The Empirical Content of Models with Multiple Equilibria in Economies with Social Interactions

Alberto Bisin Andrea Moro Giorgio Topa*

Sept. 30th, 2009[†‡]

Abstract

In this paper we propose a feasible estimation procedure for a general class of models with social interactions which might display multiple equilibria. We evaluate the efficiency and computational feasibility of different approaches to solving the curse of dimensionality implied by the equilibrium multiplicity and we implement the proposed estimation procedure using Add Health data to understand how group interactions affect teenagers' smoking behavior.

*Bisin: Department of Economics, New York University, alberto.bisin@nyu.edu. Moro: Department of Economics, Vanderbilt University, andrea@andreamoro.net. Topa: Research Group, Federal Reserve Bank of New York, giorgio.topa@ny.frb.org. The views and opinions offered in this paper do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

†We have presented preliminary versions of this research at several universities and conferences starting in 2002. The gestation of this paper has been so long that it's impossible to acknowledge all that gave us comments and encouragement. We are grateful to them nonetheless. Finally, we thank Chris Huckfeldt and Matt Denes for exceptional research assistance.

‡This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu). No direct support was received from grant P01-HD31921 for this analysis.

1 Introduction

In this paper we study identification and estimation of models with multiple equilibria in economies with social interactions. Social interactions refer to socioeconomic environments in which markets do not necessarily mediate all of agents' choices. In such environments each agent's ability to interact with others might depend on a network of relationships, e.g., a family, a peer group, or more generally a socioeconomic group. Social interactions represent an important aspect of several socioeconomic phenomena like crime, school performance, risky behavior of teen regarding sexual activity, alcohol and drug consumption, smoking, obesity, and, more generally, are related to neighborhood effects, which are important determinants of economic outcomes such as employment, the pattern of bilateral trade and economic specialization, migration, urban agglomeration and segregation.

Social interactions typically give rise to multiple equilibria because they induce externalities. Consider for example a society of agents whose preferences for smoking are stronger the higher the proportion of smokers in the population, for example because agents have preferences for conformity. In this society, we may find equilibria where few people smoke and equilibria where many people smoke if the dependence of agents' preferences on the proportion of smokers in the population is strong enough. In fact, uniqueness conditions in this class of models, like e.g., Glaeser and Scheinkman (2001)'s *moderate social influence condition*, are very strong limitations on the strength of the interactions. It is because of externalities of this kind that the canonical model of social interaction, Brock and Durlauf (2001b), also displays multiple equilibria, even in its simplest formulation.

In this paper, we refer to economies with social interactions as "societies." We consider a general society with (possibly) multiple equilibria, and assume the investigator observes data realized from one or more of the equilibria. We define the likelihood of the data conditionally on the equilibrium selection, and begin by introducing an estimator of the structural parameters of the model based on maximizing the likelihood of the data over both the set of equilibria and the set of the structural parameters. Such estimator requires the ability to compute all of the equilibria that are consistent with a given set of parameters, often a daunting computational task.

Next, we propose a computationally simpler two-step estimation procedure that does not require the computation of all feasible equilibria, or assuming an arbitrary equilibrium selection rule. Furthermore, we evaluate the efficiency and computational feasibility of the two approaches using Monte-Carlo simulations. We show, in the context of a simple econometric model of social interactions, that while less efficient, the two step method is faster by several orders of magnitude. We also show that estimation procedures based on the adoption of an arbitrary equilibrium selection rule are less efficient (and again much slower) than our two-step method, which is agnostic about equilibrium selection. Hence, our method is particularly appropriate when the investigator does not have information about how the equilibrium selection is performed.

Finally we implement the proposed estimation procedure using data on smoking behavior for different U.S. schools. We impose restrictions on which parameters are common across schools and estimate the model using data from the National Longitudinal Study of Adolescent Health ("Add Health"), a longitudinal study of a nationally representative sample of adolescents in grades 7-12 in the United States during the 1994-95 school year. We present various estimates of the model with different assumptions regarding the dependence of the equilibrium selection across schools on various observable variables, such as geo-locational factors.

In our data, an individual's smoking level is positively associated with the number of smokers within the individual's friendship network. The positive association holds even after controlling for individual characteristics such as grade, race and gender. Further, the data exhibit large variation in aggregate smoking levels across schools. These facts are consistent with social interactions of significant strenght and are suggestive that there may be scope for different schools to be in different equilibria with regard to smoking prevalence. Structural estimation of the social interactions model described above sheds light on whether the data are indeed consistent with different schools exhibiting different social equilibria.

We find that social interactions are widespread in the schools we consider, both school-wide and at the level of personal networks. Our parameter estimates are consistent with the presence of multiple equilibria in our empirical application. Simulations of the model indicate that changes in the strength of local or school-wide social interactions, changes in the number of friends in personal networks, or policies aimed at discouraging tobacco use in schools can have highly non-linear and sometimes counter-intuitive effects, with the possibility of large shifts in smoking prevalence because of the presence of multiple equilibria.

1.1 Related Literature

The identification and estimation of models with social interactions is an active area of research. In this context, the issue of identification has been clearly analyzed by Manski (1993) with regards to the linear in means model.[1] Multiplicity of equilibria is typically not an issue for this class of reduced form linear models. More generally, however, when agents' policy functions are non-linear, multiplicities arise, especially when social interactions take the form of strategic complementarities, as in the case of preferences for conformity with a reference group. In this case, *moderate social influence* assumptions, limiting the effect of social interactions, are required for uniqueness (see Glaeser and Scheinkman (2001)). In empirical work, typically, sufficient conditions for uniqueness are assumed (Glaeser, Sacerdote, and Scheinkman (1996)).

In the canonical non-linear model of social interactions – Brock and Durlauf (2001b)'s binary choice model – multiple equilibria are hard to dispel (see Soetevent and Kooreman (2007) for a generalization). In this case, social interactions are identified under functional form assumptions on the stochastic structure of preference shocks, as well as non-parametrically (Brock and Durlauf, (2007)). Brock and Durlauf (2001a), Krauth (2006), and Soetevent and Kooreman (2007) extend the binary choice analysis to economies of local interactions, that is to economies in which the social interactions occur not only at the level of the population, but also at the level of small (finite) peer groups such as friends, family,

[1] More recently, sufficient conditions for identification in this context have been proposed by Graham-Hahn (2005), Bramoulle', Djebbari, and Fortin (2007), Davezies, D'Haultfoeuille, and Fougere (2006), and Lee (2007), among others.

etc. Krauth (2006) allows for correlated effects, that is, for correlation in the preference shocks across peers, under specific parametric assumptions).

In this class of models identification does not rely on assuming a specific equilibrium selection mechanism. As a consequence, identification is obtained even if only one realization of equilibrium is observed in the data, e.g., when the social interaction occurs at the population level and only the actions of agents belonging to a single population are observed. Nonetheless, when data are available about several equilibrium realizations (populations) a specific selection mechanism is often specified in the estimation procedure; see e.g., Krauth (2006) and Nakajima, (2007)[2].

Important and related work on the econometrics of multiple equilibria has also been done in macroeconomics and in industrial organization. Dagsvik-Jovanovic (1994) study economic fluctuations in a model with two equilibria (high and low economic activity) each period; they postulate a stochastic (Markovian) equilibrium selection process over time and estimate the parameters of such process with time series data on economic activity. The adopted functional form specification allows them to derive closed form solutions of the mapping from the set of equilibria to the set of parameters, which helps contructing the sample likelihood for estimation. Imrohoroglu (1993) and Farmer and Guo (1995) estimate dynamic macroeconomic models of inflation and business cycles, respectively, with a continuum of equilibria by parametrizing equilibria with a sunspot process and recovering from data the time series of the sunspot realizations under assumptions on the properties of the process; see also Aiyagari (1995). Recent development of these methods are surveyed in Benhabib and Farmer (1998). In this paper we show that assuming a specific equilibrium selection (or sunspot) process is not always necessary, and may lead to inefficient estimates if the assumed process is not the "true" data-generating process.

An important literature which studies the issue of identification in multiple equilibrium models of industrial organization concentrates on simultaneous-move finite games of com-

[2]In fact, this distinction is not clear in the literature. Krauth (2006), for instance, claims that "an equilibrium selection rule must usually be imposed to achieve point identification of parameters" (p. 259, end of first paragraph). Nakajima (2007) makes the same point (p.9, second paragraph).

plete information where the investigator only observes the actions played by the agents, whereas the parameters to be estimated also affect the payoffs.[3] Classic examples include the entry-game studied by Bresnahan and Reiss, (1991), which has extensive applications e.g., also in labor economics. In this class of games the model is typically not identified: a continuum of parameter values is consistent with the same equilibrium realization of the strategy profile. Partial identification is however possible, as shown by Tamer (2003) for large classes of such *incomplete econometric structures*[4]. Others, as Bjorn and Vuong (1985), Bresnahan and Reiss, (1991), and Bajari, Hong, and Ryan (2007), have opted for imposing assumption which guarantee identification. Bajari, Hong, and Ryan (2007), in particular, have interesting results about estimation as well. The estimation procedure they adopt requires the computation of all equilibria of the game for any element of the parameter set and the joint estimation of the parameters of an equilibrium selection mechanism (in an ex-ante pre-specified class) which determines the probability of a given equilibrium, as in Dagsvik-Jovanovic (1994).

Bajari, Hong, Krainer, and Nekipelov (2006) and Aguirregabiria-Mira (2007) study instead, respectively, static and dynamic versions of a discrete entry-game of incomplete information. In this context, they study the properties of a two-step estimator similar in spirit to ours.[5] A version of this estimator had been introduced by Moro (2002) in the context of a model of statistical discrimination in the labor markets. In that application, the equilibrium map linking wages to the individual characteristics of workers is different across different equilibria and hence the model can be identified and estimated off cross-sectional data.

Finally, our application to teenagers' smoking behavior is also part of a large empirical literature regarding social interactions. While the evidence for strong social interactions, or

[3]See Berry and Tamer (2007) for a survey of this literature.

[4]See also, Manski and Tamer (2002), Andrews, Berry, and Jia (2004), Ciliberto and Tamer (2004), Beresteanu, Molchanov, and Molinari (2008). Echenique and Komunjer (2005) have results regarding the identification of monotone comparative statics in incomplete econometric structures.

[5]See also Pakes, Ostrovsky, and Berry (2004), Pesendorfer and Schmidt-Dengler (2004), Bajari, Benkard, and Levin (2003). See also Aguirregabiria (2004) for some foundational theoretical econometric work.

peer effects, is overwhelming, part of this literature relies on linear-in-means models, which as shown by Manski (1993) are not identified, and attribute to social interactions any effects that are possibly due instead to selection and/or common shocks. This is the case, e.g., of Wang, Fitzhugh, Westerfield, and Eddy (1995) and Wang, Eddy, and Fitzhugh (2000); see also the review in Tyas and Pederson (1998). Instrumental variable estimates, as in Norton, Lindrooth, and Ennett (1998), Gaviria and Raphael (2001), and Powell, Tauras, and Ross (2003), attempt to address these problems.[6]

Empirical studies of social interactions in smoking behavior which estimate binary choice models along the lines of Brock and Durlauf (2001b), as we noted, overcome the identification issue of the linear-in-means approach, but need to deal with multiple equilibria. Krauth (2006) employs an array of exogenously pre-specified selection mechanisms. Similarly, Soetevent and Kooreman (2007) impose a specific random - uniform, in fact - selection mechanism. This approach requires in principle the computation of all equilibria for any element of the parameter space.[7] Nakajima (2007) also exploits a selection mechanism, although the mechanism is implicitly determined by an adaptive learning mechanism.

2 A general society

Agents and groups. Consider a society populated by a finite set of agents indexed by $i \in I$.[8] Each agent $i \in I$ is characterized by a vector of exogenous characteristics X_i. Let $\mathbf{X} = (X_i)_{i \in I}$. The population is partitioned into sub-populations indexed by $n = 1, ..., N$, represented by disjoint sets I_n, such that $\bigcup_{n=1}^{N} I_n = I$. Let $\#I_n$ denote the dimensionality of set I_n and $\#I$ the dimensionality of I. Typically, different sub-populations are interpreted as

[6] See also Bauman and Fisher (1986), Krosnick and Judd (1982), and Jones (1984).

[7] Though the structure of Nash equilibria of supermodular games can be exploited to reduce the set of equilibria to be computed under strong assumptions on the support of theselection mechanism. We conjecture this approach is adopted by Kraus (2006) though information on the selection mechanism in the smoking application is not clearly reported.

[8] We impose several regularity requirements in the exposition, e.g., continuity, compactness, finiteness, even though they could be somewhat relaxed.

cities, neighborhoods, ethnic groups, schools. Each sub-population n is in turn characterized by a vector of exogenous characteristics Z_n. Let $\mathbf{Z} = (Z_n)_{n \in N}$, $\mathbf{X}_n = (X_i)_{i \in I_n}$, and $\mathbf{X} = (X_i)_{i \in I}$.[9]. Each agent $i \in I_n$ belongs to a group $g(i) \subset I_n$.[10] Local interactions, for example peer effects, will manifest themselves in group $g(i)$. Let $\#g(i)$ be the dimensionality of $g(i)$.

Choices. Each agent $i \in I$ chooses an element y_i in a compact set Y (possibly a discrete set) simultaneously. Let $\mathbf{y}_{g(i)}$, $\mathbf{y}_n$, denote respectively the vectors of choices in groups $g(i)$, I_n, respectively, and $\mathbf{y}$ denote the vector of cohices of all agents. Finally let ε_i denote a vector of idiosyncratic shocks hitting agent $i \in I$. We assume ε_i and ε_j are independent, for any $i \neq j \in I$. Let ε_n denote the vector of aggregate shocks hitting all agents $i \in I_n$. Let $\boldsymbol{\varepsilon} = (\varepsilon_n)_{n \in N}$ and let $f_{\varepsilon n}(\varepsilon_i, \varepsilon_n)$ denote the joint density of the shocks, which are assumed to be defined on compact support. We allow the distribution $f_{\varepsilon n}$ to depend on the choice vector $\mathbf{y}_n$. Typically, these shocks are preference shocks, but they could also represent technology shocks.

Equilibrium characterization. Let π_n denote an A-dimensional vector of aggregates defined at the level of sub-population n. Let in turn Π denote a vector of aggregates defined at the level of the whole population. Let $\boldsymbol{\pi} = (\pi_n)_{n \in N}$. We let π_n be implicitly determined by:

$$A_n(\mathbf{y}_n, \pi_n, \Pi, Z_n, \varepsilon_n) = 0,$$

for some vector valued smooth map A_n. Similarly, we let Π be determined by:

$$A(\mathbf{y}, \boldsymbol{\pi}, \Pi, \mathbf{Z}, \boldsymbol{\varepsilon}) = 0,$$

for some vector valued smooth map A.

Remark. *Typically, π_n contains an externality or a global social interaction effect; e.g., if Y is binary, it could contain the fraction of agents in sub-population n who choose*

[9] All characteristics can be thought of realizations from random variables in a compact support with joint density $f_{xz}(\mathbf{X}, \mathbf{Z})$.

[10] By construction $g(i)$ does not contain i. The restriction that agent i's and his/her group $g(i)$ belong to the same subpopulation is irrelevant, but it simplifies notation.

a particular element of $y \in Y$: $\frac{1}{\# I_n}\sum_{i\in I_n} I_{y_i=y}$ (where $I_{y_i=y}$ is the indicator function of $y_i = y$). If the society contains a competitive market component, then π_n typically contains the competitive price vector (for those markets which clear at the level of the sub-population n), and the function A_n encodes the aggregate excess demands in some of its components. Similarly, Π contains global externalities and prices determined at the level of the whole population.

We assume that the set of first order conditions which determine the choice of an arbitrary agent $i \in I_n$ can generally be written as:

$$y_i = y_i\left(X_i, \mathbf{y}_{g(i)}, \pi_n, Z_n, \Pi, \varepsilon_i, \varepsilon_n\right), \quad \text{for some smooth function } y_i(\cdot) \tag{1}$$

if Y is a continuous set. If instead Y is a discrete set, we assume

$$\Pr(y_i = y \in Y) = y_i\left(X_i, \mathbf{y}_{g(i)}, \pi_n, Z_n, \Pi, \varepsilon_i, \varepsilon_n\right) \text{ for some smooth function } y_i(\cdot). \tag{2}$$

Requiring the first order conditions to be satisfied jointly for any agent $i \in I$ is equivalent to requiring that $\mathbf{y}$ satisfies a Nash equilibrium of the simultaneous move anonymous game.

Remark. *Typically, the first order conditions will result from agent i's choice of y_i to maximize preferences:*

$$V\left(y_i, X_i, \mathbf{y}_{g(i)}, \pi_n, Z_n, \Pi, \varepsilon_i, \varepsilon_n\right)$$

To guarantee that the first order conditions are characterized by a continuous function $y_i(.)$ we need to restrict agent i's choice set to be compact and convex and his/her preferences to satisfy smoothness and convexity.[11]

Note that our formulation assumes that the system of first order conditions and the equilibrium conditions are bloc recursive, in the sense that X_i, ε_i enter the equilibrium conditions only through the *choice vector* $\mathbf{y}$.

[11] A technology constraint of the form

$$y_i \in Y(X_i, \mathbf{y}_{g(i)}, \pi_n, Z_n, \Pi, \varepsilon_i, \varepsilon_n)$$

on the choice set adds no complications as long as it defines a compact convex subset of Y.

Definition. *An equilibrium of the society is a vector $y \in Y^I$ which satisfies the first order conditions 1 (or 2 if Y is discrete) and the equilibrium conditions*

$$\begin{aligned} A_n(\mathbf{y}_n, \pi_n, Z_n, \Pi, \varepsilon_n) &= 0 \\ A(\mathbf{y}, \boldsymbol{\pi}, \mathbf{Z}, \Pi, \boldsymbol{\varepsilon}) &= 0 \end{aligned}$$

We assume that $(\mathbf{y}, \mathbf{X}, \boldsymbol{\pi}, \mathbf{Z}, \Pi, \boldsymbol{\varepsilon})$ are observed by the econometrician[12]. However, the econometrician does not observe the vectors $\varepsilon_i, \varepsilon_n$. Any agent $i \in I_n$ instead, observes $\varepsilon_i, \varepsilon_n$ as well as the whole vectors $\mathbf{X}, \mathbf{Z}$ before choosing y_i. Depending on the specific society we study, agent $i \in I_n$ might observe some component of $\boldsymbol{\pi}$,Π, e.g., market prices.

