 28.9 percentage points.[21] Sophisticated calculation only considers effects of *exogenous* characteristics, and most of the calculated marginal effects are slightly larger in absolute value than those of the naive calculation. In addition, through social interactions, an exogenous change of one's characteristics influences other group members' choice probabilities. The last column illustrates the effects on those "affected". They are generally small, with the largest being 0.04 percentage points.[22]

Tables 3 and 6 are the coefficient estimates of the other three specifications and corresponding calculations of marginal effects. To summarize, all estimated coefficients of own characteristics in all three specifications are significant at least at the 10% level except "years in school". Signs of these coefficients are the same as in previous two specifications and values are quantitatively similar. But it does appear that the absolute values of these estimated parameters tend to get smaller as the specification changes from spec1 to spec2/spec3 to spec4.

Specifically, the estimates of the spec3 model shows that many coefficients of contextual variables are

---

[20] In the estimation, the coefficient of the endogenous interaction effect has not been restricted. As the estimated coefficient is less than one in absolute value, the RE is unique. So the estimated endogenous effect is significant but not large enough to generate multiple equilibria.

[21] Krauth (2006) suggests most previous studies (e.g., Wang et al., 1995; Lloyd-Richardson et al., 2002; and Norton et al., 2003) may have dramatically overstated the peer endogenous effect.

[22] Even though the latter magnitude seems small, one should keep in mind that it represents, on average, the change of a 'single' individual's characteristics on another individual's smoking behavior in a group with 121 members.

significant and their signs are consistent with those of corresponding own characteristics. The friends' average age, average squared age and gender composition are insignificant. The friends' mothers' welfare condition doesn't appear significant either. On the other hand, the friends' racial composition, percentage who live with both parents and percentage who participate in sports are significant. The friends' average family background in terms of their mothers' educational levels and working status is also significant. The estimation of spec4 also gives interesting results. First, we have a larger estimate of λ than in the model with endogenous effect only. Contextual effects are different from those of spec3. On one hand, the friends' average age, average squared age and gender composition are significant. On the other hand, the friends' racial composition appears to be a lot less influential (in terms of the values of the estimated parameters) and particularly, the percentage of friends who are black becomes insignificant. The friends' mothers' working status also becomes less significant. For the model with unobserved group effect, the estimates are almost identical to those of spec4, and we cannot reject the null that $\sigma = 0$.

Table 6 reports the calculated marginal effects based on spec3 and spec4. For own characteristics, naive calculations of marginal effects for spec3 and spec4 are quantitatively similar to those of spec1 and spec2. Note that for spec3, sophisticated and naive calculations generate identical marginal effects for own characteristics, due to reasons mentioned above.

For the spec3 model, naive calculations of contextual variables' marginal effects show that other things being the same, having no black (Asian, Hispanic) friends means 8.1 (9.0, 8.6) percentage points higher probability of smoking than having all black (Asian, Hispanic) friends. Comparing to having no friends living with both parents, having all friends living with both parents reduces the probability of smoking by 7.7 percentage points. For those "affected" due to social interactions, the average marginal effects are again fairly small. The largest ones are racial variables (Black, Asian and Hispanic) with value of -0.05 percentage points.

For the spec4 model, the naive calculation shows that an adolescent's smoking probability increases by 22.9 percentage points as his/her friends' average expected smoking choices increases by 1 (for example, from a group where all friends are expected to be non-smokers to a group where half are expected to smoke). Sophisticated calculations give generally larger marginal effects for own characteristics than naive estimation does. For those "affected", the marginal effects are generally larger than those calculated based on the other two specifications (spec2 and spec3). The largest values are "living with both parents" and "Asian" (0.07 percentage points).

5. Conclusion

This paper extends Brock and Durlauf's model on binary choice with social interactions under rational expectations (Brock and Durlauf, 2001a, 2001b). The Brock and Durlauf model is a model with homogeneous rational expectations in a group setting. In their model, peers' specific observed characteristics are not in an individual's information set but only observed group characteristics are. This homogeneous assumption may be meaningful for scenarios of large group interactions. But for a small group interactions case, individuals may know each other well and may take into account individual specific characteristics in their information set in the formation of expectations on peers' behaviors. Expectations of peers can be heterogeneous. We consider models in both the group and network settings. This paper provides some theoretical aspects of such models as well as an empirical application.

In the Brock and Durlauf model, an RE represents an expectation on the group behavior. There may be multiple equilibria when social interactions are strong. In such a model, all the equilibria can be characterized as roots of a certain function on a proper interval of the real line. The function provides a simple (numerical) device to determine the number of multiple equilibria with given observed group characteristics and any possible values of parameters in the model. For the model with heterogeneous expectations, a RE will be a vector of expectations and it is a solution of a system of nonlinear equations. The number of solutions of a system will, in general, be difficult to be determined. In the group setting, it turns out there exists an aggregate scalar index which represents an equilibrium for the group as a whole, and an allocation mechanism which distributes the aggregate quantity to each individual in the group which satisfies the rational property. The number of multiple equilibria can be determined as roots of a certain function on the real line.

