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

Alberto Bisin Andrea Moro Giorgio Topa†

October 8, 2010

Abstract

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 multiplicity.

Furthermore, we implement the proposed estimation procedure using Add Health data to understand how group interactions affect teenagers' smoking behavior. We find strong social interaction effects both at the level of the friendship network and at the level of the school. We also find that multiplicity of equilibria is pervasive at the estimated parameter values. When comparing the actual level of smoking in each school with the lowest equilibrium level of smoking, we find that equilibrium selection accounts for about 15 percent additional smoking. Finally, when comparing the actual smoking with the equilibrium smoking that would occur if there were no interactions among

*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. 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. Finally, we thank Chris Huckfeldt, Matt Denes, and Ging Cee Ng for exceptional research assistance.

†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.

students, we find that student interactions surprisingly *reduce* smoking by approximately 70 percent.

1 Introduction

In this paper we study the 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 the 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 as 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 computing all feasible equilibria, nor postulating an arbitrary equilibrium selection rule. We evaluate the efficiency and computational feasibility of the two alternative procedures approaches using Monte-Carlo simulations. We show, in the context of Brock and Durlauf (2001b)'s canonical model, that while less efficient, the two step procedure 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 procedure, which is agnostic about equilibrium selection. We conclude that the two-step estimation procedure is particularly appropriate when the investigator does not have information about the equilibrium selection.

Furthermore, in this paper, we implement the proposed two-step estimation 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 (various extensions of) Brock and Durlauf (2001b)'s model. In our estimates, 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 strength 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 at the level of the school and at the level of the individuals' friendship 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 friendship or school-wide social interactions (e.g., changes in the number of friends in personal networks, or in 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.

More specifically, we find a significant heterogeneity between schools in the magnitude of the interaction effects. Simulating all equilibria in each school we find that multiple equilibria are present in 40 out of the 41 schools used for this exercise. Simulating students' behavior if all schools selected the equilibrium with the lowest smoking level, we find that selection into a higher level of smoking equilibrium accounts for about 15 percent higher smoking, on average. Finally, to quantify the effect of social interactions, we simulate the level of smoking that would occur if there were no preferences for peers' behavior. Compared to the outcome without interactions, actual smoking is 70 percent lower, on average. Somewhat contrary to standard presemptions, this result suggests that social interactions may have an important role in *reducing, rather then increasing*, smoking in adolescents.

Before going further, we wish to emphasize an important limitation of this paper. In any study of social interactions, comovements between an individual's and her peers' outcomes may be due to peer influences but also to sorting of individuals into groups according to observed and unobserved attributes. More generally, one needs to distinguish social interaction effects from other correlated unobservable factors that may induce the observed comovements in outcomes. In this paper, we sidestep this issue and focus instead on how to explicitly treat the multiplicity of equilibria that typically arise in economies with social interactions.[1]

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.[2] 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

[1]A companion paper, Bisin-Moro-Topa (2010) (in progress), addresses the possibility of correlated unobservables by explicitly modeling the selection of individuals into networks.

[2]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.

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, 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).[3]

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 constructing 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

[3] In fact, this distinction is not clear in the literature, where it is often argued that an equilibrium selection rule must usually be imposed to achieve point identification of parameters.

from data the time series of the sunspot realizations under assumptions on the properties of the process; see also Aiyagari (1995). More 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.

A related literature studies the issue of identification in multiple equilibrium models of industrial organization. It concentrates on simultaneous-move finite games of complete information where the investigator only observes the actions played by the agents, whereas the parameters to be estimated also affect the payoffs.[4] 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*.[5] 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.[6] 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

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

[5]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.

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

different equilibria, 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 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.[7]

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.[8] Nakajima (2007) also exploits a selection mechanism, although the mechanism is implicitly determined by an adaptive learning mechanism.

2 A general society

Consider a society populated by a set of agents indexed by $i \in I$. Each agent $i \in I$ is characterized by a vector of exogenous characteristics X_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 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

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

[8]The structure of Nash equilibria of supermodular games can however be exploited to reduce the set of equilibria to be computed under strong assumptions on the support of the selection mechanism. E.g., this is the approach seemingly adopted by Krauth (2006).

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

$g(i) \subset I_n$.[10] Local interactions, for example peer effects, will manifest themselves in group $g(i)$. Let $\#g(i)$ be the dimensionality of $g(i)$.

Each agent $i \in I$ chooses an element y_i in a compact set Y (possibly a discrete set). Agents' choices are simultaneous. 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 choices of all agents. Finally, let ε_i denote a vector of idiosyncratic shocks hitting agent $i \in I$, let ε_n denote the vector of aggregate shocks hitting all agents $i \in I_n$, and let $\varepsilon = (\varepsilon_n)_{n \in N}$. Let $p_{n(i)}(\varepsilon_i \mid \varepsilon)$ denote the conditional density of the shocks ε_i, for $i \in I_n$, which are defined on a compact support. We allow the distribution p_n to depend on the choice vector $\mathbf{y}_n$. We assume ε_i and ε_j are conditionally independent across $i, j \in I_n$, for any $n \in N$. Let $\mathbf{p}(\varepsilon)$ denote the density of ε. Typically, these shocks are preference shocks, but they could also represent technology shocks.

Let then π_n denote an A-dimensional vector of equilibrium aggregates defined at the level of sub-population n. Typically, π_n contains an externality or a global social interaction effect. For instance, if Y is binary, π_n could contain the fraction of agents in sub-population n who choose a particular element of $y \in Y$. If the society has a competitive market component, then π_n would typically also contain the vector of competitive equilibrium prices.

We assume complete information for simplicity, that is, any agent $i \in I$ observes $(\varepsilon_i)_{i \in I}$, ε, 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 also observe some component of π_n, e.g., market prices. The set of first order conditions which determine the choice of an arbitrary agent $i \in I_n$ can then generally be written as[11]

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

if Y is a continuous set; and as

[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.

[11] The analysis is easily extended to the case of incomplete information, e.g., where agents' information is restricted to their neighbors or their sub-populations.