2.1 Special case 1: Binary choice model: Brock and Durlauf

Agent $i \in I_n$ chooses an outcome $y_i \in \{-1, 1\}$, to maximize preferences based on its own choice, and on an aggregate of the peer's choices. In its simplest formulation interactions are only at the level of the sub-population I_n agent i belongs to:

$$\max_{y_i \in \{-1,1\}} V\left(y_i, X_i, Z_n, \pi_n, \varepsilon_i\right) = h_n(X_i, Z_n, \varepsilon_n) \cdot y_i + J_n y_i \pi_n + \varepsilon_i. \tag{3}$$

The aggregate π_n is defined to be the average choice in the population. Consequently, the single equilibrium condition for the society, $A_n(\mathbf{y}_n, \pi_n, Z_n, \Pi, \varepsilon_n) = 0$ in our general formulation, takes the simple form:

$$\pi_n = \frac{1}{\# I_n} \sum_{i \in I_n} \left(I_{y_i=1} - I_{y_i=-1}\right).$$

We assume that $h_n(X_i, Z_n, \varepsilon_n)$ is linear and we eliminate any dependence from Z_n for simplicity,

$$h_n(X_i, Z_n, \varepsilon_n) = c_n X_i + \varepsilon_n.$$

The unobserved shock distribution $f_\varepsilon(\varepsilon_i, \varepsilon_n)$ is assumed to satisfy independence $f_{\varepsilon n}(\varepsilon_i, \varepsilon_n) = f_\varepsilon(\varepsilon_i) \cdot f_n(\varepsilon_n)$. Furthermore, $f_\varepsilon(\varepsilon_i)$ depends on agent i's choice y_i and it is extreme-valued,

[12] Measurement error and sampling error can be easily added to our analysis.

that is[13],

$$\Pr\left(\varepsilon_i\left(-1\right)-\varepsilon_i\left(1\right)\leq z\right)=\frac{1}{1+\exp\left(-z\right)}.$$

The utility of each choice y_i is then:

$$V\left(y_i,X_i,\pi_n,\varepsilon_i\left(y_i\right)\right)=\left(c_nX_i+\varepsilon_n\right)\cdot y_i+J_ny_i\pi_n+\varepsilon_i\left(y_i\right).$$

From the first order conditions, it can be shown that

$$\begin{aligned}\Pr\left(y_i=1\right) &= \frac{1}{1+\exp\left(-2\phi_i\right)}\\ \Pr\left(y_i=-1\right) &= 1-\Pr\left(y_i=1\right)=\frac{1}{1+\exp\left(2\phi_i\right)}\end{aligned}$$

where

$$\phi_i\equiv c_nX_i+\varepsilon_n+J_n\pi_n.$$

Assuming as an approximation that $\#I_n$ is large enough that the Law of Large Numbers applies for each sub-population n, we obtain the following characterization of equilibrium :

$$\pi_n=\sum_{i\in I_n}\int\tanh\left(c_nX_i+\varepsilon_n+J_n\pi_n\right)f_n(\varepsilon_n)d\varepsilon_n. \tag{4}$$

It is straightforward to show that, typically, this equation has multiple solutions. In fact, assuming scalar individual characteristics and abstracting from sub-population shocks ε_n, it is shown by Brock and Durlauf (2001) that the resulting equilibrium condition is $\pi_n=\tanh\left(c_nX_i+J_n\pi_n\right)$ which generally has one or three solutions.

Note that in (3) no social interactions connect different sub-populations, e.g., no markets or externalities are formed at the global level across sub-populations. The model can be easily extended to allow preferences to depend on (i) local interactions, by adding dependence on $\pi_{g(i)}=\frac{1}{\#g(i)}\sum_{i\in g(i)}\left(I_{y_i=1}-I_{y_i=-1}\right)$; and (2) global interactions (at the level of the whole population), through dependence on $\Pi=\frac{1}{\#I}\sum_{i\in I}\left(I_{y_i=1}-I_{y_i=-1}\right)$.

[13] More generally, for an extreme value distribution,

$$\Pr\left(\varepsilon_i\left(-1\right)-\varepsilon_i\left(1\right)\leq z\right)=\frac{1}{1+\exp\left(-\beta z\right)}$$

where the parameter β is the variance of the distribution. But normalizing $\beta=1$ is without loss of generality in our setting, as it is equivalent to normalizing the units of the utility function.

2.2 Special case 2: Continuous choice model; Glaeser and Scheinkman

Agent $i \in I_n$ chooses an outcome $y_i \in [0, 1]$, as a solution of

$$\max_{y_i \in [0,1]} V\left(y_i, g\left(\mathbf{y}_{g(i)}\right), X_i, Z_n, \pi_n, \varepsilon_i, \varepsilon_n\right)$$

where $g\left(\mathbf{y}_{g(i)}\right) = \sum_{j \in g(i)} \gamma_{ij} y_j$, with $\gamma_{ij} \geq 0$, $\sum\limits_{j \in g(i)} \gamma_{ij} = 1$. An equilibrium is then a $\mathbf{y}_n \in [0, 1]^{I_n}$ such that: the first order conditions,

$$\frac{\partial V\left(y_i, \sum_{j \in g(i)} \gamma_{ij} y_j, X_i, Z_n, \pi_n, \varepsilon_i, \varepsilon_n\right)}{\partial y_i} = 0,$$

are satisfied, for any $i \in I_n$; and, at equilibrium,

$$\pi_n = \frac{1}{\# I_n} \sum_{i \in I_n} y_i.$$

The extension to encompass global interactions at the level of the whole population is straightforward.

3 Identification

In this section we study identification in the set-up of Section 2. We shall show that the conditions for identification in multiple equilibrium models are not conceptually more stringent than those which apply to models with a unique equilibrium.

For simplicity we restrict our analisys to the case in which the choice set Y is continuous.[14] Consider first the case in which no externality nor market operates at the global level across sub-populations, so that the equilibrium conditions are simply,

$$A(\mathbf{y}_n, \pi_n, Z_n, \varepsilon_n; \theta_n) = 0,$$

where $\theta_n \in \Theta$ is the vector of parameters to be estimated (Θ is a compact set) - added for clarity in the notation. We derive next conditions for θ_n to be identified only from data regarding the single sub-population n, $\mathbf{y}_n$.

[14] The case in which Y is discrete can be dealt with similar methods; see Sections 5 and 6; see also Brock and Durlauf (2001b, 2007).

An equilibrium in sub-population n also includes the first order conditions[15]

$$y_i = y_i\left(X_i, \mathbf{y}_{g(i)}, \pi_n, Z_n, \varepsilon_i, \varepsilon_n; \theta_n\right).$$

Without loss in generality, let us assume that the vector of parameters θ_n can be partitioned as $\theta_n = \left[\theta_n^{foc}, \theta_n^{eq}\right]$ so that:

$$A(\mathbf{y}_n, \pi_n, Z_n, \varepsilon_n; \theta_n) = 0$$
$$y_i = y_i\left(X_i, \mathbf{y}_{g(i)}, \pi_n, Z_n, \varepsilon_i, \varepsilon_n; \theta_n^{foc}\right)$$

In general, the *likelihood* for the random variable $\mathbf{y}_n$, given θ_n, is defined as $L(\mathbf{y}_n|\theta_n)$. In our setup, because of the possible presence of multiple equilibria, $L(\mathbf{y}_n|\theta_n)$ *is a correspondence*:

$$L(\mathbf{y}_n|\theta_n) = \int_{(\varepsilon_i,\varepsilon_n):\left\{\begin{array}{c} y_i = y_i\left(X_i, \mathbf{y}_{g(i)}, \pi_n, Z_n, \varepsilon_i, \varepsilon_n; \theta_n^{foc}\right) \\ \pi_n = \pi_n(\theta_n, \varepsilon_n) \end{array}\right\}} f_\varepsilon(\varepsilon_i, \varepsilon_n) d\varepsilon_i d\varepsilon_n \qquad (5)$$

where $\int$ denotes the Aumann integral and $\pi_n(\theta_n, \varepsilon_n)$ is the equilibrium manifold which satisfies $A_n(\mathbf{y}_n, \boldsymbol{\pi}_n, Z_n, \varepsilon_n; \theta_n) = 0$[16].

Let $\mathbf{L}(\theta_n)$ be the set of measurable likelihood functions induced by (5), so that any $l(\mathbf{y}_n|\theta_n) \in \mathbf{L}(\theta_n)$ is a measurable selection of the correspondence $L(\mathbf{y}_n|\theta_n)$.The standard sufficient condition for identification of the parameter vector θ_n at θ^0 is generalized as follows.

Definition. For all $\theta_n \in \Theta, \theta_n \neq \theta^0$,

$$\arg \max_{l(\mathbf{y}_n|\theta^0) \in \mathbf{L}(\theta^0)} l(\mathbf{y}_n|\theta^0) \notin \mathbf{L}(\theta_n).$$

[15] The dimensionality of the system of first order conditions is equal to the dimension of the vector of choice variables y_i. The dimensionality of the equilibrium conditions is equal to A, the dimension of the vector of equilibrium variables π_n.

[16] See Aliprantis for the formal definition and a discussion of the properties of such integral. Loosely speaking, the Riemann integral is not defined since $\pi_n(\theta_n^{eq}, \varepsilon_n)$ is a correspondence; the Aumann integral is defined for correspondences and is constructed by taking the union of the Riemann integrals of all measurable selections of the correspondence; it coincides with the Riemann integral when applied to a measurable function.

The conditions for the identification of θ_n^{foc} off of the first order conditions, given π_n, ε_n, are standard (and involve no issue of multiplicity of equilibria)[17]. We assume these conditions are satisfied and hence θ_n^{foc} is identified, given $\pi_n, Z_n, \varepsilon_n$. The equilibrium conditions can therefore be represented as a map (a smooth manifold, in fact) from $(\theta_n^{eq}, \varepsilon_n)$ into π_n, for given observables $\mathbf{y}_n, Z_n$ and given θ_n^{foc}. Let $\pi_n(\theta_n^{eq}, \varepsilon_n)$ be such a map.

Suppose first that *the econometrician has no prior information about the distribution of ε_n, $f_n(\varepsilon_n)$*. In this case, the realization of ε_n represents a vector of parameters which need to be estimated along with θ_n. Equilibrium is unique if $\pi_n(\theta_n^{eq}, \varepsilon_n)$ is one-to-one. As Figure 1 and 2 illustrate, this is not necessary nor sufficient for identification. Instead,

Strong condition for identification: *Identification obtains if $\pi_n(\theta_n^{eq}, \varepsilon_n)$ is onto; that is, if the inverse equilibrium map $\binom{\theta_n^{eq}}{\varepsilon_n}(\pi_n)$ is one-to-one.*

Note that this condition is unrelated to the existence of multiple or unique equilibria;[18] see panel (a) in Figure 1 for an equilibrium manifold which does not satisfy the requirement that $\binom{\theta_n^{eq}}{\varepsilon_n}(\pi_n)$ be one-to-one. Panel (b) shows instead a manifold $\pi_n(\theta_n^{eq}, \varepsilon_n)$ which is not one-to-one (as an equilibrium manifold it displays multiple equilibria), but is such that the inverse equilibrium manifold $\binom{\theta_n^{eq}}{\varepsilon_n}(\pi_n)$ is one-to-one, and hence satisfies the condition for identification.

Suppose instead that *the econometrician has prior, e.g., functional form information about the marginal distribution of ε_n, $f_n(\varepsilon_n)$.*

Weak condition for identification: *Identification obtains if i) $\theta_n^{eq}(\pi_n, \varepsilon_n)$ is one-to-one and if ii) $f_n(\varepsilon_n)$ is a known (to the econometrician) and strictly monotonic function*[19].

As for the previous case, this requirement is unrelated to the existence of multiple or unique equilibria.

[17] In particular, note that the presence of $\mathbf{y}_{g(i)}, \pi_n$ between the independent variables induces Manski's reflection problem only non-generically, typically in the linear case.

[18] Jovanovic (1989) shows that a unique reduced form is neither necessary nor sufficient condition for identification.

[19] Regularity conditions on $f_\varepsilon(\varepsilon_n)$ are necessary to avoid the non-generic case in which the likelihood of the data is maximized at multiple distinct values of the parameters. For instance, if $f_\varepsilon(\varepsilon_n)$ is assumed uniform, identification might not be guaranteed even if $\theta_n^{eq}(\pi_n, \varepsilon_n)$ is one-to-one.

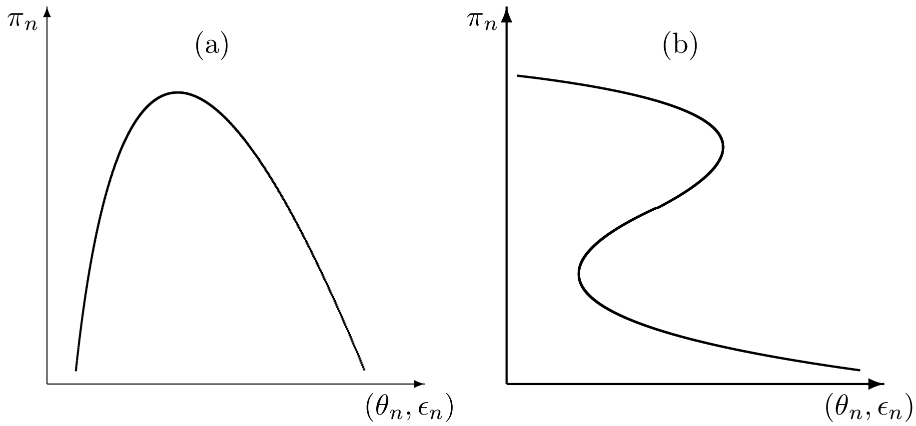

Figure 1: (a) Unique equilibrium, no identification; (b) multiple equilibria with identification

Remark. *Note that the condition that* $\theta_n^{eq}(\pi_n, \varepsilon_n)$ *be one-to-one is weaker than the condition that* $\binom{\theta_n^{eq}}{\varepsilon_n}(\pi_n)$ *be one-to-one. Furthermore note that the condition that* $f_n(\varepsilon_n)$ *be known to the econometrician can be relaxed when different sub-populations* n *are observed by the econometrician. In this case, in principle the parameters of the distribution* $f_\varepsilon(\varepsilon_n)$ *could be estimated along with* θ_n^{eq}*. Finally, the condition that* $f_n(\varepsilon_n)$ *be a strictly monotonic function can also be substantially relaxed in practical applications. Suppose for instance that* $f_n(\varepsilon_n) \sim N(0,1)$*. While in this case the* Weak condition for identification *is not satisfied, identification obtains if* $\theta_n^{eq}(\pi_n, \varepsilon_n)$ *is one-to-one and* $\theta_n^{eq}(\pi_n, \varepsilon_n) \neq \theta_n^{eq}(\pi_n, -\varepsilon_n)$.

It should be apparent that the analysis of identification in this section can be readily extended to societies with global interactions at the level of the whole population. Note also that measurement error and sampling error can be dealt with by appropriately re-defining the set of shocks $(\varepsilon_i, \varepsilon_n)$.

4 Estimation

The previous section argues that identification is no more an issue when the model has multiple equilibria than when it has a unique equilibrium. This is not the case for estimation, because the identification conditions imply the investigator's ability to compute all the feasible equilibria for every set of parameters, which computationally is ofter a daunting task. To simplify notation, with no loss of generality, we continue restricting ourselves to the environment introduced in the previous section. We also assume the *Weak condition for identification* is satisfied.

4.1 The direct estimation method

We define a direct Maximum likelihood estimator of θ_n as follows:

$$\widehat{\theta}_n = \arg\max_{\theta_n} \max_{l(\mathbf{y}_n|\theta_n)\in\mathbf{L}(\theta_n)} l(\mathbf{y}_n|\theta_n) \tag{6}$$

Recall that $\mathbf{L}(\theta_n)$ is the set of all measurable selections $l(\mathbf{y}_n|\theta_n)$ induced by the correspondence $\mathbf{L}(\mathbf{y}_n|\theta_n)$ defined by (5). Because of the possible multiplicity of equilibria, $\mathbf{L}(\theta_n)$ is very difficult to characterize. However, not surprisingly, the direct Maximum likelihood estimation of θ_n has desired properties.

Proposition. *The estimator $\widehat{\theta}_n$ is consistent and efficient.*

The estimator $\widehat{\theta}_n$ can be computed by using the following algorithm: for each $(\theta_n, \varepsilon_n)$, compute all the equilibria of the model, compute the likelihood of each equilibrium, choose the maximum among them, integrate over ε_n and maximize over θ_n. Such procedure is computationally difficult especially when the parametric form of $A_n(\mathbf{y}_n, \boldsymbol{\pi}_n; Z_n, \varepsilon_n; \theta_n) = 0$ does not allow the investigator to know in advance how many solutions the equilibrium correspondence displays.

4.2 The two-step estimation method

We now introduce a simpler two-step estimation procedure. The first step consists in computing an estimator $\widehat{\boldsymbol{\pi}}_n$ for $\boldsymbol{\pi}_n$ from data $\mathbf{y}_n$. In Glaeser and Scheinkman's society, e.g.,

$$\widehat{\boldsymbol{\pi}}_n = \frac{1}{\# I_n} \sum_{i\in I_n} y_i.$$

Let

$$l(\mathbf{y}_n, \widehat{\pi}_n|\theta_n) = \int_{(\varepsilon_i,\varepsilon_n):\ y_i = y_i\left(X_i, \mathbf{y}_{g(i)}, \widehat{\pi}_n, Z_n, \varepsilon_i, \varepsilon_n; \theta_n^{foc}\right)} f(\varepsilon_i, \varepsilon_n) d\varepsilon_i d\varepsilon_n$$

The second step then involves estimating θ_n^{foc} as

$$\widehat{\theta}_n^{foc} = \arg\max_{\theta_n} l(\mathbf{y}_n, \widehat{\pi}_n|\theta_n) \tag{7}$$

It is important to notice that the equilibrium restriction $\widehat{\pi}_n \in \pi_n(\theta_n, \varepsilon_n)$ has not been imposed yet. Furthermore, in the second step, we estimate θ_n^{eq} as

$$\widehat{\theta}_n^{eq} = \arg\max_{\theta_n^{eq}} \int_{\varepsilon_n : \theta_n^{eq} = \theta_n^{eq}(\widehat{\pi}_n, \varepsilon_n)} f_n(\varepsilon_n) d\varepsilon_n$$

Note that the estimation does not require the computation of all the equilibria as a function of θ_n^{eq}, as identification requires that $\theta_n^{eq}(\widehat{\pi}_n, \varepsilon_n)$ be one-to-one by the *Weak identification condition.*

Remark. *The estimate of* $\widehat{\theta}_n$ *obtained by the two-step method just described can be refined by iterating the estimating procedure. From* $\widehat{\theta}_n$ *a vector* $\widehat{\mathbf{y}}_n$ *can be generated from the first order conditions, and the two-step method then can be applied to* $\widehat{\mathbf{y}}_n$ *rather than to* y_n*; and so on iteratively. We do not have general conditions to guarantee that this iteration procedure converges.*[20]

4.3 Equivalence Between Estimators

We can then prove the following proposition describing two alternative sets of sufficient conditions for equivalence to hold.

Proposition. *Sufficient conditions for equivalence of the two-step and the direct estimation procedure of* π_n *are either of the following:*
(i) *for any* (π_n, θ_n) *there exists a unique* ε_n *such that* $\pi_n = \pi_n(\theta_n, \varepsilon_n)$*; moreover,* $\varepsilon_n \sim$ *uniformly; or*
(ii) $\theta_n(\pi_n, \varepsilon_n)$ *is independent of* ε_n *and takes a unique value for any* π_n*.*[21]

[20] See also Aguirregabiria and Mira (2007).

[21] It is interesting to consider a relaxation of condition ii). Suppose $\theta_n(\pi_n, \varepsilon_n)$ is independent of ε_n, but it takes a unique value π_n for π_n in a subset (possibly strict) of its domain and is not defined everywhere else. In this case equivalence does not hold, but the two-step procedure can be easily modified to guarantee equivalence.

Proof. Condition i) states that there is always a ε_n such that π_n is an equilibrium given parameters θ_n. But then if ε_n is uniform, $f(\varepsilon_n)$ is a constant and the second step is redundant. It follows that any maximizer of (7) is also a maximizer of (6).

Condition ii) states that the realization of the data depend only on π_n, not on θ_n. Hence in the first step only π_n is identified. The second condition states that the model is nonstochastic, and that the mapping from π_n to θ_n is a function. Hence, after having estimated π_n in the first step, it is possible to uniquely recover an estimate of π_n in the second step. ■

Moro (2002) has been the first to employ the two-step procedure to estimate a model with multiple equilibria. In his model identifiying restrictions are imposed so that condition ii) holds and therefore the equivalence of the two procedures follows readily.

4.4 Consistency of the Two-Step Estimator

We can now discuss the asymptotic properties of each of the two-step estimator outlined above.

Proposition. *The two-step estimator of (π_n, θ_n) is consistent if the estimator for π_n from data y_n is consistent.*

4.5 Monte Carlo analysis of estimators

We now study in detail the estimation methods in the previous section in the context of the binary choice model of Brock and Durlauf (2001), introduced in Section 2.1.