Sufficient conditions for the existence of unique equilibrium are established for both the group and network models. The system of equations which characterize the RE is a contraction mapping under those sufficient conditions. With a logistic distribution for the choice probability, the sufficient condition for a unique equilibrium is the same as that of Brock and Durlauf (2001a, 2001b).

In application, adolescents' smoking behaviors are studied. Both contextual and endogenous social effects are included, and results show there exists a significant endogenous peer effect, and many contextual variables are also significant in influencing an adolescent's smoking behavior.

Appendix: Proofs of Propositions

A.1 Proof of Proposition 1

Proof of part 1). Because F is a continuous function with its range on the interval $[0,1]$, the fixed point theorem is applicable to the vector-value mapping of $(2.2)'$ on $[0,1]^n$ to $[0,1]^n$. So there exists a fixed point $p^* \in [0,1]^n$ satisfies $(2.2)'$.

The function on the right hand side in (2.5) is a continuous function of s and its range is $[0,1]$. Hence the fixed point theorem on $[0,1]$ to $[0,1]$ is applicable, and there exists a fixed point c^* in $[0,1]$ satisfies (2.5).

Proof of part 2). Suppose $s^* \in A_s$, let $p_i^* = h_i(s^*)$ in (2.4). It follows from (2.4) that $p_i^* = F(x_i\alpha + \lambda(\frac{n}{n-1})s^* - \frac{\lambda}{n-1}p_i^*)$, $\forall i$, and, hence, $\frac{1}{n}\sum_{i=1}^n p_i^* = \frac{1}{n}\sum_{i=1}^n F(x_i\alpha + \lambda(\frac{n}{n-1})s^* - \frac{\lambda}{n-1}h_i(s^*))$. As $s^* \in A_s$, it follows $s^* = \frac{1}{n}\sum_{i=1}^n p_i^*$. By substitution, (2.4) implies

$$p_i^* = F\left(x_i\alpha + \lambda(\frac{n}{n-1})\frac{1}{n}\sum_{j=1}^n p_j^* - (\frac{\lambda}{n-1})p_i^*\right) = F\left(x_i\alpha + \frac{\lambda}{n-1}\sum_{j\neq i}^n p_j^*\right), \quad \forall\ i = 1, \cdots, n,$$

i.e., there exists a vector $p \in A_p$.

Conversely, suppose $p^* \in A_p$. Let $s^* = \frac{1}{n}\sum_{i=1}^n p_i^*$. Eq.$(2.2)'$ can be rewritten as $p_i^* = F(x_i\alpha + \lambda(\frac{n}{n-1})s^* - (\frac{\lambda}{n-1})p_i^*)$, $\forall i$. By the setting (2.4), $p_i^* = h_i(s^*)$, $\forall i = 1, \cdots, n$. By summation, $\frac{1}{n}\sum_{i=1}^n p_i^* = \frac{1}{n}\sum_{i=1}^n F(x_i\alpha + \lambda(\frac{n}{n-1})s^* - (\frac{\lambda}{n-1})h_i(s^*))$. Hence, $s^* = \frac{1}{n}\sum_{i=1}^n F(x_i\alpha + \lambda(\frac{n}{n-1})s^* - (\frac{\lambda}{n-1})h_i(s^*))$, i.e., $s^* \in A_s$.

Proof of part 3). Suppose s^* is the unique element of A_s. Suppose there exists $(\tilde{p}_1, \cdots, \tilde{p}_n)$ such that $\tilde{p}_i = F(x_i\alpha + \frac{\lambda}{n-1}\sum_{j\neq i}^n \tilde{p}_j)$. Then $\tilde{p}_i = F(x_i\alpha + \lambda(\frac{n}{n-1})\frac{1}{n}\sum_{j=1}^n \tilde{p}_j - (\frac{\lambda}{n-1})\tilde{p}_i)$. Denote $\tilde{s} = \frac{1}{n}\sum_{j=1}^n \tilde{p}_j$. One has $\tilde{p}_i = F(x_i\alpha + \lambda(\frac{n}{n-1})\tilde{s} - (\frac{\lambda}{n-1})\tilde{p}_i)$. By the setting (2.4), $\tilde{p}_i = h_i(\tilde{s})$, $\forall i = 1, \cdots, n$. Summing over i, one has $\frac{1}{n}\sum_{i=1}^n \tilde{p}_i = \frac{1}{n}\sum_{i=1}^n F(x_i\alpha + \lambda(\frac{n}{n-1})\tilde{s} - \frac{\lambda}{n-1}\tilde{p}_i)$, which is $\tilde{s} = \frac{1}{n}\sum_{i=1}^n F(x_i\alpha + \lambda(\frac{n}{n-1})\tilde{s} - (\frac{\lambda}{n-1})h_i(\tilde{s}))$. By the uniqueness of s^* in A_s, $\tilde{s} = s^*$. Corresponding to s^*