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

if instead Y is a discrete set.[12]

At equilibrium the first order conditions *1* (or 2 if Y is discrete) are satisfied jointly for any agent $i \in I$. This is equivalent to requiring that $\mathbf{y}$ satisfies a Nash equilibrium of the simultaneous move anonymous game. Furthermore, at equilibrium, for any sub-population n, π_n satisfies a set of conditions which can be represented as

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

for some vector valued smooth map A_n.

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* $\mathbf{y}_n$.

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

$$A_n(\mathbf{y}_n, \pi_n, Z_n, \varepsilon_n) = 0, \quad \text{for any } n. \qquad (3)$$

Note that in our general society sub-populations are (allowed to be) inter-related, in the sense that we did not impose any independence assumption on the shocks ε_n. On the other hand we do implicitly require that both externalities and markets do not extend across sub-population, as no $\pi_{n'}$, $n' \neq n$, enters in the equilibrium condition (3) for sub-population n. In fact, however, our analysis is unchanged if we allow for relations across

[12]Typically, the first order conditions will result from agent $i's$ choice of y_i to maximize preferences:

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

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. A technology constraint of the form

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

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

sub-populations, at the level of the general society, as follows. Let Π denote a vector of equilibrium aggregates defined at the level of the whole population. Typically, Π would contain global externalities and prices for markets extending to the whole population. Let $\pi = (\pi_n)_{n\in N}$. The equilibrium conditions, in this case, are then

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

for vector valued smooth maps A_n, A.

2.1 Special case: Brock and Durlauf's binary choice model

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:

$$\max_{y_i \in \{-1,1\}} V(y_i, X_i, Z_n, \pi_n, \varepsilon_i) = h_n(X_i, Z_n, \varepsilon_n) \cdot y_i + J_n y_i \pi_n + \varepsilon_i. \tag{4}$$

The aggregate π_n is defined to be the average choice in the sub-population n.[13] Consequently, the single equilibrium condition for the society, $A_n(\mathbf{y}_n, \pi_n, Z_n, \varepsilon_n) = 0$ in our general formulation, takes the simple form:

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

(where $I_{y_i=y}$ is the indicator function of $y_i = y$).

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 distribution of shocks ε_i is assumed to be independent of ε and identical across sub-populations n: $p_{n(i)}(\varepsilon_i \mid \varepsilon) = p(\varepsilon_i)$. Furthermore, $p(\varepsilon_i)$ depends on agent i's choice y_i and

[13] In empirical work it is often appropriate to consider local interactions, e.g., by adding dependence on $\pi_{g(i)} = \frac{1}{\# g(i)} \sum_{i \in g(i)} \left(I_{y_i=1} - I_{y_i=-1} \right)$ (see for instance our analysis of teenage smoking in Section ??).

it is extreme-valued, that is,[14]

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

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 :

$$\pi_n=\sum_{i\in I_n}\tanh\left(c_nX_i+\varepsilon_n+J_n\pi_n\right)\tag{5}$$

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 generically has either one or three solutions.

2.2 Special case: Glaeser and Scheinkman's continuous choice model

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

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

[14] More generally, for an extreme value distribution, $\Pr\left(\varepsilon_i\left(-1\right)-\varepsilon_i\left(1\right)\leq z\right)=1/(1+\exp\left(-\beta z\right))$ 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.

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,

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

3 Identification

In this section we study identification in the set-up of Section 2.[15] 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.[16] Let $\theta_n \in \Theta$ denote the vector of parameters to be estimated in sub-population n, with Θ compact. We derive conditions for $\theta = \{\theta_n\}_{n\in N}$ to be identified from the econometrician's observation of $(\pi, \mathbf{y}, \mathbf{X}, \mathbf{Z})$ as well as of the composition of each sub-population and of the whole neighborhood network, $(n(i), g(i))$ for any $i \in I$.[17] We shall also assume some parametric specification for $p_{n(i)}(\varepsilon_i \mid \varepsilon)$ but not necessarily for $\mathbf{p}(\varepsilon)$. In other words, the parameter vector $\theta_n \in \Theta$ contains the parameters which identify the specification of $p_n(\varepsilon_i \mid \varepsilon)$, for any $i \in I_n$.

Equilibrium at each n requires

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

[15] By identification we mean identification in population. Sometimes identification in population is called identifiability; see e.g., Chiappori and Ekeland (2009).

[16] 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). Also, it should be apparent that the analysis of identification in this section can be readily extended to societies with interactions both at the level of the sub-population n and at the level of the whole society.

[17] In empirical implementation it might be convenient to impose parameter constraints across sub-populations, e.g., that all parameters are identical, $\theta_n = \theta$, for any $n \in N$. The conditions for identification are of course weaker under these class of constraints.

where we explicitly include the dependence of equilibrium on $\theta_n \in \Theta$ for clarity in the notation. An equilibrium in sub-population n also includes the first order conditions:[18]

$$y_i = y_i\left(X_i, \mathbf{y}_{g(i)}, \pi_n, Z_n, \varepsilon_i, \varepsilon_n; \theta_n\right), \text{ for any } i \in I_n$$

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:

$$y_i = y_i\left(X_i, \mathbf{y}_{g(i)}, \pi_n, Z_n, \varepsilon_i, \varepsilon_n; \theta_n^{foc}\right) \tag{6}$$

Finally, let $\pi_n(\theta_n, \varepsilon_n)$ represent the equilibrium manifold which satisfies the equilibrium condition $A_n(\mathbf{y}_n, \pi_n, Z_n, \varepsilon_n; \theta_n) = 0$, given $(\mathbf{y}_n, \mathbf{X}_n, Z_n)$ (dropped from the notation for simplicity.