In this model, independence of ε_i across agents $i \in I$ implies that, for the vector of choices y_n:

$$\begin{aligned} \Pr(\mathbf{y}_n|\mathbf{X}_n, \pi_n, \varepsilon_n) &= \prod_i \Pr(y_i|X_i, \varepsilon_n, \pi_n) \sim \\ &\prod_i \exp\left((c_n X_i + \varepsilon_n) \cdot y_i + J_n y_i \pi_n\right). \end{aligned} \tag{8}$$

Equation (8) suggests the following formulation of the likelihood function as a function of the parameter vector $\theta_n = \{c_n, J_n\}$:

$$l(\mathbf{y}_n|\mathbf{X}_n, \pi_n, \varepsilon_n; \theta_n) = \prod_i \left[\Pr\left(y_i = 1|X_i, \varepsilon_n, \pi_n\right)\right]^{\frac{1+y_i}{2}} \cdot \left[\Pr\left(y_i = -1|X_i, \varepsilon_n, \pi_n\right)\right]^{\frac{1-y_i}{2}} \sim$$
$$\prod_i \left[\exp\left(c_n X_i + \varepsilon_n + J_n \pi_n\right)\right]^{\frac{1+y_i}{2}} \cdot \left[\exp\left(-c_n X_i - \varepsilon_n - J_n \pi_n\right)\right]^{\frac{1-y_i}{2}}$$

We run two sets of experiments: the first focuses on the Brock-Durlauf model in a single sub-population, $N = 1$. We compare the performance of the 2-step estimator to that of the full maximum likelihood estimator (what we call the "direct method"). The second set of experiments is run in a multiple sub-population setting, $N \geq 2$. Here we compare the properties of the 2-step method to both the direct method and another estimation method in which the multiplicity issue is addressed by explicitly incorporating an equilibrium selection mechanism into the likelihood function, as in Dagsvik and Jovanovic (1994).

Note that the slope coefficients c_n are identified in the single subpopulation case by the variation in average smoking across different values of the X's. An intercept term in c_n is only identified in the multiple subpopulation case with commmon parameters if subpopulations select at least two different equilibria, because c_n has the same effect on behavior in all equilibria, but J_n's effect is proportional to the equilibrium behavior. We don't include an intercept term in any of our specifications.

4.5.1 Results for a single subpopulation ($N = 1$)

We use a version of the Brock-Durlauf model with a single covariate $X_i \sim N(\mu_x, 1)$ and global interactions (no local interactions). Thus the model parameters are a pair $\theta \equiv (c, J)$; note that we drop the index n for simplicity as $N = 1$. We draw an artificial sample of 20,000 students (characterized by their attribute X_i) and run a Monte Carlo experiment, drawing $\mathcal{N}$ =160 vectors of the true parameters of the model. Parameter C is drawn from a uniform with support $[-0.8, 0.8]$, and parameter J from a uniform with support $[1, 3]$. For each random draw θ_j^{true} of the model parameters, $j = 1, ..., \mathcal{N}$, we use the model to generate simulated data $\widetilde{y}\left(\theta_j^{true}\right)$, choosing one single equilibrium for a given experiment (i.e., all students are acting according to the same equilibrium). For each simulated dataset

Evaluation Criterion	Direct	Two step	Direct, 2step initial est
RMSE, parameter C	0.0522	0.05043	0.05043
Bias, parameter C	-0.00144	-0.00251	-0.00251
RMSE, parameter J	0.11604	0.10708	0.10706
Bias, parameter J	-0.00749	-0.00542	-0.0039
Min time	163.37621	0.22234	157.19373
Max time	220.05349	0.34216	217.32402
Mean time	188.47448	0.26496	187.49023
Median time	188.14973	0.26686	187.09089

Table 1: Monte Carlo single subpopulation experiment - results (low-equilibrium)

$\widetilde{y}\left(\theta_j^{true}\right)$ we estimate the model parameters using both the 2-step and the direct methods, $\left\{\widehat{\theta}_j^{2s}, \widehat{\theta}_j^{d}\right\}$, $j = 1, ..., \mathcal{N}$. We then compare the properties of the two estimators, focusing on several evaluation criteria: Bias (the average difference between the estimator and the true parameter), Root Mean Squared Error (RMSE) (the root of the average of the squared differences between the estimator and the true parameter) and computational speed.

Table 1 reports the results of the experiments in which the low-level equilibrium was always chosen; results for the intermediate and high-level equilibrium are very similar and are presented in Appendix A. The second column reports properties of the direct method where the starting value θ_0 used in the likelihood maximization routine was fixed at $c = 0, J = 2$. The third column reports statistics for the 2-step method. The fourth column reports results for the direct method when the 2-step estimates $\widehat{\theta}_j^{2s}$ were used as initial values for the maximization algorithm.

The 2-step method always exhibits lower RMSE than the direct method with fixed starting values. This is surprising since the direct method represents the full maximum likelihood estimation and should therefore achieve a weakly lower RMSE. The reason for this result is that, even though we use a maximization algorithm – simulated annealing –

that is very robust to discontinuities in the objective function, in a small but non-trivial number of cases the algorithm gets "stuck" in a region of the parameters that correspond to the wrong equilibrium, which yields estimates very far from θ^{true}. To address this issue, we also use the direct method with $\widehat{\theta}_j^{2s}$ as starting values (column four): in this case, the RMSE is the same or slightly lower than in the 2-step case.

The real advantage of the 2-step nethod, however, is in computational speed. Even with this very stripped down model an estimation run with the direct method took a median time between 158 and 188 minutes (depending on the choice of equilibrium); instead, the 2-step method took roughly between 15 and 35 *seconds*. This is a vast computational advantage that enables the researcher to estimate much richer models of economic behavior than if one were to use brute force maximum likelihood.

4.5.2 Results for multiple subpopulations ($N \geq 2$)

Our second set of experiments concerns a setting with multiple sub-populations n, where all agents in a single subpopulation n, $i \in I_n$, are assumed to be in the same equilibrium but each sub-population may be in a different equilibrium. A possible approach in this sort of settings has been to postulate a selection mechanism across equilibria, which involves a specific correlation structure (in equilibria) across the different sub-populations of the society.

Dagsvik and Jovanovic (1994) and Bajari-Hong-Ryan (2006) take this approach, which enables them to write the likelihood as the product of two terms: loosely speaking, the probability of the data in a given sub-population n, conditional on parameters and on the equilibrium chosen in n; and the probability that sub-population n is in that particular equilibrium given the selection mechanism. Therefore, the likelihood is a mixture of likelihoods conditional on equilibria, where the weights are equal to the probabilities of equilibria given data. Thus the likelihood becomes a well-behaved function rather than a complicated correspondence. The downside of this approach is that the econometrician has to take a stand on the specific equilibrium selection mechanism being used.

We perform two types of experimental comparisons. First we again compare the 2-step estimator to the direct method. Secondly, we compare the 2-step estimator with the estimators obtained by postulating an equilibrium selection (we call this the D-J method, for Dagsvick and Jovanovic). Depending on the experiment, we use $n = 10$ or $n = 25$, with either 10,000 or 150 agents each.[22] To concentrate on equilibrium selection, we assume the parameters are identical across sub-populations: $(c_n, J_n) = (c, J)$, for all n, and are randomly drawn as in the single subpopulations experiments (details on the true parameter draws are in Appendix A). Suppose the equilibrium set contains at most K equilibria, indexed by $k = 1, ..., K$.Let

$$\phi_n(\pi_k) = \Pr(\text{sub-population } n \text{ is in eqm. } \pi_k \,|\, \mathbf{y}_{n-1}, \mathbf{y}_{n-2}, ...) .$$

To simulate the experimental data, we used a second order Spatially Auto-Regressive process (SAR(2)) as our equilibrium selection mechanism. The sub-population is ordered on a one-dimensional integer lattice, where "closeness" in the lattice represents "closeness" in terms of social distance and hence it justifies the correlation structure imposed on equilibrium selection.[23] Let $K = 3$, as is in fact the case in the Brock and Durlauf model we simulate. The first two sub-populations are assigned one of the three possible equilibria at random (independently), with probabilities $(p_1, p_2, 1 - p_1 - p_2)$. For $n > 2$, each sub-population n adopts the same equilibrium as sub-population $n - 1$with probability a_1, and it adopts the same equilibrium as sub-population $n - 2$ with probability a_2; with the residual probability $(1 - a_1 - a_2)$ sub-population n is assigned an equilibrium independently of the preceding sub-population (again with probabilities $(p_1, p_2, 1 - p_1 - p_2)$). The conditional probabilities $\phi_n(\pi_k)$ are computed recursively based on this particular selection mechanism.

[22]We were limited to using 150 artificial students in each sub-population n in the DJ experiment for computational reasons.

[23]In a time series context, correlation across time-periods is perhaps more natural. In a cross-sectional context one can still determine "closeness" between sub-populations by using some notion of social distance: see Conley (1999) or Conley and Topa (2003).

Correlation in eq. selection	$\alpha_1 = \alpha_2 = 1/3$		$\alpha_1 = 0.1, \alpha_2 = 0.8$	
Evaluation Criterion	Direct	Two step	Direct	Two step
(1)	(2)	(3)	(4)	(5)
RMSE, C	0.04118	0.00571	0.10920	0.09355
Bias, C	-0.00138	-0.00002	-0.00628	-0.00400
RMSE, J	0.16222	0.03810	0.35842	0.28770
Bias, J	0.02906	0.00263	0.12616	0.09020
Min time	161.998	2.25910	0.20495	0.00325
Max time	538.023	11.332	550.880	11.435
Mean time	350.937	4.958	295.898	4.2250
Median time	390.359	5.433	382.989	5.419

Table 2: Monte Carlo multiple subpopulations experiments: comparison of direct and two-step methods

Table 2 collects results for the first step of experiments, comparing 2-step and direct methods, where the evaluations are based on the results obtained from running 640 experiments where the "true" parameters are randomly drawn using thesame criteria used in the previous subsection. The second and third columns, concern an experiment in which the parameters of the SAR(2) selection mechanism were set at $a_1 = 1/3$, $a_2 = 1/3$. In this particular case, both RMSE and Bias measures are much lower for the 2-step than for the direct method. We suspect that this is a consequence of the extreme computational difficulties involved in maximizing the full likelihood in the multiple sub-populations case. As in the single sub-population case, computational speed is again roughly two order of magnutide lower for the 2-step than for the direct method.

The RMSE and Bias properties are quite sensitive to the specific parameterization of the selection mechanism. Columns 4 and 5 display results for the case in which the SAR(2) parameters are set at $a_1 = 1/10$, $a_2 = 4/5$. Here the two methods under comparison exhibit

Eval. Criterion (1)	D-J (correct) (2)	Two step (3)	D-J (misspecified) (4)	Two step (5)
RMSE, C	0.02362	0.02714	0.05569	0.02940
Bias, C	0.00096	0.00098	-0.00190	-0.00307
RMSE, J	0.11312	0.18630	0.22973	0.14404
Bias, J	-0.00741	0.09832	0.01148	0.04788
Min time	321.76167	0.04592	161.24889	0.03551
Max time	419.45020	0.13887	300.50744	0.06411
Mean time	346.77147	0.05973	174.00611	0.04943
Median time	347.29222	0.06233	173.30057	0.04832

Table 3: Monte carlo experiments: comparison between the DJ method and the two-step method

similar properties in terms of RMSE and Bias, although computational speed is again much higher for the 2-step.

Finally, we turn to two experiments that perform a comparison between 2-step and D-J methods. Both experiments were run with the same parameterization of the equilibrium selection mechanism, with $a_1 = 1/3$, $a_2 = 1/3$. In Table 3, columns 2 and 3 we report results for a situation in which the econometrician using the D-J method correctly models the selection mechanism as a SAR(2) process, whereas columns 4 and 5 focuses on the case in which the "D-J econometrician" chooses a misspecified model, assuming a SAR(1) selection process.

The comparison is interesting: in the "correctly specified" case, the D-J method performs slightly better than the 2-step in terms of RMSE and Bias, especially with regard to the social interactions parameter J. In the "misspecified" case, instead, the 2-step method does better than the D-J in terms of RMSE, and the difference in Bias is greatly reduced. This points to an obvious advantage of the 2-step method, since it does not rely on making any assumptions on the nature of the equilibrium selection process. Computational speed is always much higher for the 2-step method, even more so than in the comparison with the direct method.

5 Social interactions and smoking

In this section we estimate several different specifications of the social interactions model in Brock and Durlauf (2001), presented in Section 2.1. To this end we use the smoking component of the Add Health data. The National Longitudinal Study of Adolescent Health (Add Health) is a longitudinal study of a nationally representative sample of adolescents in grades 7-12 in the United States during the 1994-95 school year. Add Health combines longitudinal survey data on respondents' social, economic, psychological and physical well-being with contextual data on the family, neighborhood, community, school, friendships, peer groups, and romantic relationships. 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.

In the empirical application, therefore, we encode $y_i = 1$ if agent i smokes and $y_i = -1$ if he/she does not. Each sub-population n is a school. We consider only high schools, which we define as schools having students enrolled in all grades between 9 and 12. Among these, we include only the 45 schools that have at least 400 students in order to have a sufficient number of smokers and minorities in each school. Even with this restrictions, there are cases in which the parameter estimates for specific racial or ethnic groups are not estimated with any precision.

5.1 Specification of parameters

First we estimate (c_n, J_n) separately for each school n, to study the distribution of parameter estimates across schools. Secondly, we estimate a set of model specifications in which the model parameters (c_n, J_n) are either constrained to be the same across schools or to be a function of observed school attributes Z_n: $c_n = c(Z_n)$, $J_n = J(Z_n)$. Finally, we estimate specifications with multiple schools where we allow (c_n, J_n) to contain random coefficients as in (11)-(12) or in (13)-(14).

More specifically, in the case in which the parameters are specified as a function of observed school attributes, let the dimension of the vector Z_n be K, for any school n. We adopt the following linear specification:

$$c_n = c(Z_n) = \alpha_0 + \sum_{k=1}^{K} \alpha_k Z_n^k; \tag{9}$$

$$J_n = J(Z_n) = \gamma_0 + \sum_{k=1}^{K} \gamma_k Z_n^k. \tag{10}$$

In the case in which we let the parameters contain random coefficients the specification we adopt is:

$$c_n = \alpha_0 + \alpha_n, \text{ with } \alpha_n \sim N(0, \sigma_\alpha) \tag{11}$$

$$J_n = \gamma_0 + \gamma_n, \text{ with } \gamma_n \sim N(0, \sigma_\gamma) \tag{12}$$

More generally we can also include school attributes Z_n:

$$c_n = \alpha_0 + \sum_{k=1}^{K} \alpha_k Z_n^k + \alpha_n, \text{ with } \alpha_n \sim N(0, \sigma_\alpha) \tag{13}$$

$$J_n = \gamma_0 + \sum_{k=1}^{K} \gamma_k Z_n^k + \gamma_n, \text{ with } \gamma_n \sim N(0, \sigma_\gamma) \tag{14}$$

We assume that the random fixed effects $\{\alpha_n, \gamma_n\}$ are independent of each other and of the individual random terms $\varepsilon_i(y_i)$ that enter the individuals' random utilities. The idea is to specify the probability distribution of $\{\alpha_n, \gamma_n\}$ so as to put some structure on the distribution of the realized $\{c_n, J_n\}$.

5.2 Specification of social interactions

We explore also different specification of the structure of interaction inside schools. We consider first the case in which interactions are only school-wide, and then the case in which interactions are only local, that is, at the level of each agent's circle of friends, which are identified in the Add Health individual friendship network data. Finally we consider the general case in which interactions have both a school-wide and a local component.

In particular, in the general case each agent $i \in I_n$ has preferences represented by

$$V\left(y_i, \mathbf{y}_{g(i)}, \pi_n, Z_n, X_i, \varepsilon_i\right) = c_n X_i \cdot y_i + J_n^G y_i \pi_n + \sum_{j \in g(i)} J_n^L y_i \pi_{g(i)} + \varepsilon_i$$

where $\pi_{g(i)} = \frac{1}{\#g(i} \sum_{j \in g(i)} y_j$.

Only school-wide interactions are obtained with $J_n^L = 0$; while only local interaction are present with $J_n^G = 0$.

5.3 Empirical results

In what follows we present summaries of the parameter estimates for a selection of specifications. We then perform some simulation exercises to compute the estimated effect on the incidence on smoking of a given reduction in the level of social interactions.

In the general case in which interactions have both a school-wide ("global") and a local (personal network) component, and random fixed effects $\{\alpha_n, \gamma_n\}$ are added to the parameters, the likelihood of the data $\mathbf{y}_n$ given $\pi_n, \left[\pi_{g(i)}\right]_{i \in I_n}, \theta_n$ in school n is:

$$\begin{aligned} \log L(\mathbf{y}_n | \pi_n, \left[\pi_{g(i)}\right]_{i \in I_n}, \theta_n) \quad = \quad & -\sum_{i \in I_n} \left[\begin{array}{l} \left(\frac{1+y_i}{2}\right) \cdot \log\left(1 + \exp\left[-2\left(c_n X_n + J_n^G \pi_n + J_n^L \pi_{g(i)}\right)\right]\right) \\ + \left(\frac{1-y_i}{2}\right) \cdot \log\left(1 + \exp\left[2\left(c_n X_n + J_n^G \pi_n + J_n^L \pi_{g(i)}\right)\right]\right) \end{array} \right] \\ & + \Pr(\alpha_n) + \Pr(\gamma_n). \end{aligned}$$

5.3.1 School by school estimation

In this section the model parameters are estimated separately for each school, i.e., we maximize a separate likelihood for each school. See the end of section 4.5 for a brief intuition about the identification of this model. Table 4 reports the means and medians of the parameter estimates across schools for three sets of estimates: with global interactions

Variable	Global interactions Mean	Global interactions Median	Local interactions Mean	Local interactions Median	Local/Global int. Mean	Local/Global int. Median
Black	-0.8375	-0.8463	-0.7349	-0.6718	-0.7370	-0.6731
Asian	-0.1168	-0.2837	-0.0545	-0.2238	-0.0585	-0.2370
Hispanic	0.0313	-0.1607	-0.0373	-0.2172	-0.0211	-0.1699
Female	0.2670	0.2364	0.1137	0.1790	0.1260	0.1852
Age	0.1197	0.1048	0.0452	0.0439	0.1050	0.0785
Does not belong to a club	0.4033	0.3607	0.3338	0.2761	0.3397	0.2926
GPA	-0.3943	-0.3766	-0.3342	-0.3162	-0.3281	-0.3241
Mom college	-0.0606	-0.1141	-0.0578	-0.1182	-0.0351	0.1565
Dad at home	-0.2262	-0.2353	-0.2655	-0.2520	-0.2035	-0.2152
Global Interaction	2.4648	1.9281	-	-	1.0210	1.4099
Local Interactions	-	-	0.8077	0.8202	0.7962	0.8239

Table 4: Mean and median parameter estimates, all schools estimated separately

only, with local interactions only, and with both local and global interactions[24].

In all specifications the parameter estimates are qualitatively consistent with those from a reduced form logit model of smoking behavior. The signs of the median coefficients associated to each covariate are the same as in the logit and are broadly consistent with other studies of smoking: minorities (Blacks, Asians, Hispanics) tend to smoke less than Whites. Female students, older students, and students who do not participate in any school clubs, organizations or athletic teams tend to smoke more. Students who perform better academically and students whose father is present at home tend to smoke less.

Figures 2, 3, and 4 report the distribution of parameter estimates for the three specifications. Only parameter estimates with t test statistics greater than unity are plotted. There is considerable dispersion around the medians across schools, especially with regard

[24]The parameter estimates for each school are available from the authors upon request.