$$h_i(s^*) = F\left(x_i\alpha + \lambda(\frac{n}{n-1})s^* - (\frac{\lambda}{n-1})h_i(s^*)\right) = F\left(x_i\alpha + \lambda(\frac{n}{n-1})\tilde{s} - (\frac{\lambda}{n-1})h_i(s^*)\right),$$

it follows that $h_i(\tilde{s}) = h_i(s^*)$, i.e., $\tilde{p}_i = h_i(s^*)$ for all $i = 1, \cdots, n$. This shows the unique solution in $(2.2)'$ when A_s is a singleton.

For the uniqueness of (2.5), let $\tilde{s}$ such that $\tilde{s} = \frac{1}{n}\sum_{i=1}^n F(x_i\alpha + \lambda(\frac{n}{n-1})\tilde{s} - (\frac{\lambda}{n-1})h_i(\tilde{s}))$. At $\tilde{s}$, the setting (2.4) implies $F(x_i\alpha + \lambda(\frac{n}{n-1})\tilde{s} - (\frac{\lambda}{n-1})h_i(\tilde{s})) = h_i(\tilde{s})$. These imply that $\tilde{s} = \frac{1}{n}\sum_{i=1}^n h_i(\tilde{s})$. Hence,

$$h_i(\tilde{s}) = F(x_i\alpha + \lambda(\frac{n}{n-1})\frac{1}{n}\sum_{j=1}^n h_j(\tilde{s}) - (\frac{\lambda}{n-1})h_i(\tilde{s})) = F(x_i\alpha + \lambda(\frac{1}{n-1})\sum_{j\neq i}^n h_j(\tilde{s})).$$

Suppose p^* is the unique element in A_p, then $h_i(\tilde{s}) = p_i^*$, $\forall i = 1, \cdots, n$. Hence $\tilde{s} = \frac{1}{n}\sum_{i=1}^n p_i^*$ is unique when A_p is a singleton.

Proof of part 4). The mapping $p : A_s \to A_p$ is onto, which is shown in *2)*. It remains to show that this mapping on A_s is one-one. Let s_1^* and s_2^* in A_s such that $h(s_1^*) = h(s_2^*) = p^*$. If $\lambda = 0$, $s_1^* = s_2^*$ from (2.5). So without loss of generality, consider $\lambda \neq 0$. The (2.4) implies that $p_i^* = F(x_i\alpha + \lambda(\frac{n}{n-1})s_1^* - \frac{\lambda}{n-1}p_i^*) = F(x_i\alpha + \lambda(\frac{n}{n-1})s_2^* - \frac{\lambda}{n-1}p_i^*)$ for any i. By the strictly increasing property of F, $s_1^* = s_2^*$ follows. Q.E.D.

A.2 Proof of Proposition 2

Let $k_i(p_i) = F(v_i(p_i))$ where $v_i(p_i) = x_i\alpha + \lambda(\frac{n}{n-1})s - (\frac{\lambda}{n-1})p_i$ with $p_i \in [0,1]$. Define $l_i(p_i) = k_i(p_i) - p_i$. It is apparent that $l_i(0) \geq 0$ and $l_i(1) \leq 0$. The derivative of $l_i(p_i)$ is $\frac{\partial l_i(p_i)}{\partial p_i} = -\frac{\lambda}{(n-1)}f(v_i(p_i)) - 1$. If $\lambda = 0$, $h_i(s) = F(x_i\alpha)$ is the unique solution. If $\lambda > 0$, $\frac{\partial l_i(p_i)}{\partial p_i} < 0$ for all p_i. When $\lambda < 0$, the condition that $\lambda > -\frac{n-1}{\max_u f(u)}$ implies that $|\lambda| < \frac{n-1}{f(u)}$ for all u, and hence $\frac{\partial l_i(p_i)}{\partial p_i} < 0$ for all $p_i \in [0,1]$. The continuity of the function $h_i(s)$ follows from the implicit function theorem. Q.E.D.