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

$$L(\mathbf{y}|\theta) = \Pi_{i\in I} \int_{(\varepsilon_i,\varepsilon):\left\{\begin{array}{c} y_i = y_i\left(X_i, \mathbf{y}_{g(i)}, \pi_n, Z_n, \varepsilon_i, \varepsilon_n; \theta_n^{foc}\right), \text{ for any } i \in I \\ \pi_n = \pi_n(\theta_n, \varepsilon_n), \text{ for any } n \in N \end{array}\right\}} p_{n(i)}(\varepsilon_i \mid \varepsilon)\mathbf{p}(\varepsilon)d\varepsilon_i d\varepsilon \tag{7}$$

where $\int$ denotes the Aumann integral.[19] Let $\mathbf{L}(\theta)$ be the set of measurable likelihood functions induced by (7), so that any $l(\mathbf{y}|\theta) \in \mathbf{L}(\theta)$ is a measurable selection of the correspondence $L(\mathbf{y}|\theta)$.The standard sufficient condition for identification of the parameter vector θ at θ^0 can then be generalized as follows.

Identification. *For all* $\theta \in \Theta^N, \theta \neq \theta^0$,

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

[18]By construction, the dimensionality of the system of first order conditions is equal to the dimension of the vector of choice variables y_i. Also, the dimensionality of the equilibrium conditions is equal to A, the dimension of the vector of equilibrium variables π_n.

[19]See Aliprantis for the formal definition and a discussion of the properties of such integral. Loosely speaking, the Riemann integral is not defined since $\pi_n(\theta_n, \varepsilon_n)$ 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.

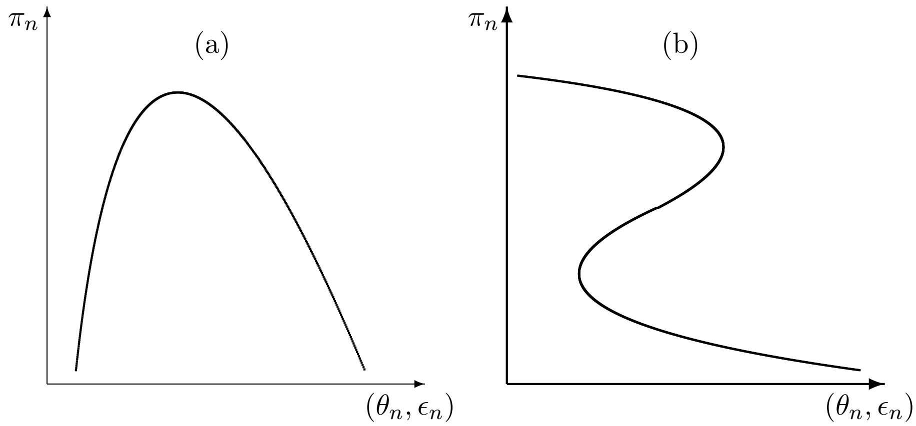

Figure 1: (a) Unique equilibrium, no identification; (b) multiple equilibria with identification

For any subpopulation n, the conditions for the identification of θ_n^{foc} off of the first order conditions, given $(\pi_n, \varepsilon_n, \mathbf{y}_n, \mathbf{X}_n, Z_n)$, are standard (and involve no issue of multiplicity of equilibria).[20] We assume these conditions are satisfied and hence θ_n^{foc} is identified, given $(\pi_n, \varepsilon_n, \mathbf{y}_n, \mathbf{X}_n, Z_n)$. The equilibrium conditions can therefore be represented, for any $n \in N$, as a map (a smooth manifold, in fact) from $(\theta_n^{eq}, \varepsilon_n)$ into π_n, for given observables $(\mathbf{y}_n, \mathbf{X}_n, Z_n)$ and for given θ_n^{foc}. Let $\pi_n(\theta_n^{eq}, \varepsilon_n)$ be such a map. Equilibrium is unique, in sub-population n, if $\pi_n(\theta_n^{eq}, \varepsilon_n)$ is one-to-one. This is however not necessary nor sufficient for identification.

Strong condition for identification of θ_n. *Identification of θ_n 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 un-related to the existence of multiple equilibria.[21] Panel (a) in Figure 1 displays an equilibrium manifold which does not satisfy the *condition for identification* that $\binom{\theta_n^{eq}}{\varepsilon_n}(\pi_n)$ be one-to-one, though equilibrium is unique. Panel (b) displays 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 it is such that the inverse equilibrium manifold $\binom{\theta_n^{eq}}{\varepsilon_n}(\pi_n)$ is indeed one-to-one, and hence it satisfies the *strong condition for identification.*

Note also that the strong condition for identification guarantees non-parametric identification with respect to the distribution of shocks ε, $\mathbf{p}(\varepsilon)$. But suppose that the econo-

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

[21] A related result appears in Jovanovic (1989), who shows that a unique reduced form is neither necessary nor sufficient condition for identification.

metrician has prior information (or is willing to make assumptions) e.g., on the functional form or on other properties of $\mathbf{p}(\varepsilon)$. In this case, weaker conditions for identification might be derived.

Weak condition for identification of θ_n. *Identification of θ_n obtains if i) $\theta_n^{eq}(\pi_n, \varepsilon_n)$ is one-to-one and if ii) $\mathbf{p}(\varepsilon)$ is strictly monotonic in its arguments.*

Some regularity condition on $\mathbf{p}(\varepsilon)$ is 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 shocks ε are independently and identically distributed across n, $\mathbf{p}(\varepsilon) = \Pi_{n\in N}\ p(\varepsilon_n)$, and $p(\varepsilon_n)$ is assumed uniform, identification might not be guaranteed even if $\theta_n^{eq}(\pi_n, \varepsilon_n)$ is one-to-one. However, it is not necessarily required that $p(\varepsilon_n)$ be strictly monotonic. If, for instance $p(\varepsilon_n) \sim N(0,1)$, identification obtains provided $\theta_n^{eq}(\pi_n, \varepsilon_n) \neq \theta_n^{eq}(\pi_n, -\varepsilon_n)$, a generic condition.

Relatedly, suppose the econometrician has information (is willing to assume) that shocks ε are independently and identically distributed.