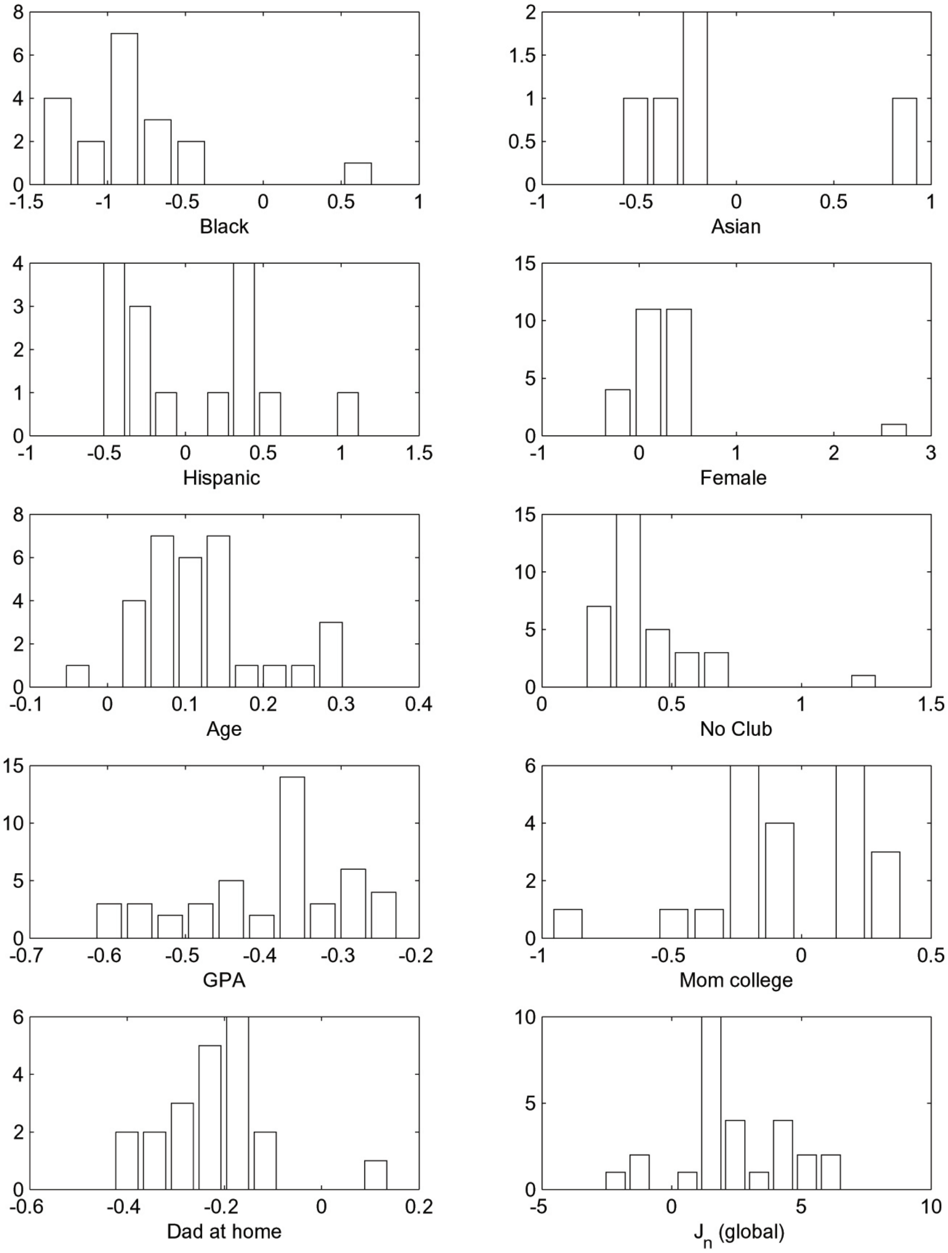

Figure 2: Distributions of parameter values, all schools separately, global interactions only

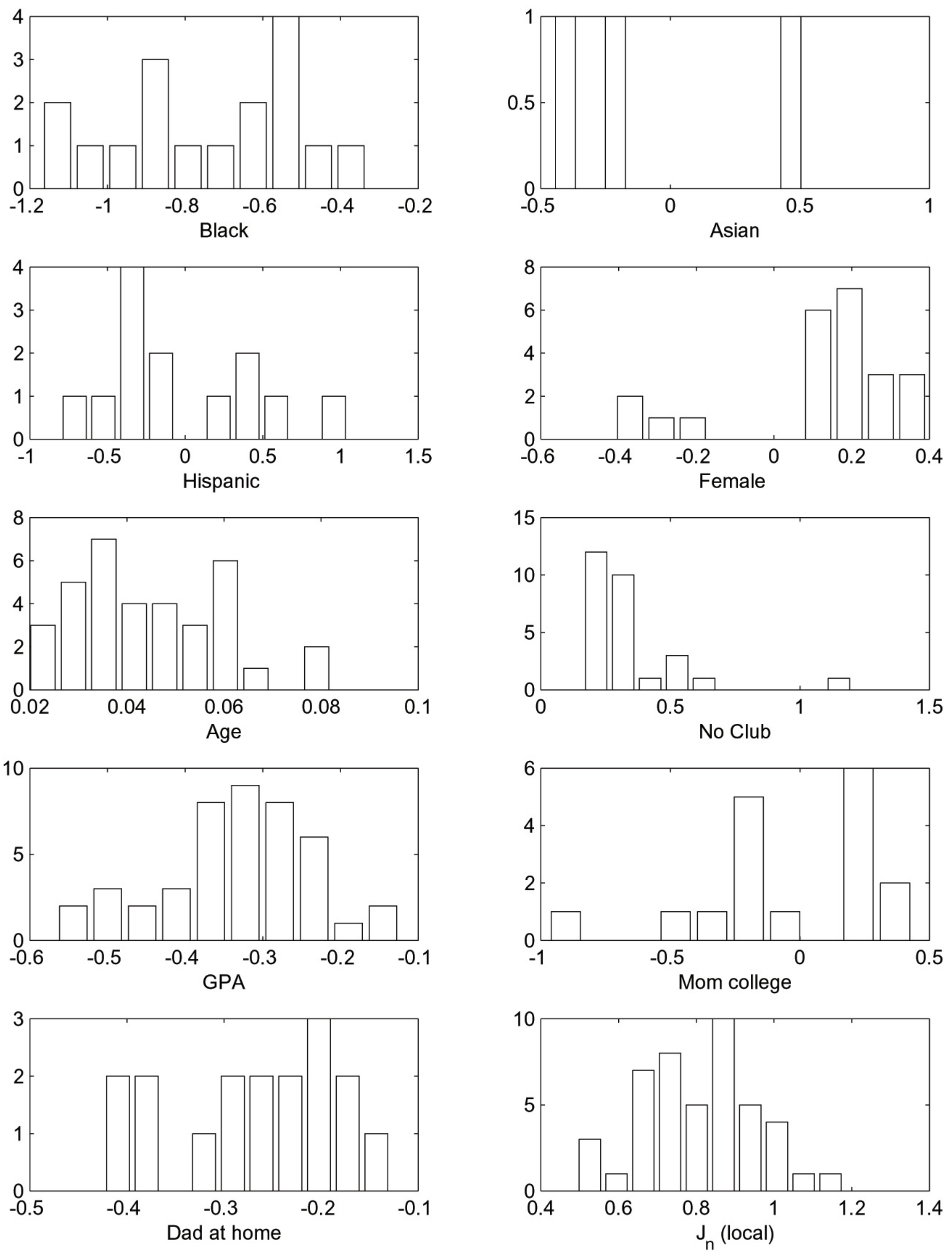

Figure 3: Distributions of parameter values, all schools separately, local interactions only

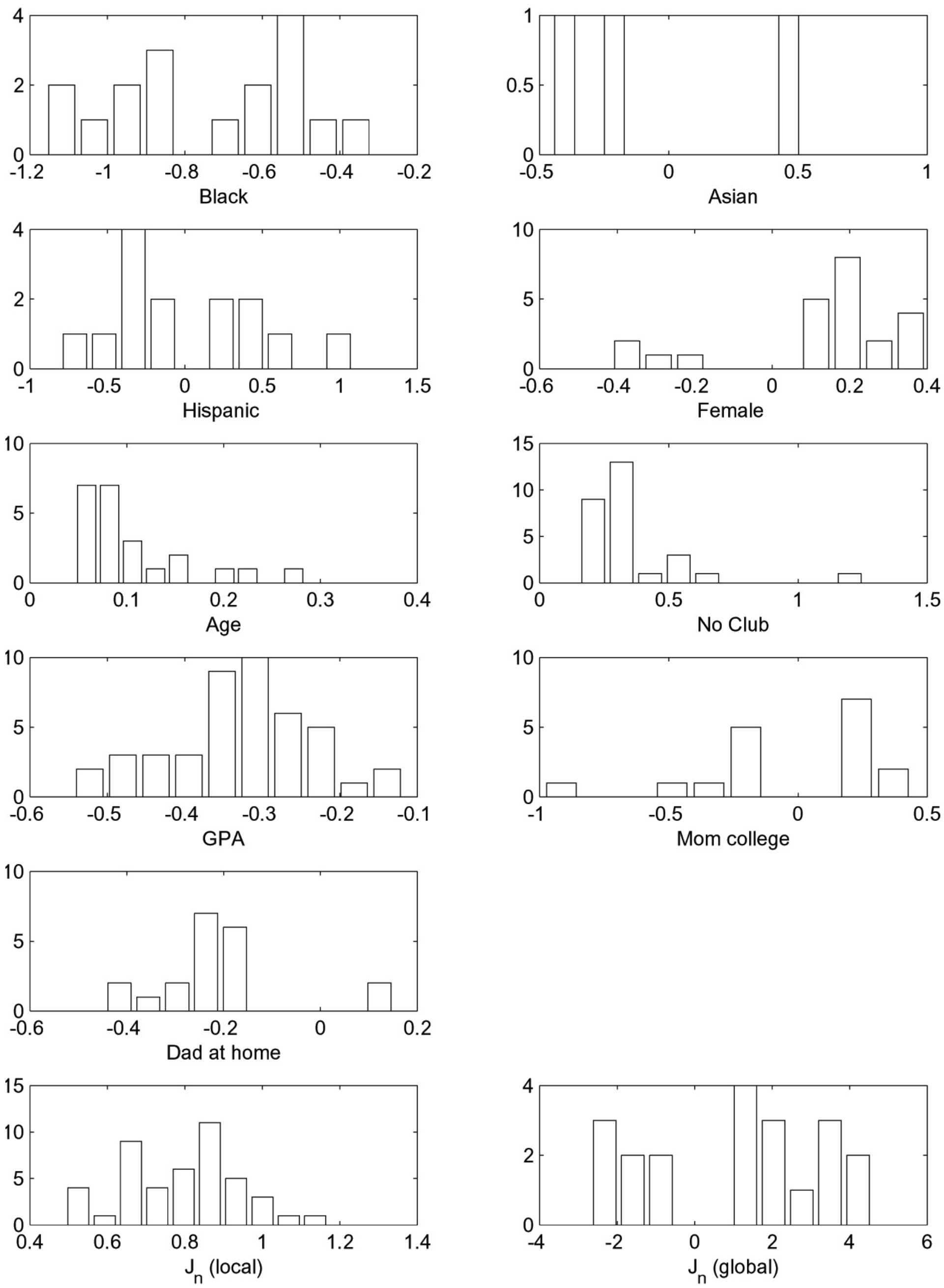

Figure 4: Distributions of parameter values, all schools separately, local and global interactions

to the racial and ethnic status variables. The results are strongest (in terms of consistency of coefficient signs across schools) for Age, Club and GPA: in almost all schools being younger (in a lower grade), belonging to a club, organization or team, and having a higher GPA are unequivocally associated with less smoking. This pattern is confirmed across all specifications we have estimated.

The school-wide interactions coefficient estimates are mostly positive and statistically significant, but again there is a wide dispersion across schools. When local interactions are omitted, the range of parameter estimates goes from about -2 to 6, indicating that some schools exhibit negative social interactions. However, positive social interactions are present in the majority of schools in our sample. This pattern is present in most specifications we have examined; as we will see below, such estimates are associated with the presence of multiple equilibria (with distinct equilibrium smoking levels) in a given school.

In the specification with only local interactions, the local interactions parameter estimates are positive in all schools and mostly statistically significant. The distribution of local interactions coefficients is tighter (less dispersed) than in the global interactions case. This finding is strongly suggestive of the presence of social interaction effects operating through individual friendship networks, although as we mentioned earlier it may also be consistent with sorting into networks along unobservable traits.

The distribution of local interaction parameter estimates does not seem to be affected much by the inclusion of the global interactions channel, whereas the estimated distribution of school-wide social effects exhibits a larger number of schools with negative global interactions than in the absence of local interactions. The medians reflect this pattern, with the median global effect falling from 1.93 to 1.41 in Table 4 while the median local interaction effect stays roughly the same at 0.82 in both specifications where it is present.

Therefore, we find a sizeable number of schools that exhibit both *positive* local interactions, and *negative* school-wide effects. We conjecture that these schools may be more stratified along racial, ethnic, or socio-economic status lines, and/or may exhibit more segregated personal networks. This finding highlights the unique nature of the Add Health data, since it allows one to contrast social interactions occurring within individual networks to those occurring within larger reference groups. Studies that do not use data containing individual network data would not be able to detect these stark differences in the nature of the interactions occurring for different definitions of reference groups[25].

5.3.2 Multiple schools

Having estimated (c_n, J_n) independently across schools, we now report the estimation results for specifications that impose some functional form on the way in which the model coefficients vary as a function of observed school-level characteristics. To do so, we estimate our social interactions model jointly for all schools in our sample, where the overall log likelihood is the sum of the individual schools' contributions.

First, as a baseline, we report estimation results for a specification in which the model parameters (c_n, J_n) are the same across all schools. This specification is obviously not a good fit of the data, since we have shown in the previous discussion that the distribution of parameter estimates across schools exhibits a significant amount of dispersion for all variables. All the same, this is a good robustness check to see if our estimation results seem consistent across specifications.

Table 5 collect results for the case with school-wide interactions only and that with both school-wide and local interactions, respectively. The estimates are qualitatively similar across the two cases, and with the distribution medians reported for the unconstrained cases in the previous Section 5.3.1. Again minorities smoke less; female and older students are more likely to smoke; students with higher GPA's, who participate in school organizations or teams, and whose fathers are present at home smoke less. Interestingly, as in the previous Section, the introduction of local interaction effects alongside school-wide effects lowers the estimated school-wide effects: the estimated J-school-wide falls by about half, from 1.34 to 0.67, while the local interaction term is equal to 0.82.

[25] See Manski (1993), on the importance of having accurate data on the extent and definition of relevant reference groups.

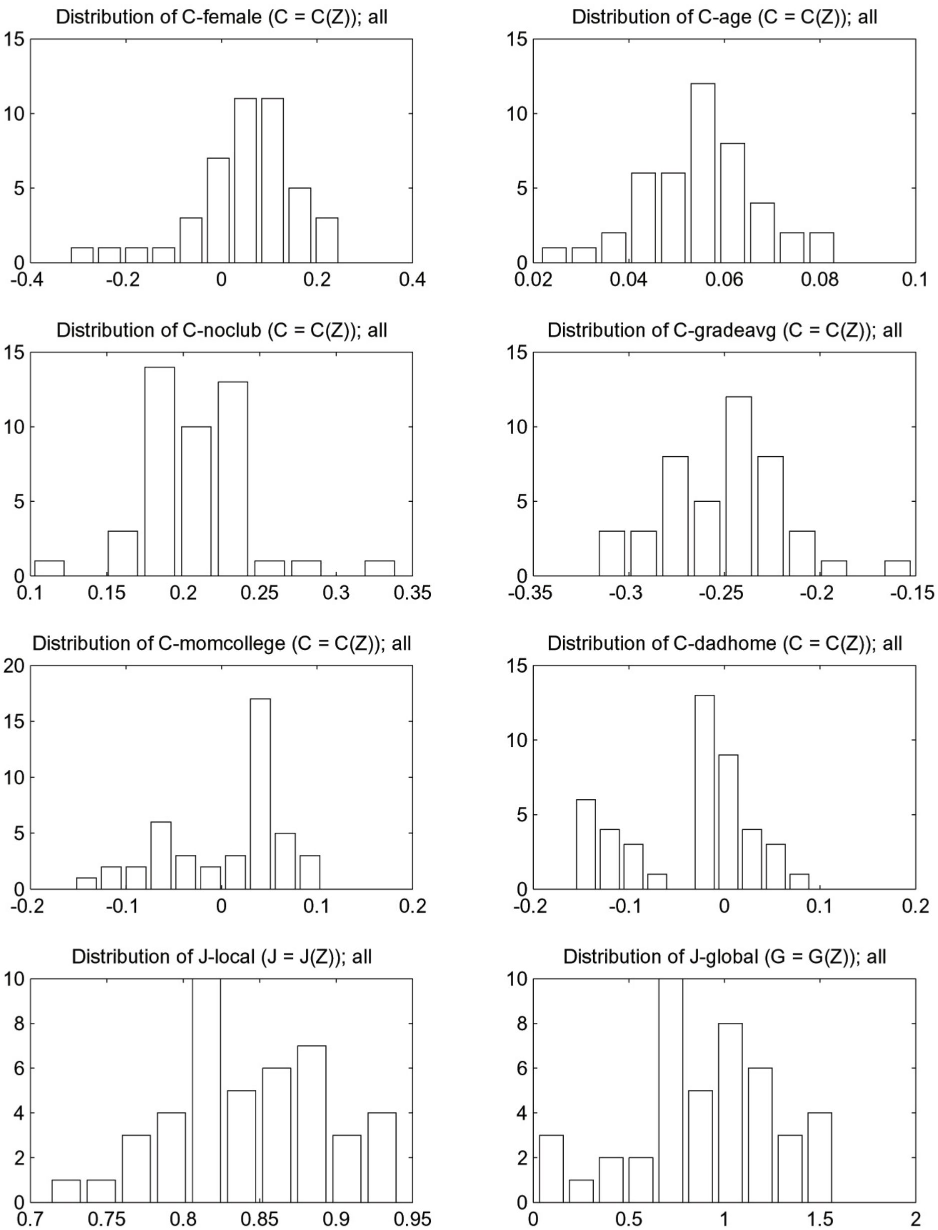

Figure 5: Distributions of parameter values, all schools, deterministic coefficients function of school characteristics

Variable	Coefficient	Std. err.	Coefficient	Std. err
Black	-0.6926	0.0343	-0.4429	0.0357
Asian	-0.1315	0.0372	-0.0712	0.0412
Hispanic	-0.2551	0.0261	-0.1664	0.0314
Female	0.0961	0.0154	0.0716	0.0177
Age	0.0684	0.0025	0.0530	0.0028
Does not belong to a club	0.3013	0.0190	0.2028	0.0221
GPA	-0.3334	0.0099	-0.2654	0.0110
Mom college	-0.0004	0.0173	0.0163	0.0195
Dad at home	-0.1205	0.0202	-0.0825	0.0222
Global interactions	1.3388	0.0585	0.6663	0.0677
Local interactions			0.8249	0.0155
Log Likelihood	-12935.04		-11253.67	

Table 5: All schools, constant parameters

Next, we turn to a specification in which parameters – while still deterministic – are specified as a function of observed school attributes, Z_n, as in equations (9) and (10). We have chosen the following list of attributes describing the presence of tobacco-related policies at a given school: whether the school enacts state-mandated training on the use and consequences of tobacco products, whether the school has implemented rules regarding the use of tobacco products by students, and whether the school has implemented rules regarding the use of tobacco products by its staff. We have also added the following list of attributes pertaining to the county in which the school is located: whether the county is urban or rural, the percentage of families under the povery line, the fraction of college educated in the population 25 and older, the fraction of female (male) adults in the labor force.

The estimation results are reported in Figure 5 and Table 6. The first observation is that once again, the distributions of estimated coefficients across schools reported in Figure 5 are qualitatively similar to those in the unrestricted specifications in Figure 4. However, the parameter estimate distributions tend to exhibit less dispersion across schools than the unconstrained ones; for instance, the distribution of the Age parameters in Figure 5 ranges from about -0.15 to about -0.32, whereas in the unrestricted case it ranges from about -0.1 to -0.5. The same is true for most of the other attributes as well as for the local interaction parameters. This observation motivates our use of the random coefficients specifications, to better capture the wide dispersion across schools of the coefficients associated to individual attributes. Further, the distribution of school-wide interaction parameters only ranges from 0 to about 1.5 and does not include any negative values, unlike the unrestricted case.