A.3 Proof of Proposition 3

As there exists a unique function $h_i(s)$ such that $h_i(s) = F(x_i\alpha + \lambda(\frac{n}{n-1})s - (\frac{\lambda}{n-1})h_i(s))$ for each s in (2.4), the derivative of $h_i(s)$ can be evaluated from this implicit equation. As

$$\frac{\partial h_i(s)}{\partial s} = f\left(x_i\alpha + \lambda(\frac{n}{n-1})s - (\frac{\lambda}{n-1})h_i(s)\right) \cdot (\frac{\lambda}{n-1})\left(n - \frac{\partial h_i(s)}{\partial s}\right),$$

it follows that

$$\frac{\partial h_i(s)}{\partial s} = \frac{\lambda(\frac{n}{n-1})f\left(x_i\alpha + \lambda(\frac{n}{n-1})s - \frac{\lambda}{n-1}h_i(s)\right)}{\left[1 + \frac{\lambda}{n-1}f\left(x_i\alpha + \lambda(\frac{n}{n-1})s - \frac{\lambda}{n-1}h_i(s)\right)\right]}.$$

Let $\bar{h}(s) = \frac{1}{n}\sum_{i=1}^{n} F\left(x_i\alpha + \lambda(\frac{n}{n-1})s - (\frac{\lambda}{n-1})h_i(s)\right)$. We shall investigate conditions on the existence and uniqueness of a fixed point of $\bar{h}(s)$, i.e., $\exists s \ni \bar{h}(s) = s$. From the setting (2.4), we have $\bar{h}(s) = \frac{1}{n}\sum_{i=1}^{n} h_i(s)$. Because $h_i(s) \in [0,1]$, $\bar{h}(s) \in [0,1]$ for all s. The derivative of $\bar{h}(s)$ is

$$\frac{\partial \bar{h}(s)}{\partial s} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial h_i(s)}{\partial s} = \frac{\lambda}{n-1}\sum_{i=1}^{n}\frac{f(u_i(s))}{[1 + \frac{\lambda}{n-1}f(u_i(s))]},$$

where $u_i(s) = x_i\alpha + \lambda(\frac{n}{n-1})s - (\frac{\lambda}{n-1})h_i(s)$. Define $q(s) = \bar{h}(s) - s$ for $s \in [0,1]$. At $s = 0$, $q(0) = \bar{h}(0) \geq 0$. At $s = 1$, $q(1) = \bar{h}(1) - 1 \leq 0$ because $0 \leq \bar{h}(s) \leq 1$ for all s. So the fixed point problem can be investigated in terms of the function $q(s)$ crossing the s-axis in $[0,1]$. If $q(s)$ is a strictly decreasing function, then the solution will be unique. The derivative of $q(s)$ is

$$\begin{aligned}\frac{\partial q(s)}{\partial s} &= \frac{\partial \bar{h}(s)}{\partial s} - 1 = \frac{\lambda}{n-1}\sum_{i=1}^{n}\frac{f(u_i(s))}{[1 + \frac{\lambda}{n-1}f(u_i(s))]} - 1 \\ &= \sum_{i=1}^{n}\frac{\{\frac{\lambda}{n-1}f(u_i(s)) - \frac{1}{n}[1 + \frac{\lambda}{n-1}f(u_i(s))]\}}{[1 + \frac{\lambda}{n-1}f(u_i(s))]} = \frac{n-1}{n}\sum_{i=1}^{n}(\frac{\lambda f(u_i(s)) - 1}{n + \lambda f(u_i(s)) - 1}).\end{aligned}$$

If $\lambda = 0$, $\frac{\partial q(s)}{\partial s} < 0$. When $\lambda > 0$, a sufficient condition for $\frac{\partial q(s)}{\partial s} < 0$ is when $\lambda \max_u f(u) - 1 < 0$, i.e., $\lambda < \frac{1}{\max_u f(u)}$. On the other hand, when $\lambda < 0$, a sufficient condition is $(n-1) + \lambda \max_u f(u) > 0$. Thus a sufficient condition for $\frac{\partial q(s)}{\partial s} < 0$ is that $-\frac{n-1}{\max_u f(u)} < \lambda < \frac{1}{\max_u f(u)}$. Q.E.D.

A.4 Proof of Proposition 4

The gradient matrix of $H(p)$ is

$$\frac{\partial H(p)}{\partial p'} = \begin{pmatrix} 0 & f\left(u_1\left(p\right)\right)\frac{\lambda}{(n-1)} & \cdots & f\left(u_1\left(p\right)\right)\frac{\lambda}{(n-1)} \\ \vdots & \vdots & \vdots & \vdots \\ f\left(u_n\left(p\right)\right)\frac{\lambda}{(n-1)} & f\left(u_n\left(p\right)\right)\frac{\lambda}{(n-1)} & \cdots & 0 \end{pmatrix}.$$

By using the maximum row sum matrix norm $\| \cdot \|_\infty$,

$$\| \frac{\partial H(p)}{\partial p'} \|_\infty = |\lambda| \max_{i=1,\cdots,n} f(u_i(p)) \leq |\lambda| \max_u f(u).$$

Under the condition that $|\lambda| < \frac{1}{\max_u f(u)}$, the $H(p)$ is a contraction mapping. This is so, because for any two vectors p and q, by the mean value theorem,

$$\| H(p) - H(q) \|_\infty = \| \frac{\partial H(\bar{p})}{\partial p'}(p-q) \|_\infty \leq c \| p - q \|_\infty$$

where $\bar{p}$ lies between p and q, and $c = |\lambda| \max_u f(u)$.[23] Under the condition that $|\lambda| < \frac{1}{\max_u f(u)}$, $c < 1$ and, hence, $H(p)$ is a contraction mapping. Thus, under such a condition on λ, p^* exists and is unique as it is the unique fixed point of $H(p)$. Q.E.D.