Condition for identification of $\mathbf{p}(\varepsilon)$. *Suppose shocks ε are independently and identically distributed across n, that is, $\mathbf{p}(\varepsilon) = \Pi_{n\in N} p(\varepsilon_n)$. Identification of all moments up the N-th of $p(\varepsilon_n)$ obtains if $\binom{\theta_n^{eq}}{\varepsilon_n}(\pi_n)$ is one-to-one (the strong condition for identification of θ_n is satisfied) for all $n \in N$.*

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 requires that the econometrician be able compute all the feasible equilibria for every set of parameters, which is ofter a daunting computational task. In this section, we introduce two alternative estimation methods for societies with potentially multiple equilibria. The *direct estimation method* has all the large and small sample properties of a standard maximum likelihood method, though it requires the computation of all the feasible equilibria for every set of parameters. The alternative estimator we propose, a *two-step estimation method,* preserve the large sample properties of the direct method, while being computationally straightforward.

Suppose that, for any sub-population $n \in N$, the econometrician observes (y_i, X_i) for a random sample of the individuals $i \in I_n$, Z_n, and also (the econometrician can compute) $\widehat{\pi}_n$, a point estimate of π_n.[22] We assume that, for any $n \in N$, the weak condition for identification of θ_n is satisfied, but of course the analysis carries true if the strong condition is satisfied.

4.1 The direct estimation method

We define a direct (maximum likelihood) estimator of θ as follows:

$$\widehat{\theta} = \arg\max_{\theta} \max_{l(\mathbf{y}|\theta) \in \mathbf{L}(\theta)} l(\mathbf{y}|\theta) \tag{8}$$

Recall that $\mathbf{L}(\theta)$ is the set of all measurable selections $l(\mathbf{y}|\theta)$ induced by the correspondence $\mathbf{L}(\mathbf{y}|\theta)$ defined by (7). Because of the possible multiplicity of equilibria, $\mathbf{L}(\theta)$ is very difficult to characterize. However, under standard regularity conditions,

the direct Maximum likelihood estimation estimator $\widehat{\theta}$ is consistent and efficient.

The estimator $\widehat{\theta}$ can be computed by using the following algorithm: *1.* Pick a sub-population $n \in N$; *2.* For each $(\theta_n, \varepsilon_n)$, compute all the equilibria of the model; *3.* Compute the likelihood of each equilibrium, choose the maximum among them; *4.* Repeat for all $n \in N$; *5.* Integrate over ε and maximize over θ. Such procedure is computationally difficult especially when the society is sufficiently complex that a closed form characterization of equilibrium is impossible.

4.2 The two-step estimation method

We now introduce a simpler two-step estimation procedure. The first step consists in computing the point estimate $\widehat{\pi}_n$ for π_n.[23] For any $i \in I_n$, let

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

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

[22] We also assume that the econometrician observes, for any individual i in the sample, which sub-population $n(i)$ he is assigned to, his network $g(i)$, and the vector $\mathbf{y}_{g(i)}$.

[23] In Glaeser and Scheinkman's society, e.g., $\widehat{\pi}_n$ could be the sample mean y_i.

$$\widehat{\hat{\theta}}_n^{foc} = \arg\max_{\theta_n^{foc}} l(\mathbf{y}_n, \widehat{\pi}_n | \theta_n^{foc}) \tag{9}$$

Furthermore, in the second step, we estimate θ_n^{eq} as

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

Note that the estimation of $\widehat{\hat{\theta}}_n^{eq}$ does not require the computation of all the equilibria as a function of θ_n^{eq}, as $\theta_n^{eq}(\hat{\pi}_n, \varepsilon_n)$ is one-to-one by the weak condition for identification.[24] It is straightforward to conclude the following.

Consistency of the *two-step estimation method*. *As long as $\widehat{\pi}_n$ is a consistent estimator of π_n, the two-step estimator $\widehat{\hat{\theta}}_n$ is in turn a consistent estimator of θ_n.*

Furthermore, in specific empirical implementations, the small sample properties for the *two-step estimator* can be improved by means of several expedients. First of all, the two-step method can be iterated as e.g., in Aguirregabiria and Mira (2007). Second, in constructing the likelihood $l(\mathbf{y}_n, \widehat{\pi}_n | \theta_n)$ the econometrician could account for the distribution of the estimator $\widehat{\pi}_n$, due e.g, to sampling error. Finally, the whole vector $\widehat{\theta}$ can be jointly estimated if shocks ε_n are known (or assumed) to be correlated over $n \in N$; or if parameter constraints across the sub-populations are imposed, as in Section??

Moro (2002) has been the first to employ the two-step method to estimate a model with multiple equilibria. In his model identifiying restrictions require that shocks are independently and identically distributed across sub-populations $n \in N$, and that $\theta_n(\pi_n, \varepsilon_n)$ is independent of ε_n and takes a unique value for any π_n. Under these conditions it is the case that the direct and the two-step estimation methods formally coincide.

4.3 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

[24] If $\mathbf{p}(\varepsilon)$ is assumed decreasing, then $\widehat{\hat{\theta}}_n^{eq}$ corresponds to $\theta_n^{eq}(\pi_n, \varepsilon_n)$ evaluated at $\pi_n = \hat{\pi}_n$ and at the smallest value for ε_n.

choices y_n:

$$\begin{aligned} \Pr\left(\mathbf{y}_n|\mathbf{X}_n,\pi_n,\varepsilon_n\right) &= \prod_i \Pr\left(y_i|X_i,\varepsilon_n,\pi_n\right) \sim \\ &\prod_i \exp\left(\left(c_n X_i+\varepsilon_n\right)\cdot y_i + J_n y_i \pi_n\right). \end{aligned} \tag{10}$$

Equation (10) suggests the following formulation of the likelihood function as a function of the parameter vector $\theta_n = \{c_n, J_n\}$:

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

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.