Some of the parameter estimates for the $c(Z_n)$ and $J(Z_n)$ functions are noteworthy. For instance, the mostly positive coefficients associated to female students are greatly reduced in neighborhoods with high poverty levels and high female labor force participation, suggesting that female smoking may be related to higher socio-economic status. The positive relationship between student age and smoking is again stronger in high poverty areas. The negative association between academic achievement and smoking is stronger in more highly educated neighborhoods. The finding that Dad's presence at home is associated with less smoking is reinforced in neighborhoods with high female labor force participation, perhaps indicating that one parent's presence and control is even more crucial when the other parent works outside the home.

Interestingly, a school's tobacco-related policies can have a large impact on the strength of the social interaction terms. The presence of tobacco rules for students is associated with lower school-wide interactions parameters, whereas tobacco rules for the staff seem to increase the strength of school-wide interactions but slightly reduce the strength of local interactions. Of course these school policy variables are largely endogenous, but the fact that tobacco policies are related to stronger or weaker social interaction terms is important and we will come back to it in our discussion of counter-factual experiments.

Finally, it is worth noting that neighborhood poverty levels have a huge impact on

	Female	Age	No club
	0.4412 (0.1274)	-0.0055 (0.0063)	0.2358 (0.2049)
% Urban	-0.0081 (0.0637)	0.0128 (0.0079)	-0.0004 (0.0731)
% Poverty	-1.2608 (0.3502)	0.1345 (0.0430)	0.0356 (0.4976)
% College over 25	0.3816 (0.3348)	0.0552 (0.0216)	0.3184 (0.4615)
Female Labor Force P.R	-0.5580 (0.3287)	0.0369 (0.0131)	0.5364 (0.4951)
Male Labor Forece P.R.	0.0063 (0.1833)	-0.0171 (0.0076)	-0.5182 (0.3240)
Tobacco training	-0.0781 (0.0385)	0.0146 (0.0060)	0.0006 (0.0547)
Tobacco student policy	0.1094 (0.0820)	-0.0176 (0.0070)	0.2000 (0.1112)
Tobacco staff policy	-0.1404 (0.0776)	0.0275 (0.0064)	-0.1699 (0.1027)

	Grade point average	Mom college	Dad Home
Intercept	-0.1204 (0.0342)	-0.2170 (0.1730)	0.0169 (0.1043)
% Urban	0.0238 (0.0325)	0.1395 (0.0724)	-0.0034 (0.0719)
% Poverty	0.1027 (0.1801)	-0.2135 (0.4198)	0.2747 (0.4173)
% College	-0.3549 (0.1238)	-0.3789 (0.3567)	0.2357 (0.3816)
Female Labor Force P.R.	-0.1927 (0.0764)	-0.1708 (0.4820)	-0.9254 (0.2598)
Male Labor Force P.R	0.0487 (0.0455)	0.5053 (0.2663)	0.2454 (0.1423)
Tobacco training	-0.0029 (0.0258)	-0.0218 (0.0458)	0.1247 (0.0484)
Tobacco student policy	-0.0301 (0.0421)	0.0130 (0.1161)	-0.1311 (0.0970)
Tobacco Staff policy	0.0171 (0.0389)	-0.0360 (0.1169)	0.1392 (0.0907)

	Local interactions	Global Interactions
Intercept	0.9120 (0.1332)	-0.2492 (0.1553)
% Urban	-0.0844 (0.0549)	0.5883 (0.1727)
% Poverty	-0.4960 (0.3429)	3.8357 (0.8214)
% College	0.0548 (0.3040)	0.0668 (0.5481)
Female Labor Force P.R.	-0.0976 (0.3559)	-1.5061 (0.2835)
Male Labor Force P.R	0.2248 (0.2083)	0.6361 (0.2075)
Tobacco training	-0.0416 (0.0357)	0.5226 (0.1460)
Tobacco student policy	-0.0099 (0.0741)	-0.5529 (0.1555)
Tobacco Staff policy	-0.0790 (0.0749)	0.7528 (0.1464)

Table 6: All schools, deterministic coefficients function of school characteristics (standard errors in parenthesis, log likelihood -11130.3049)

school-wide interactions estimates. This suggests that some of the variation in smoking across schools that is unexplained by observed student attributes and is attributed in the estimation to school-wide interaction effects may be in fact a school-wide fixed effect related to the area's socio-economic status. Therefore, as mentioned earlier, it stresses the value of having individual friendship network data to estimate local social interaction effects.

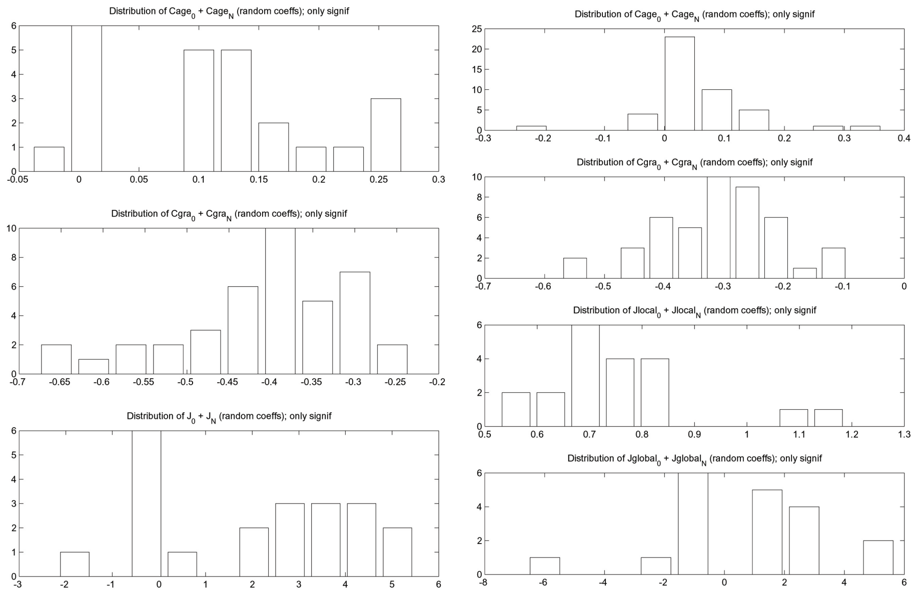

Figure 6: Distributions of parameter values, random coefficients. Left three panes: global interactions only; right four panes: both local and global interactions

5.3.3 Multiple schools, random coefficients

As mentioned in the previous section, letting (c_n, J_n) depend on observed school characteristics in a deterministic fashion might not be sufficient to capture the wide variation of parameter estimates across schools. Therefore, we also use the random coefficient specification described in Section 5.1, to better capture the dispersion in coefficient values across schools. We first estimate a version with only an intercept and a random term, as in (11)-(12), and then augment it with school level characteristics Z_n as in (13)-(14). For computational feasibility, we only use two individual student attributes, namely age and grade point average, since the introduction of random coefficients raises the number of parameters to be estimated considerably.

Figure 6 reports estimated distributions for two specifications without Z_n, in the case with only school-wide interactions (left three panes) and with both school-wide and local interactions (right four panes). Results are qualitatively consistent with the specifications

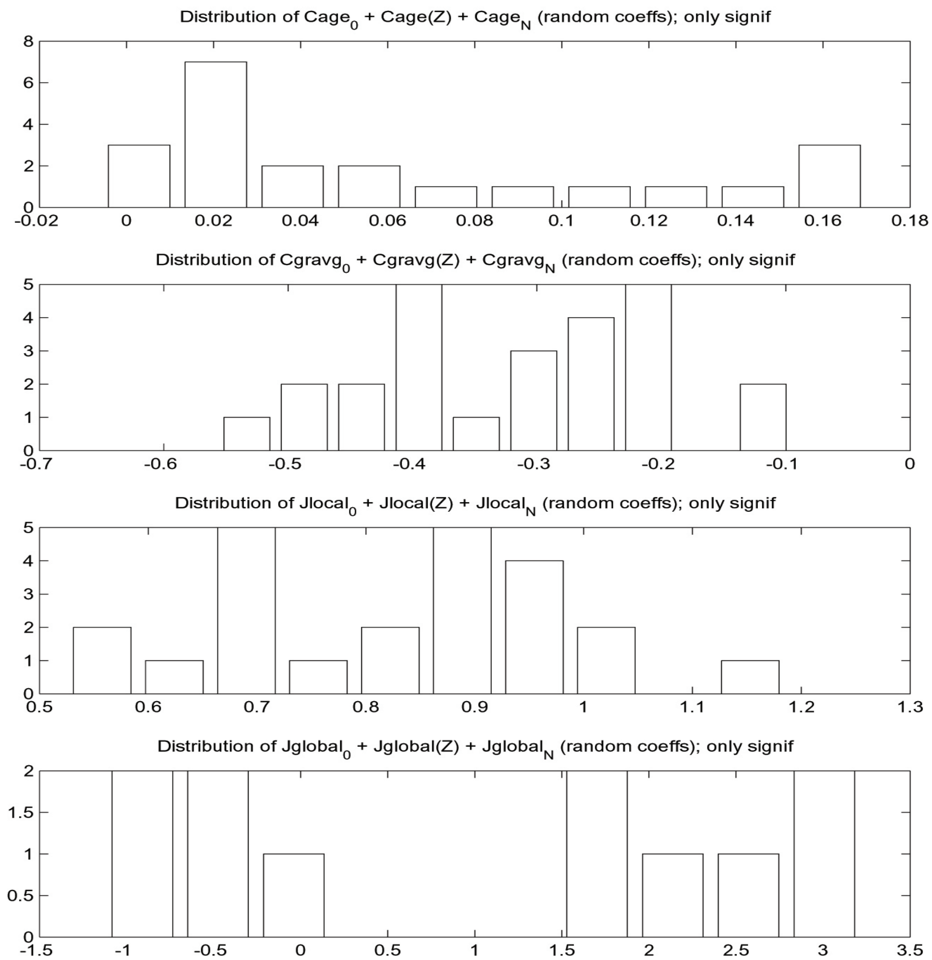

Figure 7: Distributions of parameter values, random coefficients and school characteristics

estimated so far. As expected, the range of values for the age and GPA coefficients is wider than in the deterministic case (Figure 5) and very similar to that of the unrestricted school by school specifications (Figures 2 and 4). The distributions of the school-wide only, and school-wide/local parameter estimates are also very similar to those in the unrestricted cases. Again, introducing local interactions effects generally lowers the school-wide interactions estimates, with some schools exhibiting negative school-wide coefficients.

Figure 7 and Table 7 report estimation results for the full model with random coefficients and school attributes Z_n. The Table only reports the parameters of the deterministic portion

	Age	GPA	Local Inter.	Global Inter.
Intercept	0.041 (0.005)	-0.167 (0.028)	1.919 (0.121)	1.564 (0.155)
% Urban	0.048 (0.008)	0.011 (0.040)	-0.229 (0.135)	0.202 (0.193)
% Poverty	0.021 (0.041)	0.228 (0.204)	-3.694 (0.655)	-2.379 (0.950)
Female Labor FPR	0.013 (0.013)	0.052 (0.061)	-0.823 (0.265)	-0.218 (0.318)
Tobacco training	-0.000 (0.006)	-0.087 (0.031)	-0.052 (0.114)	-0.171 (0.152)
Tobacco stud. policy	-0.009 (0.007)	0.055 (0.034)	0.100 (0.108)	-0.663 (0.169)
Tobacco staff policy	-0.007 (0.007)	-0.203 (0.038)	-0.166 (0.127)	-0.021 (0.172)

Table 7: All schools, random coefficients (standard errors in parenthesis, log likelihood -11034.074)

of the $c(Z_n)$ and $J(Z_n)$ functions, i.e. the $(\alpha_0, \alpha_k, \gamma_0, \gamma_k)$ parameters. Again, the estimated distributions in Figure 7 are remarkably similar to those estimated in the unrestricted case, with positive association between smoking and age, negative association with academic achievement, significantly positive local interactions effects and a bimodal distribution for school-wide effects (with negative coefficients in some schools).

Table 7 also shows roughly similar patterns to the deterministic specification in Table 6: the age effect on smoking is higher in high poverty neighborhoods; the negative coefficients associated to grade average are weakened in high poverty areas (which moves in opposite direction to education levels). Remarkably, tobacco rules for school staff again reduce the strength of local interactions, whereas tobacco rules for students weaken the estimated school-wide interactions estimates. As we will see in the next Section however, such policy effects do not necessarily imply less smoking.

6 Counterfactual experiments

In this section we use our estimation results to perform two sets of exercises. Firstly, given the estimates, we wish to compute the set of possible equilibria for all schools, to see whether in fact the estimated model parameters lie in regions of the parameter space

School ID	N	Global int. (J)	Percentage of smokers				
			Data	Simulated eq.	Equil. 1	Equil. 2	Equil. 3
77	1191	1.51	10.3	10.1	10.5	42.2	93.1
56	1243	3.07	20.1	17.9	3.9	19.2	99.9
57	504	0.55	23.4	23.4	24.0	-	-

Table 8: Equilibrium smoking levels in a sample of schools

that give rise to multiplicity of smoking equilibria. A corollary of this analysis is that we are able to determine whether a given school is in the equilibrium with the lowest possible smoking prevalence, or if there are other equilibria for that school and with the same parameter estimates that exhibit lower levels of smoking. Secondly, we want to perform counter-factual exercises to simulate the effect of lowering or increasing the strength of the interactions parameters on smoking. Since we have seen that some tobacco policies are associated with lower or higher interactions effects, this seems like an interesting analysis to gain some insight on the possible effects of such policies.

6.1 The prevalence of multiplicity

We use the model specification with random effects (no school covariates Z_n) and school-wide interactions only, to perform our simulations. This specification makes it manageable to compute the equilibrium mapping because it uses only two student attributes, thus making it easier to numerically integrate over the empirical distribution of these covariates (see equation (4)). At the same time, we have seen that this specification seems to capture well the variability of parameter estimates found in the unrestricted cases. Here we focus on the school-wide interactions case as an illustration, again for computational ease. Later we study the case with both local and school-wide interactions.

Table 8 reports the equilibria computed for a sample of schools, as an illustration (results for all schools are in Appendix B). The first and second columns report school ID and number of observations; the third column contains the estimated school-wide interactions

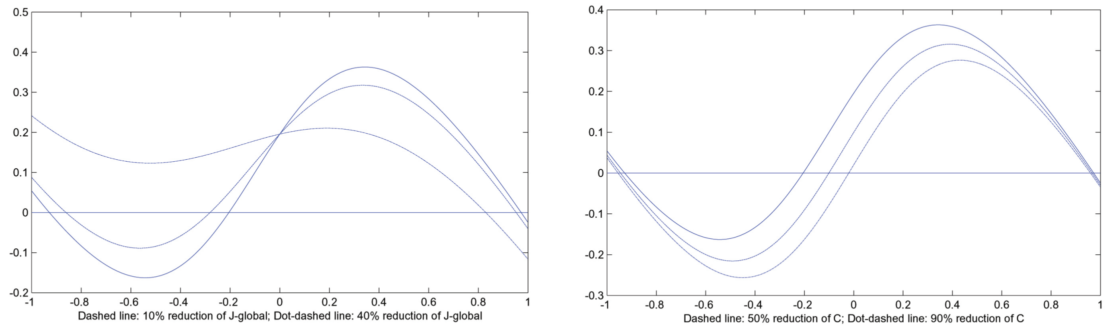

Figure 8: Equilibrium mapping as a function of global interactions J (left pane) and "cost" C_n (right pane). The mapping becomes flatter' as J decreases and moves down as C_n decreases

parameter; average smoking in each school in the data is reported in column 4; column 5 reports the simulated smoking prevalence given the equilibrium that is closest to the average smoking level in the data; columns 6-8 report the computed equilibria in the case of multiple equilibria, or the single equilibrium that arises for that school.

Actual smoking averages are consistent with one of the simulated equilibria in all cases; small differences are due to the fact that we had to discretize the support of the GPA variable for computational reasons. So this model specification captures well the large variation in smoking behavior across schools. Multiple equilibria are present in 30 out of 45 cases, given our parameter estimates (see, e.g, schools #77 and 56 in the table). This validates our approach as it shows that in this application and given the data multiple equilibria are prevalent. Finally, in 21 out of 30 schools, the actual fraction of smokers is consistent with the intermediate equilibrium (this is the case for schol #56). Thus in roughly two out of three schools where multiple equilibria arise, there exists one other equilibrium with a lower (and typically substantially lower) smoking prevalence. This is again very important from a policy perspective as it raises the question of whether it is feasible to move a school from one equilibrium to another, and if so how.

Before turning to the counter-factual exercises, we illustrate the effects of changes in the

School ID	Interactions	Percentage of smokers			
		Data	Equil 1	Equil 2	Equil 3
	Baseline	10.3	10.5	42.2	93.1
77	J reduced 5%		13.5	40.4	91.6
	J reduced 10%		18.1	37.2	89.7
	Baseline	20.1	3.9	19.2	99.9
56	J reduced 5%		7.1	13.8	99.9
	J reduced 10%				99.9

Table 9: Simulations with global interaction estimates

strength of social interactions J_n in Figure 8. The dotted line in the left pane shows how the equilibrium mapping moves for a representative school following a 10% reduction in the level of J_n (the mapping has been "rotated" so that the 45 degree line is horizontal; the axes domain is $\{-1, +1\}$, with -1 corresponding to no students smoking and +1 to all smoking). The mapping becomes "flatter", and as a result the equilibria (points where the mapping crosses the 45 degree line) move in interesting ways: specifically, the low equilibrium moves up, whereas the intermediate and the high equilibria both move down. Thus, depending on whether a given school is in the low or the intermediate equilibrium, the same reduction in social interactions may increase or decrease the equilibrium level of smoking (assuming the school stays in the same equilibrium). In addition, as J_n decreases even further (dash-dotted line), the equilibrium mapping ends up crossing the 45 degree line only once (at the previous "high" equilibrium), thus inducing a very large change in smoking prevalence.

These patterns are confirmed in our counter-factual simulations. For simplicity, Table 9 reports two examples to illustrate our findings (the full results of our exercise are available from the authors). The top panel of Table 9 focuses on school #77. Here the prevalence of smokers in our sample is 10.3%, and the school is in the low equilibrium. As the strength of social interactions is reduced by 5% and 10%, the fraction of smokers in the low equilibrium actually rises to 13.5% and 18%, respectively (consistently with Figure 8, smoking declines in both the intermediate and the high equilibria). The bottom panel tells a different story: here, because the school finds itself in the intermediate equilibrium, a 5% reduction in J_n is accompanied by a reduction in smoking prevalence (from 19.2% to 13.8%). A further reduction in J_n to 90% of its starting level makes two of the three equilibria disappear, leaving only the high smoking one: thus the school jumps to a situation in which almost every student smokes. Interestingly, if social interactions are turned off entirely (i.e., reducing J_n to zero), smoking prevalence actually declines to 87.4%.

These strong non-linearities in the effect of a reduction in social interactions stress the importance of estimating a structural model that explicitly takes into account the possibility of multiple equilibria induced by the positive feedbacks among agents' actions. These are very important from a policy perspective: as a mere illustration, if the adoption of rules for tobacco use by school staff indeed reduces the strength of school-wide social interactions, such a policy may have the unintended consequence – in some schools – of actually increasing smoking prevalence in the school.

A final caveat concerns equilibrium selection. Both the model we use in this application and our estimation approach are silent on the actual mechanism through which an equilibrium is selected. It could very well be that equilibria are "sticky" and that it is very difficult to move a school from one equilibrium to another. Even so, our illustration shows that the same policy may have very different effects in different schools, depending on the specific equilibrium the school finds itself in.

6.2 Counterfactual experiments with both local and school-wide interactions

Given our school-by-school estimates with the entire set of demographic controls, we focus on a single school and perform two sets of simulation exercises using the Brock-Durlauf (BD) model with both local and school-wide interactions. First, given our estimates for a representative school, we simulate the smoking process over time for a variety of different scenarios: changing the intensity of local and/or school-wide interactions; introducing a tax on cigarettes; changing the number of friends in students' friendship networks.

Secondly, again given our estimated parameter values, we simulate the smoking process for an artificial school, to study the impact of different network structures for instance looking at perfectly segregated vs perfectly integrated personal networks.