[23] Because the mean value theorem is applicable to functions but not vector-value functions, the mean value theorem shall be separately applied to each component of $H(p)$. Thus, strictly speaking, each row of the gradient matrix of $H(p)$ may be evaluated at a different $\bar{p}$. The upper bound is applicable without change. The above simplicity is conventional in econometric writings as long as it did not cause additional complication.

REFERENCES

Akerlof, G. A. (1997), "Social Distance and Social Decisions." *Econometrica* 65, 1005-1027.

Anselin, L. (1988), *Spatial Econometrics: Methods and Models.* Kluwer Academic Publishers, The Netherlands.

Bearman, P.S., J. Jones and J.R. Udry (1997), *The National Longitudinal Study of Adolescent Health: Research Design.* http://www.cpc.unc.edu/projects/addhealth/design.

Becker, G. (1974), "A Theory of Social Interactions." *Journal of Political Economy* 82, 1063-1093.

Becker, G. (1996), *Accounting for Tastes.* Harvard University Press, Cambridge, MA.

Becker G. S., M. Grossman, and K. Murphy (1994), "An Empirical Analysis of Cigarette Addiction." *American Economic Review* 84, 396-418.

Bernheim, B.D. (1994), "A Theory of Conformity." *Journal of Political Economy* 102, 841-847.

Brock, W. and S. Durlauf (2001a), "Discrete Choice with Social Interaction." *Review of Economic Studies* 68, 235-260.

———, (2001b), "Interaction-Based Models" in *Handbook of Econometrics 5*, J. Heckman and E. Leamer eds., Elsevier, pp. 3297-3380.

———, (2003), "Multinomial Choice with Social Interactions." NBER Technical Working Papers 0288.

Case, A.C. (1991), "Spatial Patterns in Household Demand", *Econometrica* 59, 953-965.

Case, A. C. and L. F. Katz (1991), "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths." NBER Working Paper No. 3705.

Chaloupka F. J., and M. Grossman (1996), "Price, Tobacco Control Policies and Youth Smoking." NBER Working Paper No. 5740.

Cliff, A., and J.K. Ord (1973), *Spatial Autocorrelation.* London: Pion.

Department of Health and Human Services (1994). *Preventing Tobacco Use Among Young People: A Report of the Surgeon General.* Washington, D.C.: U.S. Government Printing Office.

Duncan, G. J., J. Boisjoly and K. M. Harris (2001), "Sibling, Peer, Neighbor, and Schoolmate Correlations as Indicators of the Importance of Context for Adolescent Development." *Demography* 38, 437-447.

Gaviria, A., and S. Raphael (2001), "School-Based Peer Effects and Juvenile Behavior." *Review of Economics and Statistics*, Vol. 83, No. 2., pp. 257-268.

Glaeser, E., B. Sacerdote and J. Scheinkman, (1996), "Crime and Social Interactions." *Quarterly Journal of Economics* 111, 507-548.

Gouriéroux, C., Monfort, A., (1996), *Simulation-Based Econometric Methods.* Oxford University Press, Oxford.

Greene, W. H. (2000), *Econometric Analysis*, Fourth Edition, Prentice Hall.

Gruber, J. (2001), "Youth Smoking in the 1990's: Why Did It Rise And What Are the Long-run Implications?." *American Economic Review* 91, 85–90.

Gruber, J and J. Zinman (2000), "Youth Smoking in the U.S.: Evidence and Implications." NBER Working Paper 7780.

Haynie, D. L. (2002), "Friendship Networks and Delinquency: The Relative Nature of Peer Delinquency." *Journal of Quantitative Criminology* 18, 99-134.

Harris, J. E. and B. G. López-Valcácel (2004), "Asymmetric Social Interaction in Economics: Cigarette Smoking Among Young People in the United States, 1992-1999." NBER Working Paper, 10409.

Heckman, J.J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46: 931-960.

Krauth, B. V. (2006), "Simulation-based Estimation of Peer Effects." *Journal of Econometrics* 133, 243–271.

Lee, L-f., (2007), "Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects." *Journal of Econometrics* 140, 333-374.

Li, J. and L-f., Lee (2006), "Binary Choice under Social Interactions: an Empirical Study With and Without Subjective Data on Expectations." Manuscript, Department of Economics, Ohio State University, Columbus.

Lin, X. (2007), "Peer Effects and Student Academic Achievement: An Application of Spatial Autoregressive Models with Group Unobservables." Working paper, Department of Economics, Tsinghua University, Beijing, China.