4.3.1 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,

Evaluation Criterion	Direct	Two step	Direct, 2step initial est
RMSE, parameter C	0.05056	0.05043	0.05043
Bias, parameter C	-0.00315	-0.00251	-0.00251
RMSE, parameter J	0.10768	0.10708	0.10706
Bias,parameter J	-0.00184	-0.00542	-0.00390
Min time	136.57404	0.23075	132.76532
Max time	183.95280	0.36781	181.82902
Mean time	157.22378	0.30190	158.34783
Median time	156.88555	0.29983	158.34823

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

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 $\widetilde{y}\left(\theta_j^{true}\right)$ 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 available from the authors upon request. 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 (the difference is slight in the case in which the first equilibrium is picked, and larger in the other two cases, especially when the intermediate equilibrium is selected). 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 method, however, is in computational speed. Even with this very stripped down model an estimation run with the direct method took a median time between 156 and 159 minutes (depending on the choice of equilibrium); instead, the 2-step method took roughly between 27 and 30 *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 only.

4.3.2 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). In the direct vs. two-step method experiment, we use $n = 20$, with 5,000 agents in each subpopulation; in the D-J vs. two-step method comparison, we use $n = 300$, with 200 agents in each subpopulation.[25] 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 sub-population experiments. 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.[26] 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.

Table 2 collects results for the first set of experiments, comparing 2-step and direct methods, where the evaluations are based on the results obtained from 160 runs where the “true” parameters are randomly drawn using the same 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,

[25] We have a larger number of cities in the D-J experiment because we wanted to also study the method’s ability to estimate the equilibrium selection mechanism; consequently, we were limited to having only 200 artificial students in each sub-population n for computational reasons.

[26] 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	α_1=α_2=1/3		α_1=0.1,α_2=0.8	
Evaluation Criterion	Direct	Two step	Direct	Two step
(1)	(2)	(4)	(5)	(6)
RMSE, C	0.04849	0.00343	0.02468	0.02330
Bias, C	0.00530	-0.00006	-0.00239	-0.00093
RMSE, J	0.18000	0.02769	0.04941	0.04216
Bias, J	0.03824	0.00705	0.00584	-0.00023
Min time	133.07780	1.23348	130.13754	1.30992
Max time	234.96883	6.49528	224.77100	7.60333
Mean time	159.35291	1.57270	154.92274	1.65008
Median time	157.15724	1.53797	152.82602	1.59336

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

Econometrician Conjectures	Correct: $\alpha_1, \alpha_2 > 0$		Misspecified: $\alpha_2 = 0$	
Evaluation Criterion	D-J	Two step	D-J	Two step
(1)	(2)	(3)	(4)	(5)
RMSE, C	0.00580	0.00736	0.00567	0.00721
Bias, C	-0.00015	-0.00027	-0.00032	-0.00036
RMSE, J	0.02455	0.04497	0.02390	0.04466
Bias, J	0.00235	0.02343	0.00253	0.02302
Min time	330.61444	0.60659	277.07853	0.62736
Median time	352.19753	0.74459	297.17443	0.73311
Max time	417.93318	2.32693	382.53198	5.82907

Table 3: Monte carlo experiments: misspecification of the equilibrium transition probabilities

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

Finally, we turn to two experiments that perform a comparison between 2-step and D-J methods. In each experiment, we wish to evaluate the performance of the D-J method in the case in which the econometrician correctly specifies the equilibrium selection process, as opposed to a situation in which the equilibrium selection is misspecified relative to the truth. In the first experiment, the true selection mechanism is a SAR(2) process, with $a_1 = 0$, $a_2 = 0.95$, whereas the econometrician assumes a SAR(1) specification in the misspecified case. In the second experiment, the true equilibrium selection mechanism imposes $p_1 = p_2 = 1/3$ and $a_1 = a_2 = 1/3$, but the econometrician assumes $p_2 = 0$ (i.e., the intermediate equilibrium is never chosen) in the misspecified case.[27]

The results of these two experiments are reported in Tables 3 and 4, respectively. In each Table, columns 2 and 3 report results for a situation in which the econometrician adopting the D-J method uses the correct specification, whereas columns 4 and 5 focus on the case in which the "D-J econometrician" chooses a misspecified model. Notice that when using the 2-step method one does not need to take a stand on the selection mechanism.

The comparison is interesting: in the first experiment, the D-J method performs somewhat better than the 2-step in terms of RMSE, especially with regard to the social interactions parameter J, in both the correct and the misspecified case. In the second experiment, however, the 2-step method performs dramatically better than the D-J one, with RMSE measures roughly one order of magnitude smaller for the former. In both cases, computational speed remains much higher for the 2-step method, even more so than in the comparison with the direct method. These Monte Carlo experiments therefore highlight another advantage of the 2-step method (in addition to speed): since it does not involve explicitly specifying an equilibrium selection, it is more robust to potential misspecification of this sort than methods that rely on taking a stand on the selection mechanism.

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

[27] The second experiment is inspired by models with myopic dynamics in which the intermediate equilibrium is typically unstable.

Econometrician Conjectures	Correct: $P_1, P_2 > 0$		Misspecified: $P_2 = 0$	
Evaluation Criterion	Jovanovic	Two step	Jovanovic	Two step
(1)	(2)	(3)	(4)	(5)
RMSE, C	0.00517	0.00912	0.07751	0.00911
Bias, C	0.00001	-0.00042	-0.01124	-0.00043
RMSE, J	0.02523	0.08002	0.74713	0.07953
Bias, J	0.00437	0.05484	0.65119	0.05394
Min time	333.29544	0.59714	271.94123	0.62225
Median time	356.65684	0.72343	287.05614	0.73007
Max time	453.14298	3.49153	398.08752	0.84737

Table 4: Monte Carlo experiments: misspecification of the equilibrium selection probabilities

(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, urban or rural setting, school size, school type, and ethnic composition.

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 (13)-(14) or in (15)-(16). In all cases, we allow for an intercept term in c_n.