Because we allow for both local and school-wide interactions, it is very difficult to compute the model equilibria and to obtain a closed form solution for the stationary distribution of the model. This is why we resort to simulations to compute the long run fraction of smokers. Specifically, we simulate the BD model as a first-order Markov process in discrete time, where the agents' state in each period (a configuration of smokers and non-smokers in the school) is a function of the state of all agents at the previous period.

All agents change state at each period, based on the smoking configuration in the previous period. Our simulation results do not change if we instead allow only one randomly drawn agent to change her state in each period. We let the process run for many periods (2,400) and report the long run average out of the last 2,000 iterations.

6.2.1 Results for a representative school

We pick school #56 as a representative school, both in terms of demographics and in terms of its parameter estimates. Figures 9-13 report the results of our counterfactual experiments to show the importance of multiple equilibria and non-linearities. Figure 9 reports the equilibria we find for several values of the local interactions parameter (displayed in the vertical axis). The horizontal lines display the basin of attraction of the low level equilibrium. For example, at the estimated value of J, when setting an initial fraction of smokers between 0 and 16%, the procedure converged to a low-level equilibrium with 1.05 percent of students smoking. For all other levels of initial fraction of smokers, the procedure converged to the high level of smoking, always to 100% of students smoking. Figure 10 repeats the exercise by varying the global interaction parameter, Figure 11 by

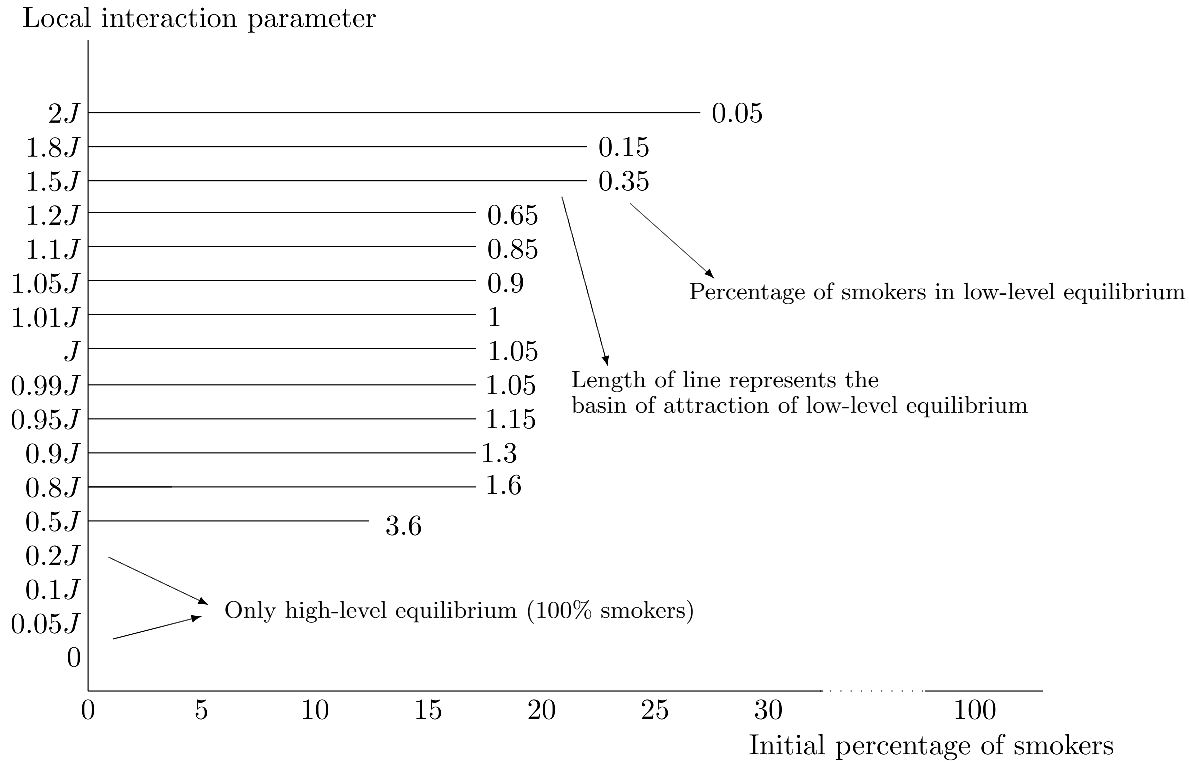

Figure 9: Percentage of smokers and basins of attraction of low-level equilibrium in a representative school for different values of the local interaction parameter (High level equilibrium at 100% smokers in every simulation)

changing the number of friends, Figure 12 by changing both local and global interactions, and Figure 13 by changing the cost of smoking.

Overall, multiple equilibria are pervasive. In our simulations, only the high and low equilibria appear in the long run. The middle equilibrium does not seem to be stable, even though for certain choices of parameter values and for certain starting points (in terms of smoking averages) the process seems to tend to an intermediate equilibrium for a few iterations. Eventually however it goes to one of the two extreme equilibria.

Changes in the strength of local and/or school-wide interactions, or in the number of friends in students' personal networks, all go in the same direction (see Figures 9-11). As we reduce J, G, or the number of friends, the basin of attraction of the low equilibrium (i.e. the set of initial conditions from which the process reaches it) shrinks, until it eventually

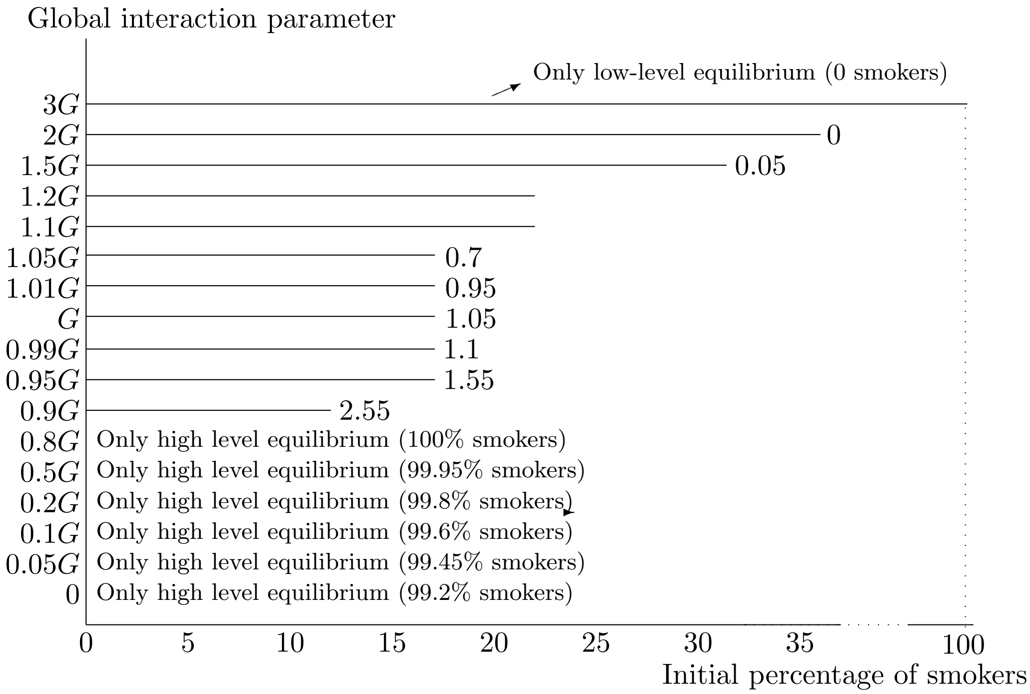

Figure 10: Percentage of smokers and basins of attraction of low-level equilibrium in a representative school for different values of the global interaction parameter. High level equilibrium at 100% smokers unless indicated

disappears; this is consistent with the left panel of Figure 8 (which illustrates movements in the equilibrium mapping following a change in G). The fraction of smokers increases slowly at the low equilibrium, until it jumps to a near-totality of smokers in the high equilibrium when the low equilibrium disappears. Local and social interactions appear to be "strategic complements" (Figure 12): keeping the strength of local (respectively, school-wide) interactions fixed, the basin of attraction of the low equilibrium shrinks as the strength of school-wide (respectively, local) interactions decreases, and viceversa.

Figure 13 reports the effect of changes in the individual cost of cigarettes (measured in "utils"): specifically, an increase in the cost of smoking is modeled as a lowering of the intercept in the random utility of smoking of each student. This experiment changes the long run level of smoking in the direction that one would expect from the right panel in Figure 8. As the utility cost of smoking goes down (i.e., the intercept of C_i rises), the basin of attraction of the low equilibrium shrinks and the fraction of smokers in the school rises slightly, until the process jumps to the high equilibrium where almost everyone smokes.

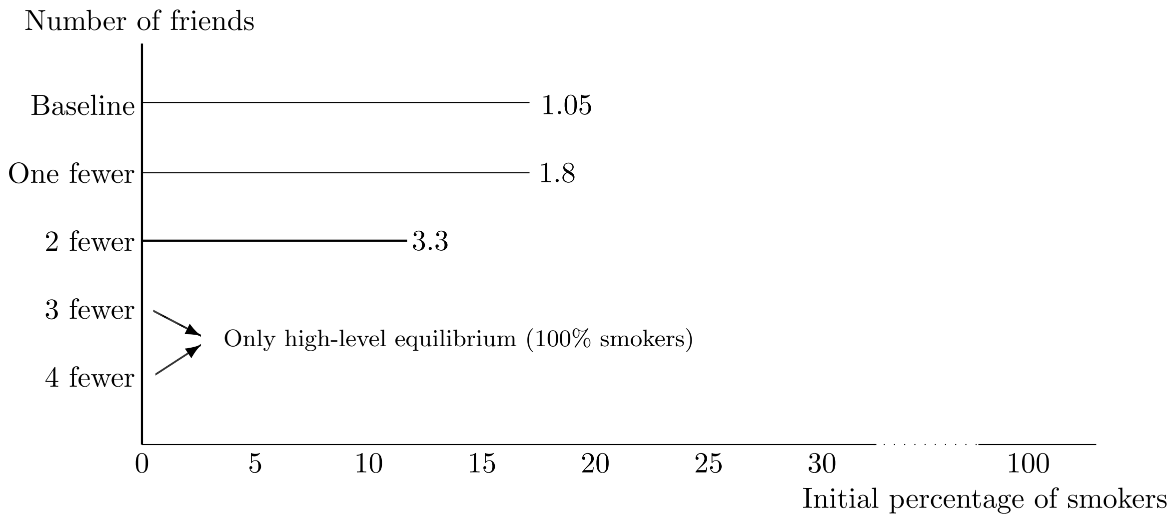

Figure 11: Percentage of smokers and basins of attraction of low-level equilibrium in a representative school for different number of friends

6.2.2 Results for an artificial school

For these experiments we construct an artificial school with similar characteristics to the actual school we used above. We consider a school with 800 students disposed on a circle. Every student i has R friends, defined as the R students directly to the right of i. The baseline number of friends is $R = 4$ (which is the median number of friends for students in our representative school #56), but we vary this parameter in the simulations. Note that friendship ties are modeled as directed links: student i "names" student j as a friend, but not viceversa. Initially all students are homogeneous, and have the same characteristics as the median student in school #56: non-Black, non-Asian, non-Hispanic, female, 16 years old; participates in school clubs and associations; has a 3.0 GPA; mom has less than a college education; dad lives at home.

Figures 14-16 report the results of our experiments on the strength of local and global interactions (changes in J, G), and on the number of friends R. The artificial school behaves very similarly to the actual one. Reductions in local interactions J, in school-wide interactions G or in the number of friends R all induce the basin of attraction of the low equilibrium to become smaller and eventually disappear. While the low equilibrium per-

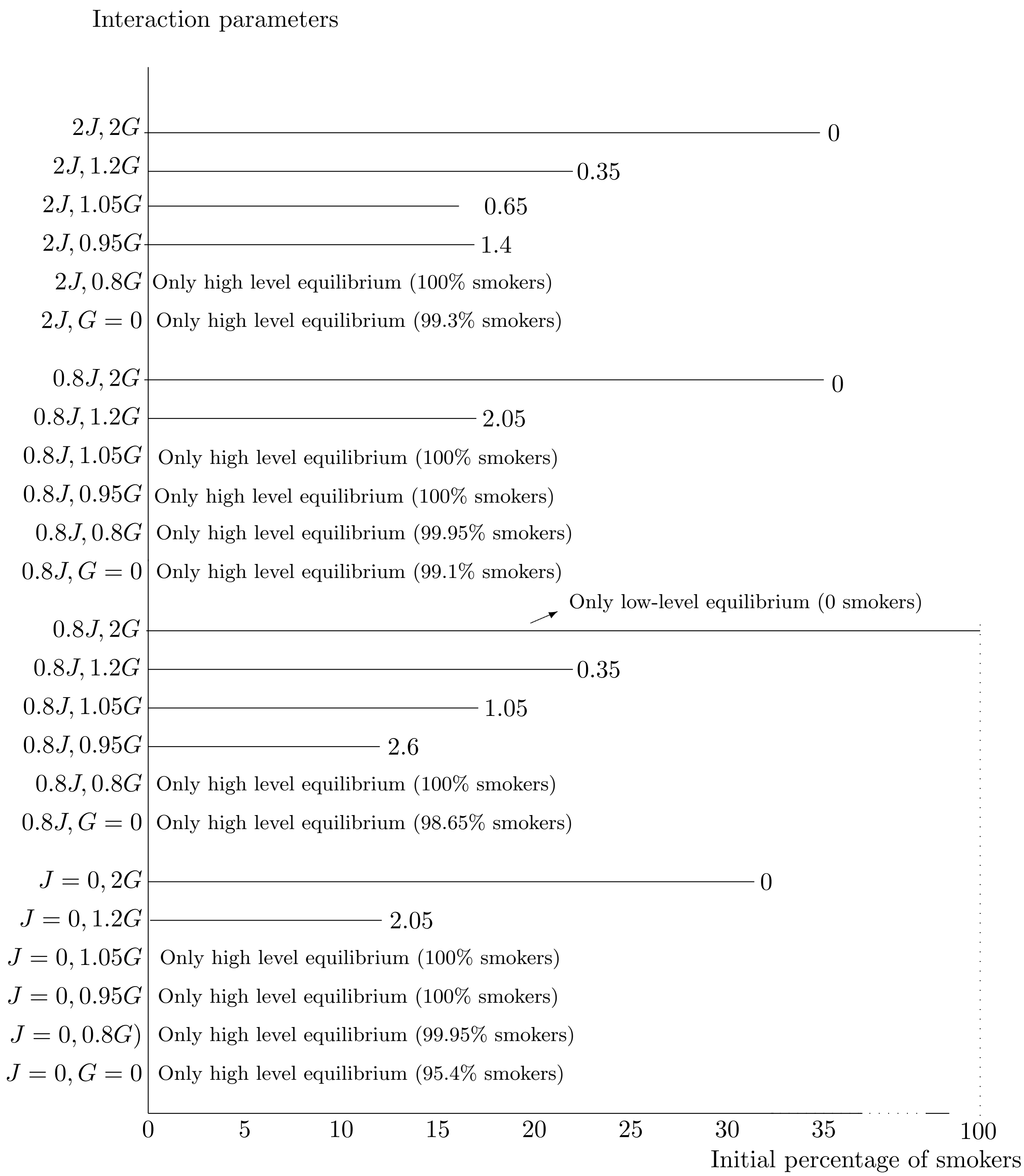

Figure 12: Percentage of smokers and basins of attraction of low-level equilibrium in a representative school for different values of local and global interactions. High level equilibrium always 100% smokers unless indicated

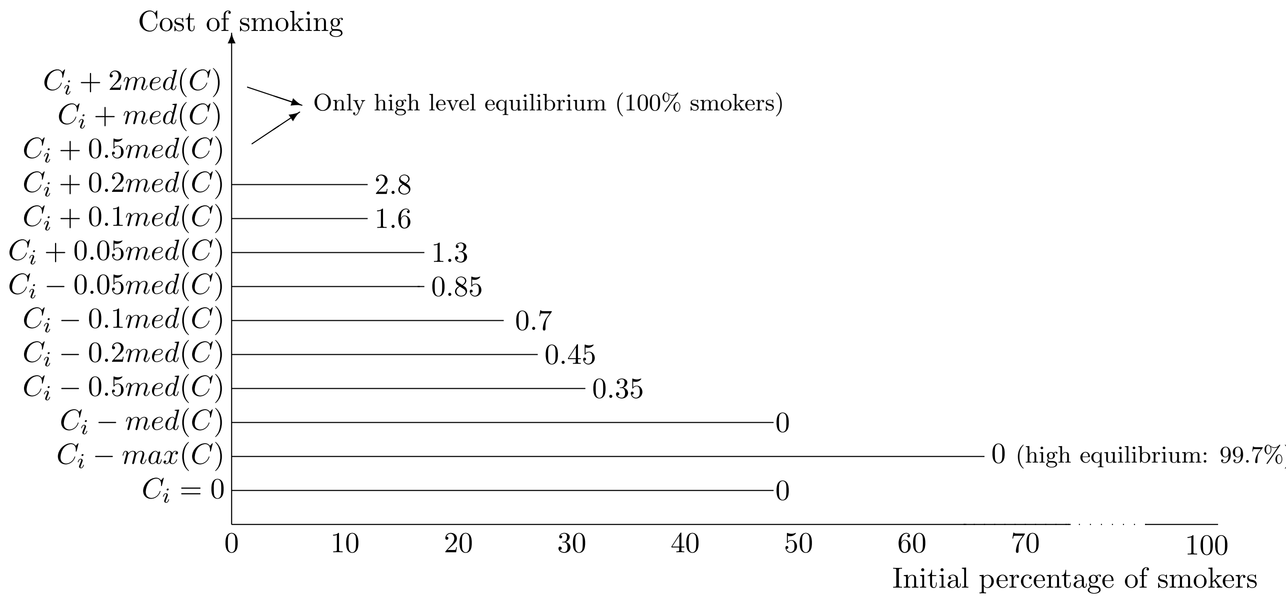

Figure 13: Percentage of smokers and basins of attraction of low-level equilibrium in a representative school for different "cost" values of smoking

sists, reductions in these parameters are associated with a slight increase in the fraction of smokers in the school. As with the actual school, holding fixed any two parameters in (J, G, R) while reducing the third one has the same effect on the basin of attraction of the low equilibrium and on the fraction of smokers in this equilibrium while it persists.

Figure 17 reports the effects of changing the demographic characteristics of all artificial students in the school; agents are still homogeneous here. These experiments are equivalent to shifting the intercept of the individual utility of smoking (conceptually, it is the same as introducing a "tax/subsidy" on cigarettes). For Black, male, or high-GPA students the intercept shifts down relative to our baseline student: this induces the basin of attraction of the low smoking equilibrium to widen, and the level of smoking in the low equilibrium falls slightly. Changing our baseline artificial student to an older student, or to a student who does not participate in school clubs, induces the opposite effect. An artificial school made up of students that combine all the demographic characteristics associated with higher propensities to smoke is characterized by the disappearance of the low smoking equilibrium.