Lloyd-Richardson, E. E., G. Papandonatos, A. Kazura, C. Stanton, R. Niaura (2002), "Differentiating Stages of Smoking Intensity among Adolescents: Stage-Specific Psychological and Social Influences." *Journal of Consulting and Clinical Psychology* 70, 998–1009.

Maddala, G.S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press.

Manski, C. (1993), "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60, 531-542.

Manski, C. (2000), "Economic Analysis of Social Interactions." *Journal of Economic Perspectives* 14, 115-136.

Manski, C. (2004), "Measuring Expectations." *Econometrica* 72, 1329-1376.

McFadden D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior." In P. Zarembka, ed., *Frontiers in Econometrics.* New York: Academic Press.

Moffitt, R.A. (2001). "Policy Interventions, Low-level Equilibria, and Social Interactions." in S. Durlauf and H. P. Young, eds., *Social Dynamics*, Cambridge, Mass.: MIT Press, pp. 45-82.

Norton, E. C., R. C. Lindrooth, S. T. Ennett (2003), "How Measures of Perception from Survey Data Lead to Inconsistent Regression Results: Evidence from Adolescent and Peer Substance Use." *Health Economics* 12, 139–148.

Powell, L. M., J. A. Tauras and H. Ross (2005), "The Importance of Peer Effects, Cigarette Prices and Tobacco Control Policies for Youth Smoking Behavior." *Journal of Health Economics* 24, 950-968.

Sacerdote, B. (2001), "Peer Effects With Random Assignment: Results For Dartmouth Roommates." *Quarterly Journal of Economics* 116, 681-704.

Soetevent, A.R., and P. Kooreman (2007), "A Discrete-Choice Model with Social Interactions: With An Application to High School Teen Behavior." *Journal of Applied Econometrics* 22, 599-624.

Tamer, E. (2003), "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria." *Review of Economic Studies* 70, 147-165.

Wang, M. Q., E. C. Fitzhugh, R. C. Westerfield, J. M. Eddy (1995), "Family and Peer Influences on Smoking Behavior among American Adolescents: An Age Trend." *Journal of Adolescent Health* 16, 200–203.

Weinberg, B., P. B. Reagan and J. J. Yankow (2004), "Do Neighborhoods Affect Work Behavior? Evidence from the NLSY79." *Journal of Labor Economics* 24, 891-924.

Table 1. Definitions and Summary Statistics of Variables

Mean, Standard deviation, and Correlation Coefficients
(Standard deviations are in parentheses)

Variable	Definition	Own	Contextual	ρ
Smoking	1 if ever smoke; 0 otherwise	-0.542	-0.666	0.272
(Dependent var.)		(0.659)	(0.570)	
Age	Age	15.054	11.067	0.040
		(1.690)	(6.683)	
$Age^2/10$	Squared-age/10	22.948	16.727	0.185
		(5.066)	(10.718)	
Years in school	Number of years in current school	2.511	1.922	0.497
		(1.424)	(1.584)	
Male	1 if male; 0 otherwise	0.488	0.330	0.310
		(0.500)	(0.358)	
Black	1 if black; 0 otherwise	0.182	0.127	0.665
		(0.386)	(0.307)	
Asian	1 if Asian; 0 otherwise	0.066	0.047	0.481
		(0.248)	(0.172)	
Hispanic	1 if Hispanic; 0 otherwise	0.138	0.091	0.465
		(0.345)	(0.240)	
Other race	1 if American Indian or other race;	0.056	0.039	0.072
	0 otherwise	(0.230)	(0.130)	
White	*1 if white; 0 otherwise*	0.558	0.438	0.534
		(0.497)	(0.443)	
Live with both parents	1 if living with both parents; 0 otherwise	0.731	0.563	0.143
		(0.443)	(0.417)	
Sports club	1 if participating in any kind of sport;	0.526	0.424	0.265
	0 otherwise	(0.499)	(0.395)	
Mom education less than HS	1 if mom's education less than high school;	0.100	0.066	0.127
	0 otherwise	(0.300)	(0.178)	
Mom education more than HS	1 if mom's education more than high school;	0.412	0.330	0.226
	0 otherwise	(0.492)	(0.356)	
Mom education missing	1 if the information about mom's education	0.107	0.070	0.057
	is missing; 0 otherwise	(0.309)	(0.174)	
Mom education HS	*1 if mom's education high school;*	0.380	0.275	0.091
	0 otherwise	(0.486)	(0.321)	
Mom's job is professional	1 if mom is a doctor, lawyer, scientist, teacher,	0.262	0.213	0.116
	executive, director and the like; 0 otherwise	(0.440)	(0.282)	
Mom other jobs	1 if mom's occupation is not among the	0.358	0.272	0.062
	"Professional" or "Stay home" ; 0 otherwise	(0.479)	(0.311)	
Mom job missing	1 if the information about mom's job	0.090	0.060	0.025
	is missing; 0 otherwise	(0.286)	(0.160)	
Stay home	*1 if mom is a homemaker, retired*	0.282	0.192	0.029
	or does not work; 0 otherwise	(0.450)	(0.272)	
Mom on welfare	1 if mom receives public assistance, such as	0.009	0.005	0.041
	welfare; 0 otherwise	(0.093)	(0.046)	

Note: The variables in italic font are categories left out in estimation (to avoid dummy variables trap).