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:

$$c_n = c(Z_n) = \alpha_0 + \sum_{k=1}^{K} \alpha_k Z_n^k; \tag{11}$$

$$J_n = J(Z_n) = \gamma_0 + \sum_{k=1}^{K} \gamma_k Z_n^k. \tag{12}$$

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

$$c_n = \alpha_0 + \alpha_n, \text{ with } \alpha_n \sim N(0, \sigma_\alpha) \tag{13}$$

$$J_n = \gamma_0 + \gamma_n, \text{ with } \gamma_n \sim N(0, \sigma_\gamma) \tag{14}$$

More generally we can also include school attributes Z_n:

$$c_n = \alpha_0 + \sum_{k=1}^{K} \alpha_k Z_n^k + \alpha_n, \text{ with } \alpha_n \sim N(0, \sigma_\alpha) \tag{15}$$

$$J_n = \gamma_0 + \sum_{k=1}^{K} \gamma_k Z_n^k + \gamma_n, \text{ with } \gamma_n \sim N(0, \sigma_\gamma) \tag{16}$$

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. In most cases, we consider the general case in which interactions have both a school-wide and a local component, that is, at the level of each agent's circle of friends, which are reported in the Add Health individual friendship network data. However, when we estimate the model parameters separately for each school, the global interaction parameter is not separately identified from the intercept term in agents' random utility (see Section 5.3). Therefore, in this case we estimate the case with local interactions only. Finally, as we discuss later, in order to study the effects of changes in the parameters on the equilibrium mapping we also need to estimate a specification with school-wide interactions only, so that we have a closed form solution for the equilibrium mapping.

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 Identification of parameters

In the general case presented above, where the estimation is run on data from all schools jointly, the parameters $\left(c_n, J_n^L, J_n^G\right)$ are separately identified by the variation in demographic attributes within schools, in smoking prevalence across personal networks, and in school-wide smoking prevalence across schools. The parameter ruling school-wide interactions, J_n^G, is identified separately from the intercept in c_n – which is assumed constant across schools – by the variation in the fraction of smokers across schools. The parameter for local interactions, J_n^L, is identified by the variation in smoking behavior across personal friendship networks both within and across schools.

In a single school setting, the school-wide interaction parameter is identified separately from the slope coefficients in c_n as long as there is some variation in demographic attributes within the school. Similarly, the local interaction parameter is identified as long as there

Variable	Local interactions Mean	Median
Black	-0.7358	-0.6737
Asian	-0.0576	-0.2356
Hispanic	-0.0206	-0.1690
Female	0.1248	0.1844
Age	0.1084	0.0795
Does not belong to a club	0.3394	0.2917
GPA	-0.3281	-0.3231
Mom college	-0.0237	0.1585
Dad at home	-0.2056	-0.2174
Local Interactions	0.7964	0.8241
Constant	-0.7270	-0.8506

Table 5: Mean and median parameter estimates, all schools estimated separately

is some variation in smoking prevalence across individual networks. However, J_n^G is not separately identified from the intercept term in c_n. Therefore, in the school-by-school estimation that we present below, we focus on the specification with local interactions only, arbitrarily setting $J_n^G = 0$.

5.4 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 changes 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 π_n, $\left[\pi_{g(i)}\right]_{i \in I_n}, \theta_n$ in school n is:

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

5.4.1 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 4.3 for a brief intuition about the identification of this model. Table 5 reports the means and medians of the parameter estimates across schools for the specification with local interactions only, since as discussed earlier the global interaction parameter is not separately identified from the constant in this case[28].

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:[29]ee for instance Gruber and Zinman (2000) or Gilleskie and Strumpf (2005). 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. Finally, the local interaction parameter estimate is positive and statistically significant in all schools. Its median size is roughly two and a half times as large as that of a student's GPA (in absolute value).

Figure 2 reports the distribution of parameter estimates for this specification. Only parameter estimates with t test statistics greater than unity are plotted. There is considerable dispersion around the medians across schools, especially with regard to the Hispanic category and the education level of the mother. 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. As noted, the local interaction parameter estimates are positive and statistically significant at the 5% level in all schools. 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.