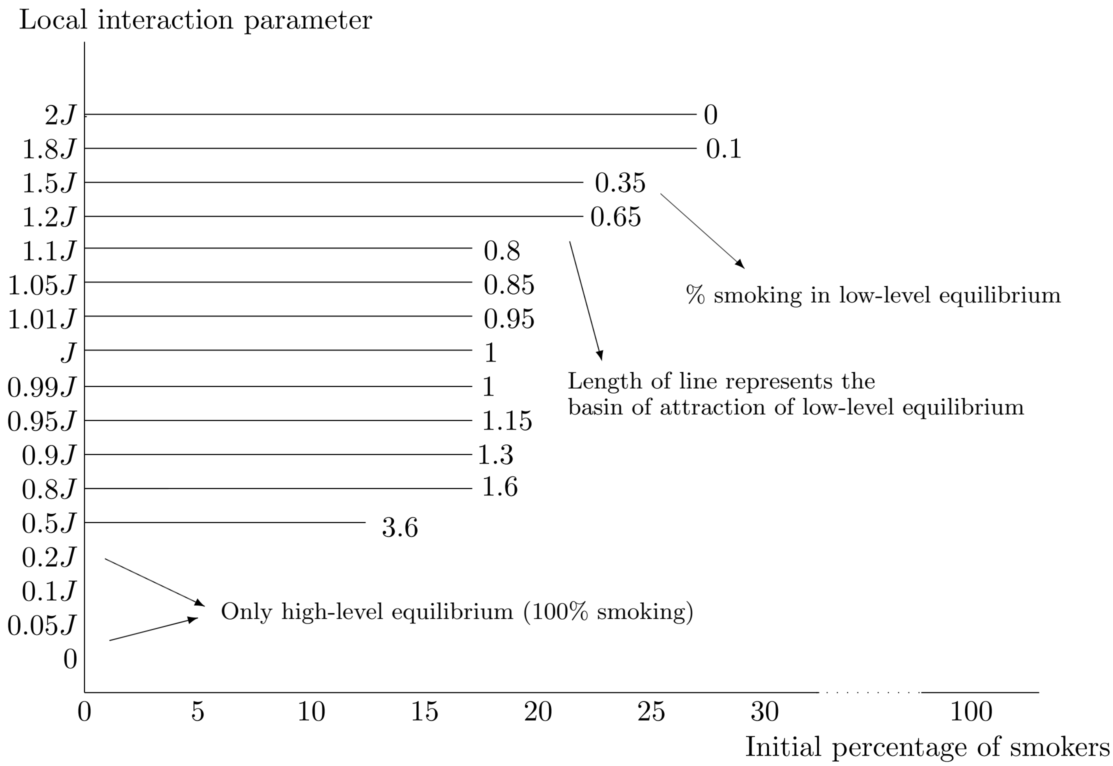

Figure 14: Percentage of smokers and basins of attraction of low-level equilibrium in artificial school for different values of the local interaction parameter. High level equilibrium at 100% smokers in all simulations

Figures 18 and 19 look at the effects of introducing two different types of students in the artificial school: type L students are the least likely to smoke based on their individual characteristics; type M students are the most likely to smoke. We vary both the fraction and the arrangement of the two types within the artificial school (i.e., along the circle). In Figure 18 we vary the relative fraction of the two types, but maintain an arrangement that is "perfectly integrated": $LMLM...$, or $LLLMLLLM...$, or $LLLLLLLMLLLLLLLM....$ In Figure 19 instead, we simulate perfect "segregation": the school is divided into two subgroups, one where everyone is of type L, the other where everyone is of type M. Local interactions only occur within the two subgroups (that is, friendship networks are perfectly segregated), but global interactions still occur school-wide (hence across groups).

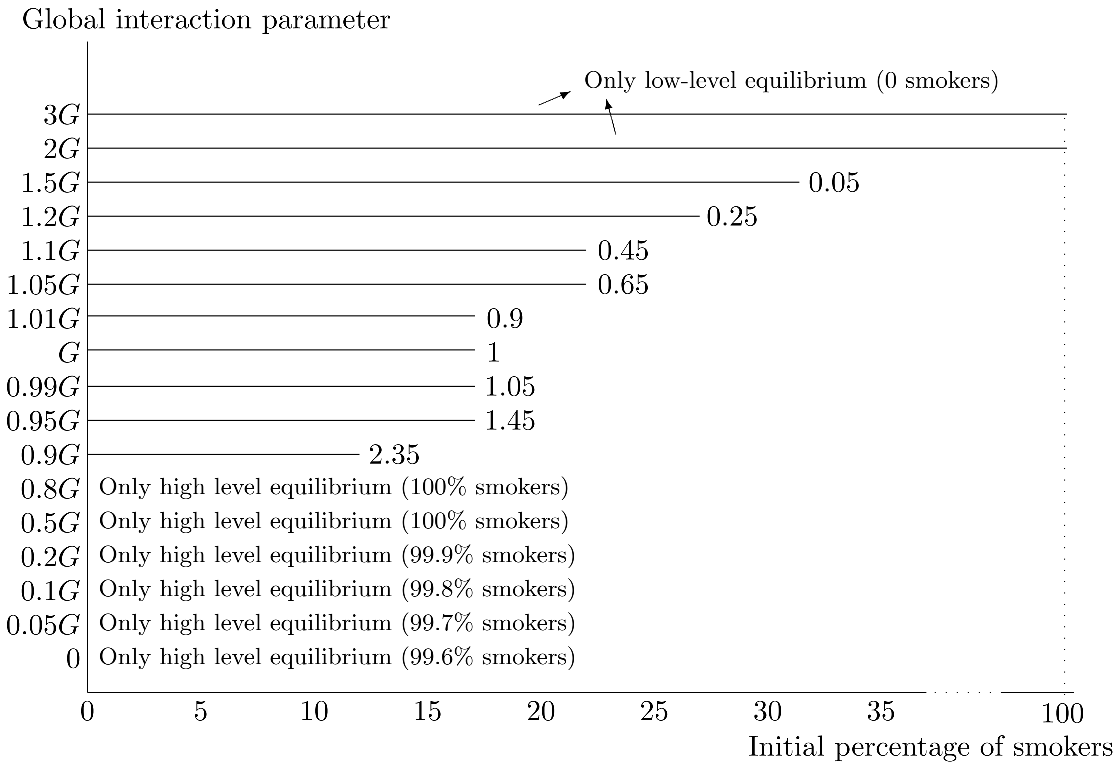

Figure 15: Percentage of smokers and basins of attraction of low-level equilibrium in artificial school for different values of the global interaction parameter. High level equilibrium at 100% smokers unless indicated

When the fraction of types is $(L = 1/2, M = 1/2)$ the "high smokers" are too many for the low equilibrium to arise, and the only surviving equilibrium is the high one where everyone smokes. This occurs regardless of the distribution of types along the circle and of the number of friends. As the fraction of M students falls, the low equilibrium starts to appear, its basin of attraction becomes greater, and the fractions of smokers in the low equilibrium falls slightly.

When the fractions are $(L = 3/4, M = 1/4)$ the arrangement of types along the circle begins to matter. As the strength of local interactions J increases relative to school-wide interactions, the low equilibrium become more likely (its basin of attraction widens). Moreover, a new, intermediate equilibrium starts appearing, determined by the fraction of (L, M)

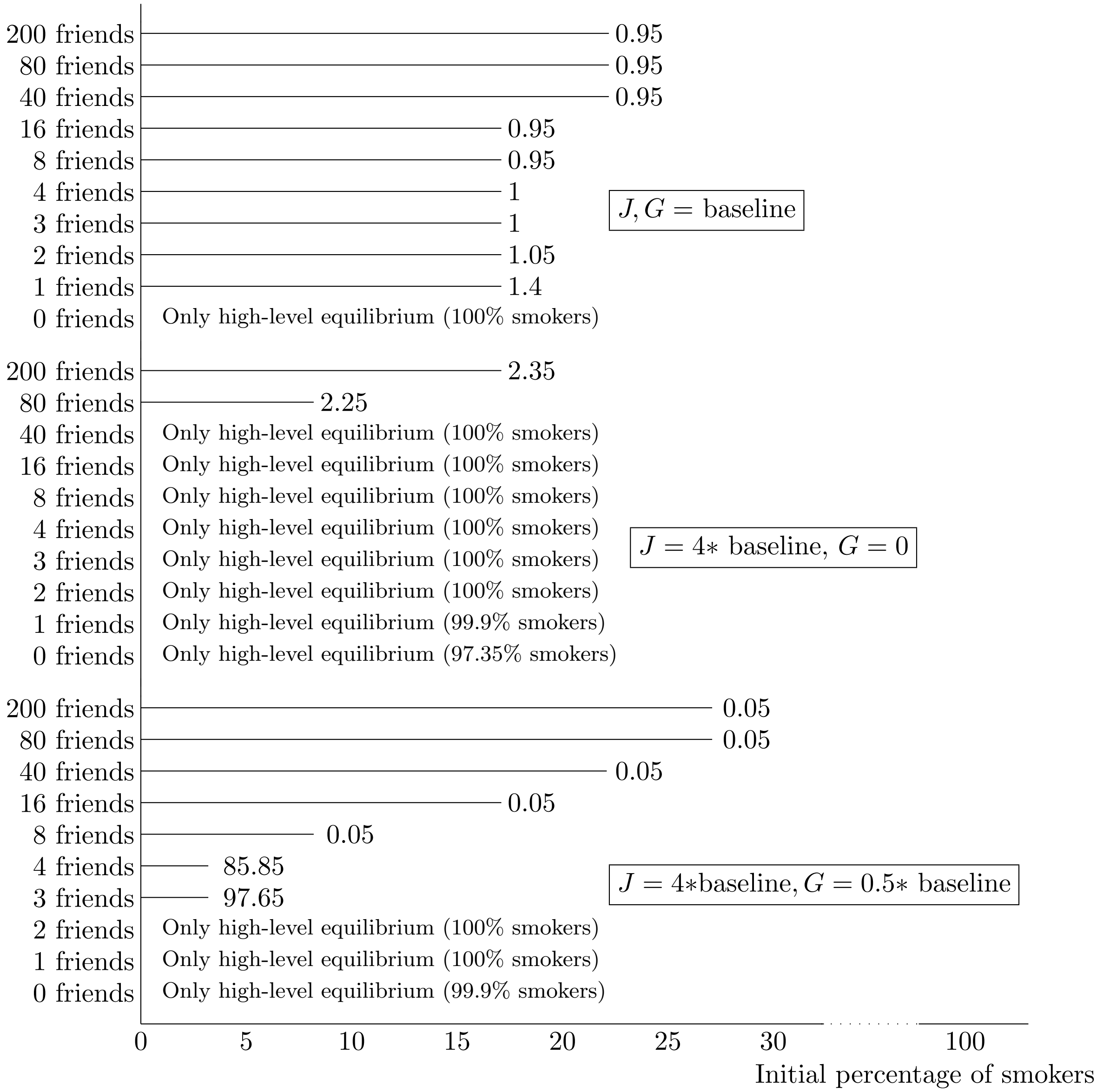

Figure 16: Percentage of smokers and basins of attraction of low-level equilibrium in artificial school for different number of friends and interaction parameters. High level equilibrium at 100% smokers unless indicated

types as shown in Figure 19. This intermediate equilibrium appears only under perfect segregation, but not under perfect integration – in the experiments we have carried out. Thus the structure of personal networks within a school (e.g., integrated vs. segregated along certain demographic traits) seems to have an impact on the smoking prevalence arising in equilibrium. Within each configuration $\{\%L, \%M, J, G\}$ as the number of friends rises the

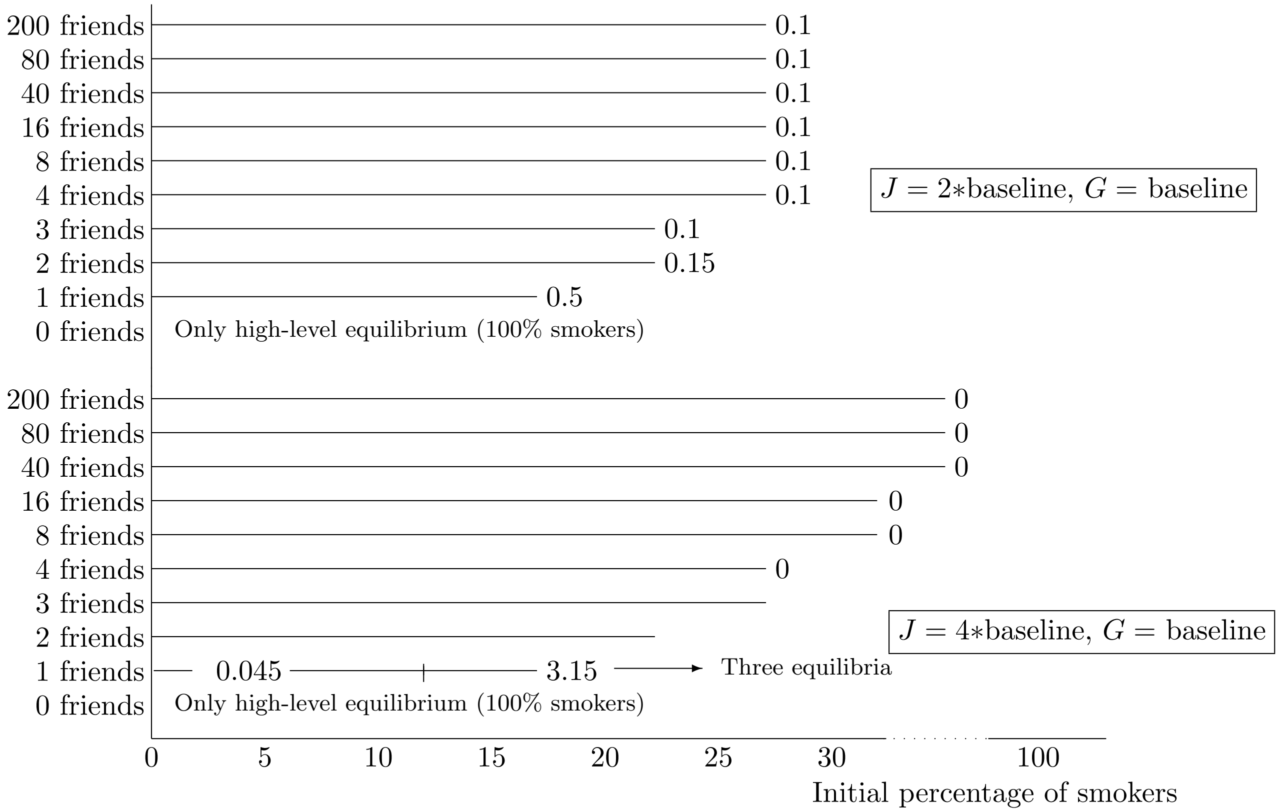

Figure 16: (continued) Percentage of smokers and basins of attraction of low-level equilibrium in artificial school for different number of friends and interaction parameters. High level equilibrium at 100% smokers unless indicated

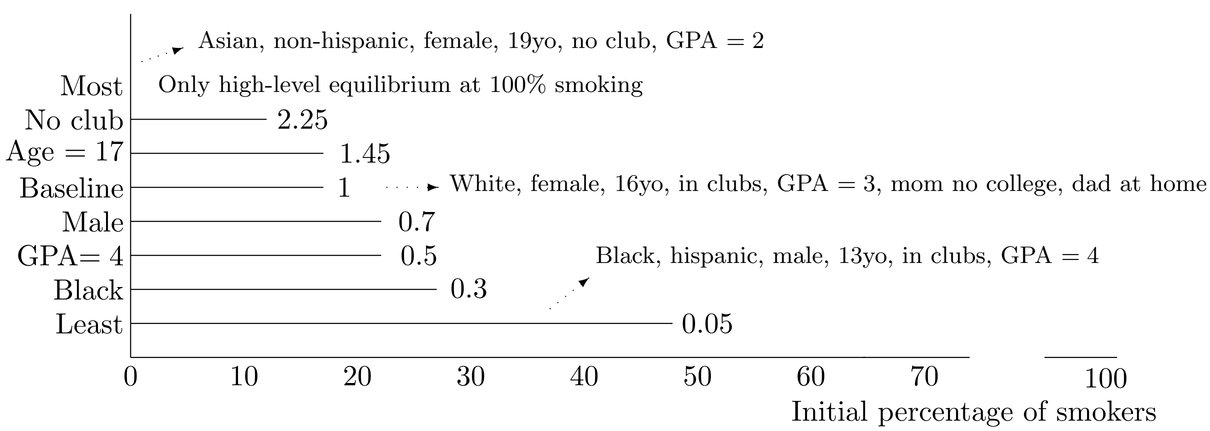

Figure 17: Percentage of smokers and basins of attraction of low-level equilibrium in artificial school for different artificial (homogenous) student. High level equilibrium at 100% smokers unless indicated

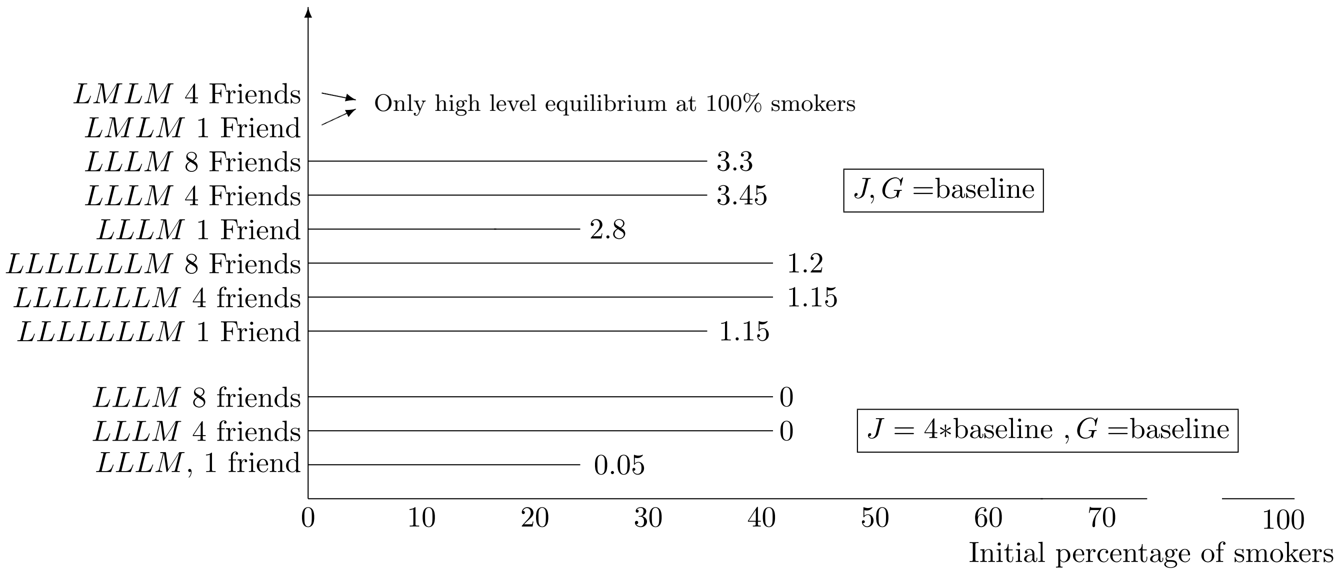

Figure 18: Percentage of smokers and basins of attraction of low-level equilibrium in artificial school for different arrangements of students on the circle. High level equilibrium at 100% smokers in all simulations

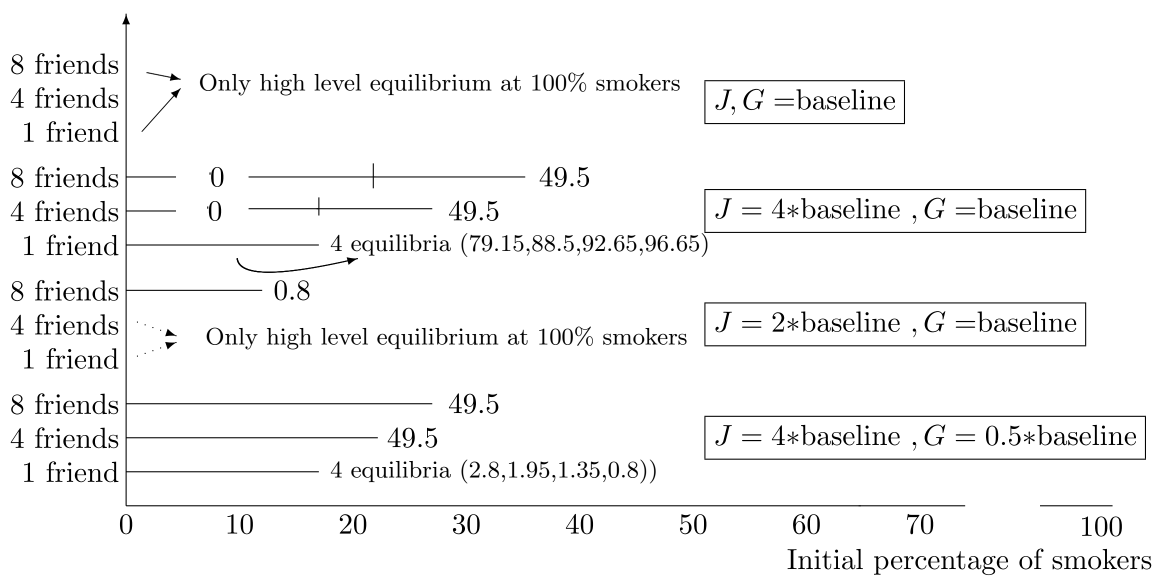

Figure 19: Percentage of smokers and basins of attraction of low-level equilibrium in artificial school for different groupings of students. Type of students: L=50%, M=50%. High level equilibrium at 100% smokers unless indicated