Table 2. Parameter Estimates: No Social Effects and Endogenous Effect Only.

(Standard errors in parentheses. *:10% significant; **:5% significant; ***:1% significant.)

	1		2	
	No Social Effects		Endog. Only	
λ			.374***	(.013)
Own Characteristics:				
Age	1.018***	(.049)	.894***	(.009)
$\text{Age}^2/10$	-.298***	(.016)	-.266***	(.003)
Years in school	-.032***	(.003)	-.009***	(.002)
Male	.022**	(.009)	.004	(.006)
Black	-.611***	(.015)	-.566***	(.009)
Asian	-.324***	(.020)	-.303***	(.014)
Hispanic	-.285***	(.014)	-.275***	(.008)
Other race	.098***	(.018)	.094***	(.017)
Live with both parents	-.168***	(.011)	-.160***	(.010)
Sports club	-.154***	(.009)	-.135***	(.007)
Mom education less than HS	.043**	(.016)	.044***	(.014)
Mom education more than HS	-.075***	(.011)	-.067***	(.009)
Mom education missing	-.105***	(.017)	-.112***	(.019)
Mom's job is professional	.063***	(.013)	.069***	(.012)
Mom other jobs	.077***	(.011)	.081***	(.010)
Mom job missing	.070***	(.018)	.060***	(.017)
Mom on welfare	.115**	(.049)	.103**	(.045)
Constant	-8.764***	(.370)	-7.527***	(.071)
2*log-likelihood value	-76373.60		-75933.70	

Table 3. Parameter Estimates: Endogenous, Contextual, and Correlated Effects.

(Standard errors in parentheses. *:10% significant; **:5% significant; ***:1% significant.)

	3		4		5	
	Contextual Only		Both Effects		Correlated Effect	
λ			.664***	(.020)	.663***	(.038)
σ					.005	(.018)
Own Characteristics:						
Age	.992***	(.050)	.774***	(.013)	.780***	(.043)
$Age^2/10$	-.295***	(.016)	-.226***	(.005)	-.228***	(.014)
Years in school	-.003	(.004)	-.001	(.004)	-.001	(.004)
Male	.020*	(.010)	.019**	(.009)	.019*	(.010)
Black	-.546***	(.020)	-.536***	(.018)	-.536***	(.020)
Asian	-.238***	(.023)	-.227***	(.023)	-.227***	(.023)
Hispanic	-.233***	(.016)	-.218***	(.015)	-.218***	(.016)
Other race	.099***	(.019)	.091**	(.021)	.092***	(.019)
Live with both parents	-.155***	(.011)	-.143***	(.012)	-.143***	(.011)
Sports club	-.123***	(.010)	-.114***	(.010)	-.114***	(.010)
Mom education less than HS	.038**	(.016)	.034**	(.017)	.034**	(.016)
Mom education more than HS	-.050***	(.011)	-.038***	(.013)	-.039***	(.011)
Mom education missing	-.107***	(.017)	-.104***	(.021)	-.104***	(.017)
Mom's job professional	.071***	(.014)	.066***	(.014)	.066***	(.014)
Mom other jobs	.076***	(.012)	.071***	(.013)	.071***	(.012)
Mom job missing	.058***	(.018)	.050***	(.019)	.050***	(.018)
Mom on welfare	.100**	(.049)	.097*	(.054)	.097**	(.049)
Contextual Characteristics:						
Age	.011	(.007)	.070***	(.005)	.070***	(.007)
$Age^2/10$	.003	(.004)	-.030***	(.003)	-.030***	(.004)
Years in school	-.023***	(.006)	-.011**	(.005)	-.010*	(.005)
Male	-.020	(.017)	-.023*	(.013)	-.023	(.015)
Black	-.240***	(.026)	.034	(.025)	.033	(.030)
Asian	-.265***	(.035)	-.091***	(.032)	-.091***	(.034)
Hispanic	-.253***	(.025)	-.083***	(.024)	-.083***	(.025)
Other race	.118***	(.035)	.076***	(.029)	.077**	(.031)
Live with both parents	-.226***	(.020)	-.122***	(.017)	-.122***	(.020)
Sports club	-.118***	(.016)	-.027**	(.013)	-.027*	(.016)
Mom education less than HS	.059*	(.030)	.041*	(.023)	.041	(.026)
Mom education more than HS	-.148***	(.020)	-.088***	(.016)	-.088***	(.018)
Mom education missing	-.065***	(.032)	.0003	(.027)	.0003	(.028)
Mom's job professional	.081***	(.021)	.039**	(.019)	.039*	(.022)
Mom other jobs	.082***	(.021)	.031*	(.018)	.032*	(.019)
Mom job missing	.124***	(.034)	.082***	(.027)	.082***	(.030)
Mom on welfare	.002	(.106)	-.056	(.091)	-.056	(.095)
Constant	-8.435***	(.385)	-6.752***	(.094)	-6.796***	(.333)
2*log-likelihood value	-75691.82		-75549.72		-75549.72	

Table 4. Summary of Representative Group's Characteristics.