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

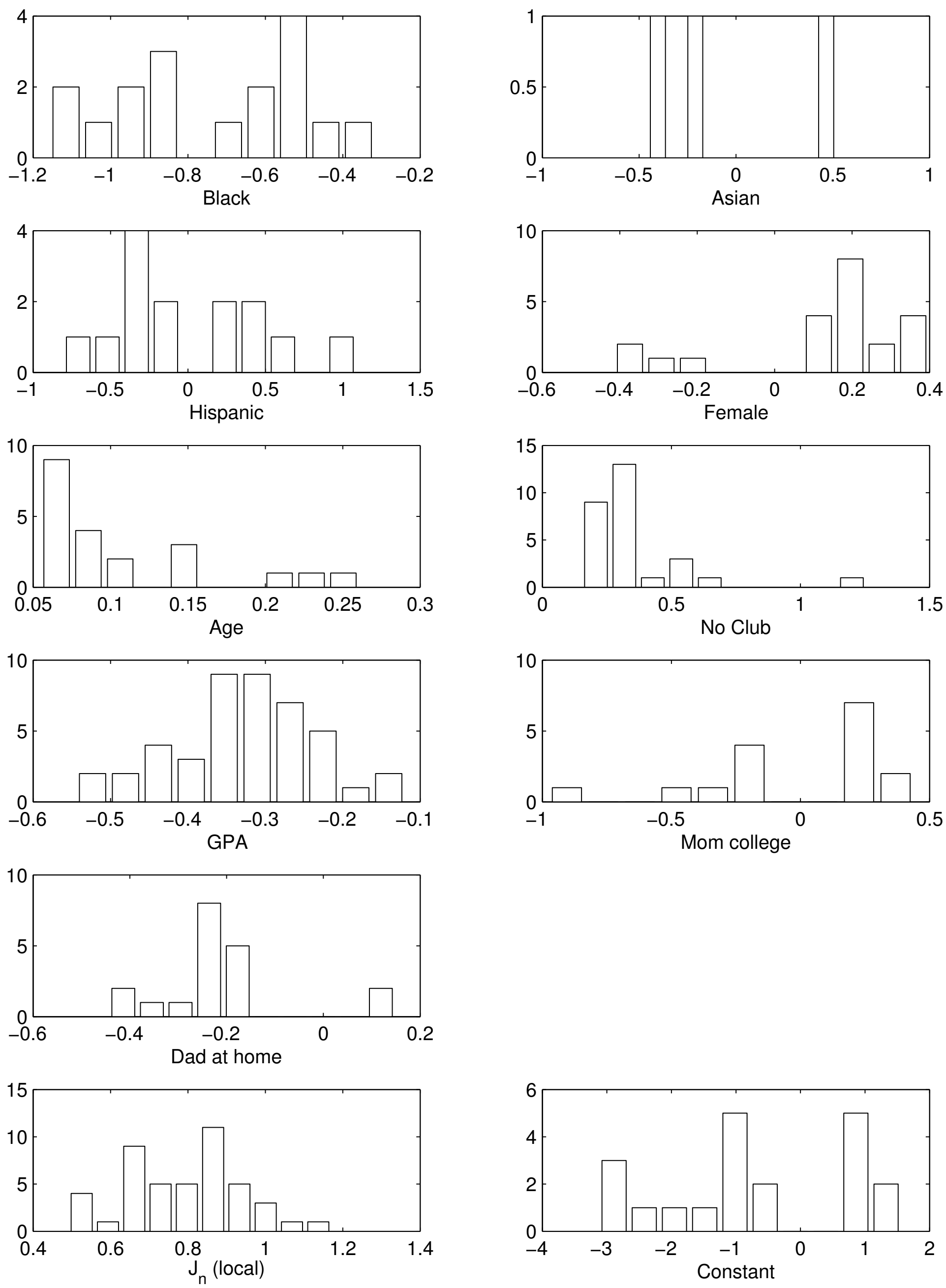

Figure 2: Distributions of parameter values, all schools separately, local interactions only

Variable	Coefficient	Std. err.
Black	-0.4415	0.0355
Asian	-0.0664	0.0392
Hispanic	-0.1622	0.0288
Female	0.0693	0.0174
Age	0.0441	0.0060
Does not belong to a club	0.2022	0.0208
GPA	-0.2695	0.0121
Mom college	0.0166	0.0199
Dad at home	-0.0877	0.0230
Global interactions	0.7191	0.0717
Local interactions	0.8252	0.0152
Constant	0.1889	0.1062
Log Likelihood	-11252.50	

Table 6: All schools, constant parameters

5.4.2 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 6 collect results for the general case with both school-wide and local interactions. The estimates are qualitatively similar to the distribution medians reported for the unconstrained case in the previous Section 5.4.1. Again minorities tend to 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. The parameter estimate for local interactions is very similar in magnitude to the median estimate for the case in which each school is treated separately. The school-wide interaction parameter, which is identified in this specification because of the variation in smoking prevalence across schools, is slightly smaller than that for local interactions, but still positive and statistically significant at the 1% level.

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 (11) and (12). 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 poverty line, the fraction of college educated in the population 25 years and older, the fraction of female (male) adults in the labor force.

The estimation results are reported in Figure 3 and Table 7. The first observation is that the distributions of estimated coefficients across schools reported in Figure 3 are qualitatively similar to those in the unrestricted specifications in Figure 2. 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 3 ranges from about 0.01 to about 0.08, whereas in the unrestricted case it ranges from about 0.05 to 0.25. The same is true for most of the other attributes as well as for the local interaction parameters: the latter ranges between about 0.7 and 0.95, whereas in the school by school estimation it ranges between 0.5 and 1.2 approximately. 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. Finally, the distribution of school-wide interaction estimates exhibits a much wider dispersion than that of local interactions, ranging between 0 and 2.

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 higher poverty levels, lower fraction of college graduates, and higher