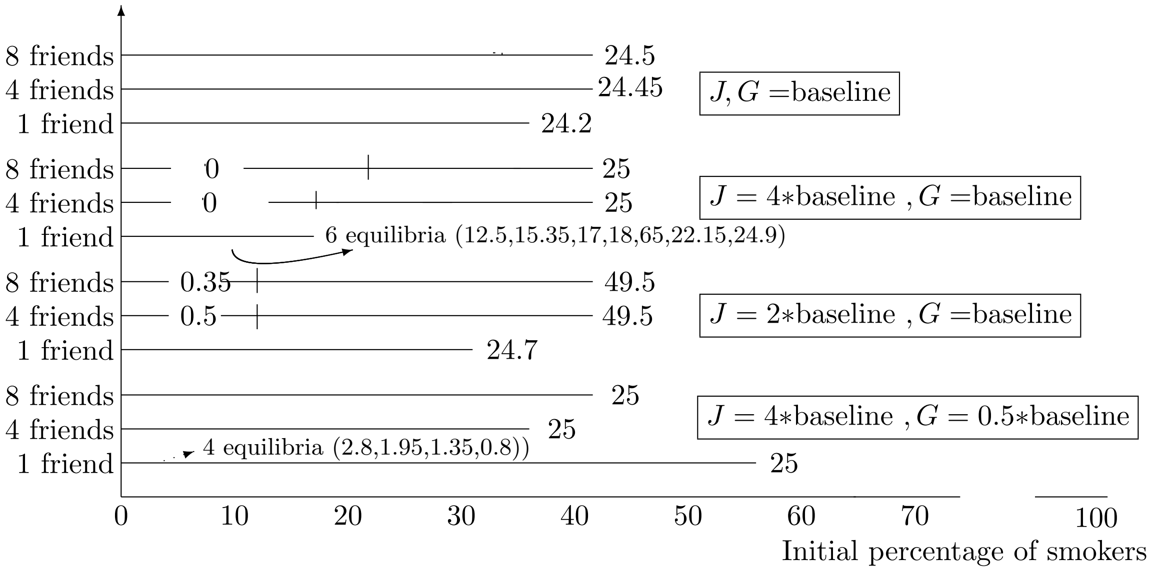

Figure 19: (continued) Percentage of smokers and basins of attraction of low-level equilibrium in artificial school for different groupings of students. Type of students: L=75%, M=25%. High level equilibrium at 100% smokers unless indicated

basin of attraction of the low equilibrium widens again.

7 Conclusion

In this paper we present a general framework to study models with multiple equilibria in economies with social interactions. We show that point identification of model parameters is conceptually distinct from the presence of multiple equilibria, and derive some general conditions for identification. We then present a two-step estimation strategy that, while less efficient than the direct maximum likelihood estimator, has two significant advantages: first, it is computationally feasible as it is several orders of magnitude faster than the direct or the Dagsvik-Jovanovic (D-J) method; second, it does not rely on making explicit assumptions about the nature of the selection mechanism across equilibria, as in the D-J method.

We then apply our estimation approach to a version of the Brock-Durlauf binary choice model with social interactions, using data on teenage smoking. We find statistically significant evidence of both school-wide and local (within personal networks) interactions. Our estimates are consistent across specifications that take into account school and local neighborhood attributes. Given our estimates, multiple equilibria are prevalent in our data: the estimated parameter values for about two thirds of our schools give rise to multiplicity. In many cases where multiple equilibria are present, there exists one other equilibrium with a lower smoking prevalence than that observed in a given school.

Having estimated the parameters of a structural model enables us to run several counterfactual experiments. In specifications with school-wide interactions only, we show that reductions in the strength of social interactions may increase or decrease smoking prevalence depending on whether the school is in the low or intermediate equilibrium. Large reductions in school-wide interactions eventually make equilibrium multiplicity disappear, with only the high smoking equilibrium surviving. Thus, tobacco policies aimed at students – which are associated with lower school-wide social interactions according to our estimates – may have the counter-intuitive and undesirable effect of actually increasing smoking prevalence in a school, sometimes by a large amount.

Simulations in settings where both school-wide and local interactions are present show that reductions in local and/or global interactions, or in the number of friends, make the basin of attraction of the low smoking equilibrium shrink until it disappears; while this equilibrium persists, smoking prevalence rises slightly. Reductions in the utility cost of smoking also make the basin of attraction of low equilibria become smaller, and smoking prevalence increase. When the low equilibrium disappears, the fraction of smokers in a school jumps up dramatically to an equilibrium where almost everyone smokes. Finally, the arrangement of students within a school, and hence the structure of personal networks (e.g. segregation vs. integration by demographics), can influence the number and types of smoking equilibria that may arise.

These results should be taken as illustratory of the kind of policy experiments that one may carry out having obtained structural parameter estimates for a behavioral model of smoking with social interactions that exhibits multiple equilibria. However, they should not be taken literally: here we assume that personal networks are given and that tobacco use policies are exogenous, the model does not incorporate dynamic features such as addiction, and we do not take a stand on the way in which agents select a specific equilibrium. This is left for future work. In a companion paper, we study settings where friendship networks are endogenous: agents sort on the basis of both observable and unobservable attributes, which in turn may be related to one's propensity to smoke. It turns out that taking this endogeneity explicitly into account can help identify the extent of social interactions.

A Additional Results of Monte Carlo Experiments

Monte Carlo single subpopulation experiment - results (intermediate equilibrium)

Evaluation Criterion	Direct	Two step	Direct, 2step initial est
RMSE, parameter C	0.07971	0.00808	0.00812
Bias, parameter C	-0.00479	0.00005	0.00008
RMSE, parameter J	0.73115	0.68944	0.68802
Bias, parameter J	0.2121	0.10599	0.10499
Min time	121.42235	0.1987	127.70132
Max time	189.26107	0.35191	188.28
Mean time	159.08873	0.27914	158.77346
Median time	159.83816	0.27635	159.29816

Monte Carlo single subpopulation experiment - results (high equilibrium)}

Evaluation Criterion	Direct	Two step	Direct, 2step initial est
RMSE, parameter C	0.04735	0.04631	0.0463
Bias, parameter C	0.00455	0.0048	0.0048
RMSE, parameter J	0.09303	0.08689	0.08692
Bias, parameter J	0.00058	0.00302	0.0015
Min time	137.35112	0.23884	135.03178
Max time	180.90666	0.3706	179.13963
Mean time	159.40678	0.30305	158.65599
Median time	159.63261	0.30731	158.13407

B Equilibrium smoking in all schools

School ID	N	Local int. (J)	Percentage of smokers					
			Data	Simulated eq.	Equil. 1	Equil. 2	Equil. 3
77	1191	1.51	10.3	10.1	10.5	42.2	93.1
56	1243	3.07	20.1	17.9	3.9	19.2	99.9
57	504	0.55	23.4	23.4	24.0	-	-
33	553	1.89	24.8	20.8	11.3	22.5	98.7
29	788	4.21	16.9	16.1	1.6	16.6	100.0
62	1763	0.43	24.8	26.8	25.3	-	-
82	526	3.95	21.9	19.8	1.3	21.3	100.0
72	1125	-1.28	23.6	21.3	23.8	-	-
76	843	2.82	17.8	17.4	5.6	17.0	99.9
44	1142	0.69	18.8	18.1	19.3	-	-
71	775	0.88	21.7	18.3	22.1	-	-
40	523	3.93	24.7	22.0	0.9	24.3	100.0
79	1142	0.52	20.1	23.9	20.7	-	-
34	470	1.55	17.4	16.8	19.6	29.2	93.6
41	1644	2.02	6.9	6.7	6.8	27.1	99.1
86	1152	0.78	19.5	18.8	20.5	-	-
259	1064	2.52	25.9	22.5	4.4	24.9	99.8
43	451	2.35	4.0	4.0	4.0	27.5	99.7
35	594	2.11	22.2	20.9	9.2	20.7	99.3
78	798	1.38	15.0	16.3	16.9	47.6	85.2
20	610	0.83	14.6	15.6	15.0	-	-
268	1001	-0.32	9.8	10.4	9.9	-	-
53	1181	0.84	20.6	20.8	20.8	-	-
193	406	3.21	13.3	12.6	5.7	12.2	100.0
42	827	2.86	19.7	20.2	4.0	19.2	99.9
74	538	2.44	23.2	21.4	5.9	22.2	99.7
50	1235	1.37	18.8	20.5	19.9	28.8	92.6
162	424	1.78	6.8	6.6	7.7	42.8	95.8
31	1086	2.02	25.0	20.4	10.0	22.8	99.0
91	987	-0.75	26.1	27.7	26.4	-	-
267	534	0.98	19.3	21.0	20.6	-	-
65	1516	1.81	7.3	6.7	7.7	34.2	97.6
67	598	1.97	25.4	25.6	8.4	24.8	99.0
13	504	6.06	15.3	15.7	0.5	15.1	100.0
23	462	2.83	23.4	20.6	3.4	22.6	99.9
15	462	1.29	16.2	16.2	17.3	-	-
25	451	2.78	13.7	7.5	7.8	12.4	99.9
52	639	0.44	7.2	5.5	7.4	-	-
58	640	-2.33	33.4	33.1	33.8	-	-
269	799	4.41	28.3	30.4	0.3	28.1	100.0
270	487	1.53	16.8	14.4	17.8	22.4	96.1
47	515	2.77	26.2	22.5	3.4	25.7	99.8
271	823	3.4	11.1	8.9	6.7	9.1	100.0
18	466	2.27	20.0	19.1	9.1	17.9	99.6
87	516	2	25.0	23.6	11.9	21.7	98.6

References

Aguirregabiria, V. (2004): "Pseudo Maximum Likelihood Estimation of Structural Models Involving Fixed-Point Problems," *Economics Letters*, 84(3), 335-40.

Aguirregabiria, V. and P. Mira (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75(1), 1-53.

Aiyagari, S.R. (1995): "Comments on Farmer and Guo's 'The Econometrics of Indeterminacy: An Applied Study'," *Federal Reserve Bank of Minneapolis*, *Research Department Staff Report* 196.

Akerlof, G., (1997), "Social Distance and Social Decisions," *Econometrica*, 65, 1005-27.

Andrews, D.W.K., S.T. Berry, and P. Jia (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location, " mimeo.

Bajari, P., H. Hong, and S. Ryan (2007): "Identification and Estimation of a Discrete Game of Complete Information," Duke University, mimeo.

Bajari, P., L. Benkard, and J. Levin (2003): "Estimating Dynamic Models of Imperfect Competition," Duke University, mimeo.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2006): "Estimating Static Models of Strategic Interactions," mimeo.

Bauman, K. and L. Fisher (1986): "On the Measurment of Friend Behavior in Research on Friend Influence and Selection: Findings from Longitudinal Studies of Adoloscent Smoking and Drinking," Journal of Youth and Adolescence, 15, 345-53.

Becker, G. (1991): "A Note on Restaurant Pricing and Other Examples of Social Influence on Price," *Journal of Political Economy*, 99, 1109-16.

Benabou, R. (1993): “Workings of a City: Location, Education and Production,” *Quarterly Journal of Economics*, 108, 619-52.

Beresteanu, A., I. Molchanov, and F. Molinari (2008): ”Sharp Identification Regions in Games,” cemmap working paper CWP15/08.

Bernheim, D. (1994): “A Theory of Conformity,” *Journal of Political Economy*, 102, 841-77.

Berry, S.T. and E. Tamer (2007): ”Identification in Models of Oligopoly Entry,” in Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, vol. II, Chap. 2, 46-85, Cambridge: Cambridge University Press.

Bifulco, R., J. Fletcher, and S. Ross (2009): ”The Effect of Classmate Characteristics on Individual Outcomes: Evidence from the Add Health,” University of Connecticut Working Paper 2008-21R

Bisin, Alberto, Giorgio Topa, and Thierry Verdier, “An Empirical Analysis of Religious Homogamy and Socialization in the U.S.,” mimeo, NYU, 2000.

Bjorn, P.A. and Q.H. Vuong (1985): ” Simultaneous Equation Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation,” Caltech, Working Paper 537.

Blume, L. and S. Durlauf, (1998): “Equilibrium Concepts for Social Interaction Models,” mimeo, Cornell University.

Blume, L. and S. Durlauf, (2001): “The Interactions-Based Approach to Socioeconomic Behavior,” in S. Durlauf and P. Young (eds.), *Social Dynamics*, Brookings

Institution Press and MIT Press, Cambridge, MA.

Breshnahan, T. F. and P. C. Reiss, “Empirical Models of Discrete Games”, *J. of Econometrics* **64,** 57-81, 1991

Brock, W.A. and S. N. Durlauf (2001a): “Interaction-Based Models,” in J. J. Heckman and E. E. Leamer, eds., *Handbook of Econometrics*, North-Holland, Chapter V.

Brock, W. and S. Durlauf (2001b): “Discrete Choice with Social Interactions”, *The Review of Economic Studies*, 68, 235-260.

Brock, W. and S. Durlauf (2007): “Identification of Binary Choice Models with Social Interactions,” *Journal of Econometrics*, 140, 52-75.

Bramoulle’, Y., H. Djebbari, and B. Fortin (2007): ”Identification of Peer Effects through Social Network,” IZA Discussion Paper 2652.

Calvo-Armengol, A., E. Patacchini, and Y. Zenou (2009): ”Peer Effects and Social Networks in Education,” *The Review of Economic Studie*s, forthcoming.

Case, A. and L. Katz (1991): “The Company You Keep: The Effect of Family and Neighborhood on Disadvantaged Families,” NBER Working Paper 3705.

Ciliberto, F. and E. Tamer (2004): ”Market Structure and Multiple Equilibria in Airline Markets,” mimeo.

Crane, J. (1991): “The Epidemic Theory of Ghettoes and Neighborhood Effects of Dropping Out and Teenage Childbearing,” *American Journal of Sociology*, 96, 1229-59.

Dagsvik, J. and B. Jovanovic, “Was the Great Depression a Low-Level Equilibrium?” *European Economic Review* 38 (December 1994), 1711–29.

Davezies, L., X. D’Haultfoeuille, and D. Fougere (2006): ”Identification of Peer Effects Using Group Size Variation,” CEPR Discussion Paper 5865.

Durlauf, S. (1993): “Nonergodic Economic Growth,” *Review of Economic Studies*, 60, 349-66.

Durlauf, S. (1996): “Neighborhood Feedbacks, Endogenous Stratification, and Income Inequality,” in Barnett, W., G. Gandolfo, and C. Hillinger (eds.), *Dynamic Disequi-*

librium Modelling: Proceedings of the Ninth International Symposium on Economic Theory and Econometrics, Cambridge: Cambridge University Press.

Echenique, F. and I. Komunjer (2005): "Testing Models with Multiple Equilibria by Quantile Methods," California Institute of Technology Social Science Working Paper 1244.

Ellison, G. and D. Fudenberg (1993): "Rules of Thumb for Social Learning," *Journal of Political Economy*, 101, 612-44.

Farmer, R. and J.-T. Guo (1995): "The Econometrics of Indeterminacy: An Applied Study," *Carnegie-Rochester Series in Public Policy*, 43, 225-272.

Galichon, A. and M. Henry (2008): "Inference in Models with Multiple Equilibria," Universite' de Montreal, mimeo.

Gaviria, A. and S. Raphael (2001): "School-Based Peer Effects and Juvenice Behavior," *Review of Economics and Statistics*, 83(2), 257-68.

Glaeser, Edward, and Jose' A. Scheinkman, "Measuring Social Interaction," in S. Durlauf and P. Young (eds.), *Social Dynamics*, Cambridge, MA: Brookings Institution Press and MIT Press. (2001)

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

Graham, B.S. and J. Hahn (2005): "Identification and Estimation of the Linear-in-Means Model of Social Interactions," *Economics Letters*, 88, 1-6.

Heckman, James, "Sample Selection Bias as a Specification Error with an Application to the Estimation of Labor Supply Functions," in James P. Smith (Ed.), *Female Labor Supply: Theory and Estimation*, Princeton University Press, 1980.

Imrohoroglu, S. (1993): "Testing for Sunspot Equilibria in the German Hyperinflation," *Journal of Economics Dynamics and Control*, 17(2), 289-317.

Ioannides, Y.M., (2003): “Topologies of Social Interactions,” mimeo, Tufts University.

Jones, A. (1994), “Health, Addiction, Social Interaction, and the Decision to Quit Smoking,” J*ournal of Health Economics*, 13, 93-110.

Jovanovic, B. (1989): “Observable Implications of Models with Multiple Equilibria”, *Econometrica*, 57 (1989), 1431-1437.

Kelly, M. (1997): “The Dynamics of Smithian Growth,” *Quarterly Journal of Economics*, 112, 939-64.

Krauth, B.V. (2006): ”Simulation-Based Estimation of Peer Effects,” *Journal of Econometrics*, 133, 243-71.

Krosnick, J. and C. Judd (1982): “Transitions in Social Influence in Adolescence: Who Induces Cigarette Smoking,” *Develomental Psycholog*y, 81, 359-68.

Lee, L.f (2007): ”Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects,” *Journal of Econometrics*, 140, 333-74.

Manski, C.F. and E. Tamer (2002): ”Inference on Regressions with Interval Data on a Regressor or Outcome,” *Econometrica*, 70, 519-546.

Moody, J. (2001): ”Race, School Integration, and Friendship Segregation in America,” *American Journal of Sociology*, 107(3), 679-716.

Moro, A., “The Effect of Statistical Discrimination on Black-White Wage Inequality: Estimating a Model with Multiple Equilibria,” mimeo, University of Minnesota, 2001.

Nakajima, R. (2007): ”Measuring Peer Effects on Youth Smoking Behavior,” *Review of Economic Studies*, 74(3), 897-935.

Norton, E.C., R.C. Lindrooth, and S.T. Ennett (1998): ”Controlling for the Endogeneity of Peer Substance Use on Adolescent Alcohol and Tobacco Use,” *Health Economics*, 7(5), 439-53.

Pakes, A., M. Ostrovsky, and S.T. Berry (2004): "Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples)," NBER Working Paper 10506.

Pakes, A., and D. Pollard, "Simulation and the Asymptotics of Optimization Estimators", *Econometrica*, 57 (1989), 1027-1057.

Pesendorfer, M. and Schmidt-Dengler (2004): "Least Squares Estimators of Dynamic Games," LSE, mimeo.

Powell, L.M., J.A. Tauras, and H. Ross (2003): "Peer Effects, Tobacco Control Policies, and Youth Smoking Behavior," University of Illinois at Chicago, mimeo.

Schelling, T. (1972): "A Process of Residential Segregation: Neighborhood Tipping," in A. Pascal (ed.), *Racial Discrimination and Economic Life*, Lexington, MA: Lexington Books.

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," *The Review of Economic Studies*, 70, 147-65.

Tyas, S.L. and L.L. Pederson (1998): "Psychosocial Factors Related to Adolescent Smoking: A Critical Review of the Literature," *Tobacco Control*, 7(4), 409-20.

Wang, M.Q., E.C. Fitzhugh, R.C. Westerfield, and J.M. Eddy (1995): "Family and Peer Influences on Smoking Behavior among American Adolescent: An Age Trend," *Journal of Adolescent Health*, 16(3), 200-03.

Wang, M.Q., J.M. Eddy, E.C. Fitzhugh, (2000): "Smoking Acquisition: Peer Influence and Self-selection," *Psychological Reports*, 86, 1241-46.