Variable	Mean	(s.d.)
Age	15.281	(.849)
$\text{Age}^2/10$	23.422	(2.673)
Years in school	2.231	(.990)
Male	0.446	(.499)
Black	0.322	(.469)
Asian	0.083	(.277)
Hispanic	0.099	(.300)
Other race	0.099	(.300)
Live with both parents	0.760	(.429)
Sports club	0.537	(.501)
Mom education less than HS	0.182	(.387)
Mom education more than HS	0.388	(.489)
Mom education missing	0.157	(.365)
Mom's job is professional	0.347	(.478)
Mom other jobs	0.413	(.495)
Mom job missing	0.149	(.357)
Mom on welfare	0.091	(.289)

Table 5. Marginal Effects: Models with no Social Effects or Endogenous Effect only
(in percentage point)

	No Social Effects	Endogenous Effect Only		
	naive	naive	sophisticated	
			changer	affected
Peers' expected average choice		12.461		
Own Characteristics				
Age	3.797	2.968	3.022	.006
Years in school	-.531	-.300	-.304	-.001
Male	.739	.130	.132	-.000
Black	-20.466	-18.867	-18.986	-.036
Asian	-9.196	-10.102	-8.858	-.018
Hispanic	-8.262	-9.162	-8.154	-.017
Other race	3.372	3.135	3.306	.007
Live with both parents	-5.757	-5.314	-5.548	-.011
Sports club	-4.948	-4.485	-4.405	-.009
Mom education less than HS	1.444	1.476	1.512	.003
Mom education more than HS	-2.453	-2.242	-2.247	-.005
Mom education missing	-3.342	-3.715	-3.608	-.008
Mom's job professional	2.147	2.309	2.402	.005
Mom other jobs	2.513	2.705	2.711	.006
Mom job missing	2.383	2.006	2.071	.004
Mom on welfare	3.992	3.442	3.642	.008

Table 6. Marginal Effects: Models with Social Effects
(in percentage point)

	Contextual Effects			Both Social effects		
	naive	sophisticated		naive	sophisticated	
		changer	affected		changer	affected
Own Characteristics:						
Age	3.376	3.376	.004	3.143	3.337	.009
Years in school	-.092	-.092	-.005	-.048	-.0931	-.004
Male	.673	.673	-.004	.666	.623	-.004
Black	-18.845	-18.845	-.050	-18.484	-19.700	-.050
Asian	-7.312	-7.312	-.050	-7.822	-7.892	-.067
Hispanic	-7.172	-7.172	-.049	-7.532	-7.617	-.064
Other race	3.491	3.491	.028	3.154	3.789	.044
Live with both parents	-5.402	-5.402	-.046	-4.920	-5.919	-.069
Sports club	-4.056	-4.056	-.024	-3.920	-4.192	-.029
Mom education less than HS	1.283	1.283	.012	1.170	1.426	.020
Mom education more than HS	-1.676	-1.676	-.029	-1.325	-1.759	-.037
Mom education missing	-3.494	-3.494	-.013	-3.575	-3.687	-.019
Mom's job is professional	2.441	2.441	.017	2.291	2.649	.026
Mom other jobs	2.566	2.566	.016	2.446	2.712	.023
Mom job missing	1.981	1.981	.026	1.729	2.198	.038
Mom on welfare	3.517	3.517	.0003	3.361	3.479	-.000
Peers' expected choice	—	—	—	22.921	—	—
Contextual Characteristics:						
Age	.386	—	—	2.429	—	—
Years in school	-.766	—	—	-.362	—	—
Male	-.661	—	—	-.794	—	—
Black	-8.139	—	—	1.173	—	—
Asian	-8.993	—	—	-3.126	—	—
Hispanic	-8.576	—	—	-2.857	—	—
Other race	4.519	—	—	2.633	—	—
Live with both parents	-7.657	—	—	-4.196	—	—
Sports club	-3.993	—	—	-.914	—	—
Mom education less than HS	2.003	—	—	1.422	—	—
Mom education more than HS	-5.003	—	—	-3.036	—	—
Mom education missing	-2.193	—	—	.010	—	—
Mom's job is professional	2.739	—	—	1.332	—	—
Mom other jobs	2.769	—	—	1.08	—	—
Mom job missing	4.193	—	—	2.812	—	—
Mom on welfare	0.058	—	—	-1.929	—	—