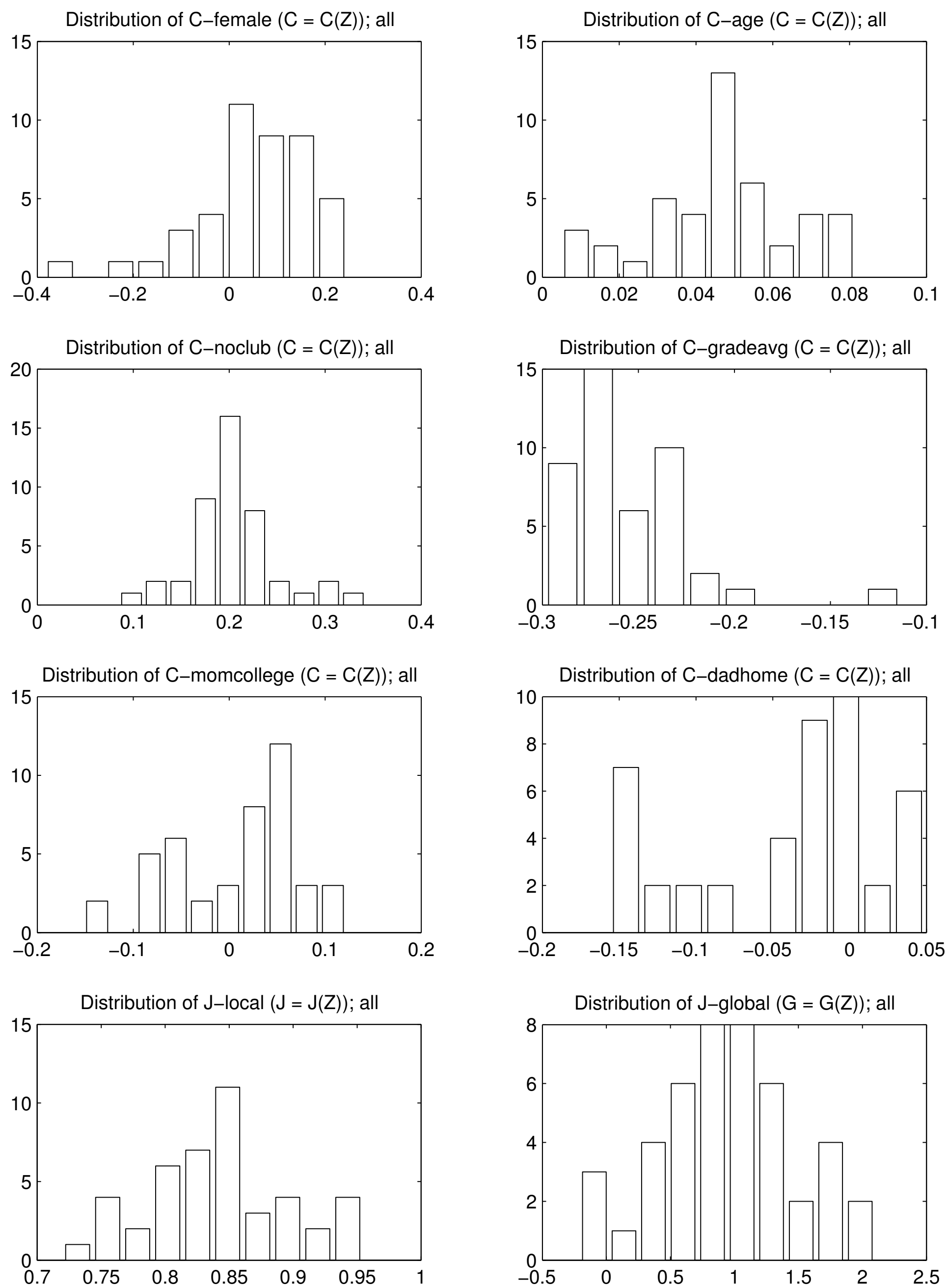

Figure 3: Distributions of parameter values, all schools, deterministic coefficients function of school characteristics

	Female	Age	No club
Intercept	1.4904 (0.1300)	-0.0037 (0.0067)	-0.5867 (0.2113)
% Urban	0.0034 (0.0601)	0.0333 (0.0080)	-0.0122 (0.0726)
% Poverty	-1.6764 (0.3644)	0.1857 (0.0400)	0.4575 (0.4327)
% College over 25	0.6127 (0.3485)	0.0370 (0.0238)	0.2228 (0.4275)
Female Labor Force P.R.	-1.4867 (0.2978)	-0.0900 (0.0133)	0.9968 (0.5351)
Male Labor Force P.R.	-0.8336 (0.1841)	0.0106 (0.0076)	0.2586 (0.3100)
Tobacco training	-0.0888 (0.0379)	0.0223 (0.0058)	0.0122 (0.0482)
Tobacco student policy	0.0922 (0.0880)	-0.0179 (0.0068)	0.2056 (0.1004)
Tobacco staff policy	-0.1237 (0.0882)	0.0388 (0.0061)	-0.1809 (0.0956)
	GPA	**Mom college**	**Dad Home**
Intercept	0.1427 (0.0325)	-0.3879 (0.1720)	-0.2364 (0.0966)
% Urban	0.0001 (0.0370)	0.1338 (0.0730)	-0.0185 (0.0707)
% Poverty	0.0711 (0.1740)	-0.0321 (0.3909)	0.3613 (0.3915)
% College over 25	-0.0967 (0.1322)	-0.4370 (0.3739)	0.2692 (0.3856)
Female Labor Force P.R.	-0.4700 (0.0734)	-0.4900 (0.4452)	-0.6471 (0.2632)
Male Labor Force P.R.	-0.1830 (0.0449)	0.9316 (0.2464)	0.4012 (0.1495)
Tobacco training	-0.0128 (0.0245)	-0.0264 (0.0457)	0.1201 (0.0485)
Tobacco student policy	-0.0375 (0.0386)	-0.0075 (0.1104)	-0.1367 (0.0996)
Tobacco staff policy	0.0106 (0.0402)	-0.0205 (0.1065)	0.1388 (0.0923)
	Local Inter.	**Global Inter.**	**Constant**
Intercept	0.3394 (0.1373)	2.5743 (0.1562)	0.1882 (0.1099)
% Urban	-0.1048 (0.0517)	0.9509 (0.1878)	
% Poverty	-0.2967 (0.3193)	4.4323 (0.7985)	
% College over 25	-0.0294 (0.2953)	1.2494 (0.6009)	
Female Labor Force P.R.	0.5911 (0.3383)	-6.8403 (0.3243)	
Male Labor Force P.R.	0.5800 (0.2016)	-0.8632 (0.1961)	
Tobacco training	-0.0403 (0.0397)	0.6166 (0.1487)	
Tobacco student policy	0.0038 (0.0821)	-0.6505 (0.1554)	
Tobacco staff policy	-0.0951 (0.0816)	1.0344 (0.1506)	

Table 7: All schools, deterministic coefficients function of school characteristics (standard errors in parenthesis, log likelihood -11130.5654)

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 higher poverty areas. The finding that Dad's presence at home is associated with less smoking is reinforced in neighborhoods with higher 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 training programs and tobacco rules for the staff seem to increase the strength of school-wide interactions but slightly reduce the strength of local interactions (however the latter estimates are not statistically significant). Of course these school policy variables are largely endogenous, but the finding that tobacco policies are related to stronger or weaker social interaction terms is important and we will return to it in our discussion of counter-factual experiments.

Finally, it is worth noting that neighborhood poverty levels and female labor force participation have a very large impact on 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 and other attributes. This possibility stresses the value of having individual friendship network data to estimate local social interaction effects.

5.4.3 Multiple schools, random coefficients

As mentioned in the previous section, letting (c_n, J_n) depend on observed school characteristics in a deterministic fashion may not be sufficient to capture the wide variation of parameter estimates across schools found in the unrestricted case. 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 (13)-(14), and then augment it with school level characteristics Z_n as in (15)-(16). 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 4 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 estimated so far. As expected, the range of values for the age, GPA and local interaction coefficients is wider than in the deterministic case (Figure 3) and very similar to that of the unrestricted school by school specification (Figure 2). Introducing local interactions effects generally lowers the school-wide interactions estimates relative to the case with school-wide interactions only.

Interestingly, we find a very large dispersion in the strength of school-wide social interaction effects, ranging roughly between zero and eight. We conjecture that this dispersion may reflect differences across schools in terms of neighborhood attributes, or in the extent of

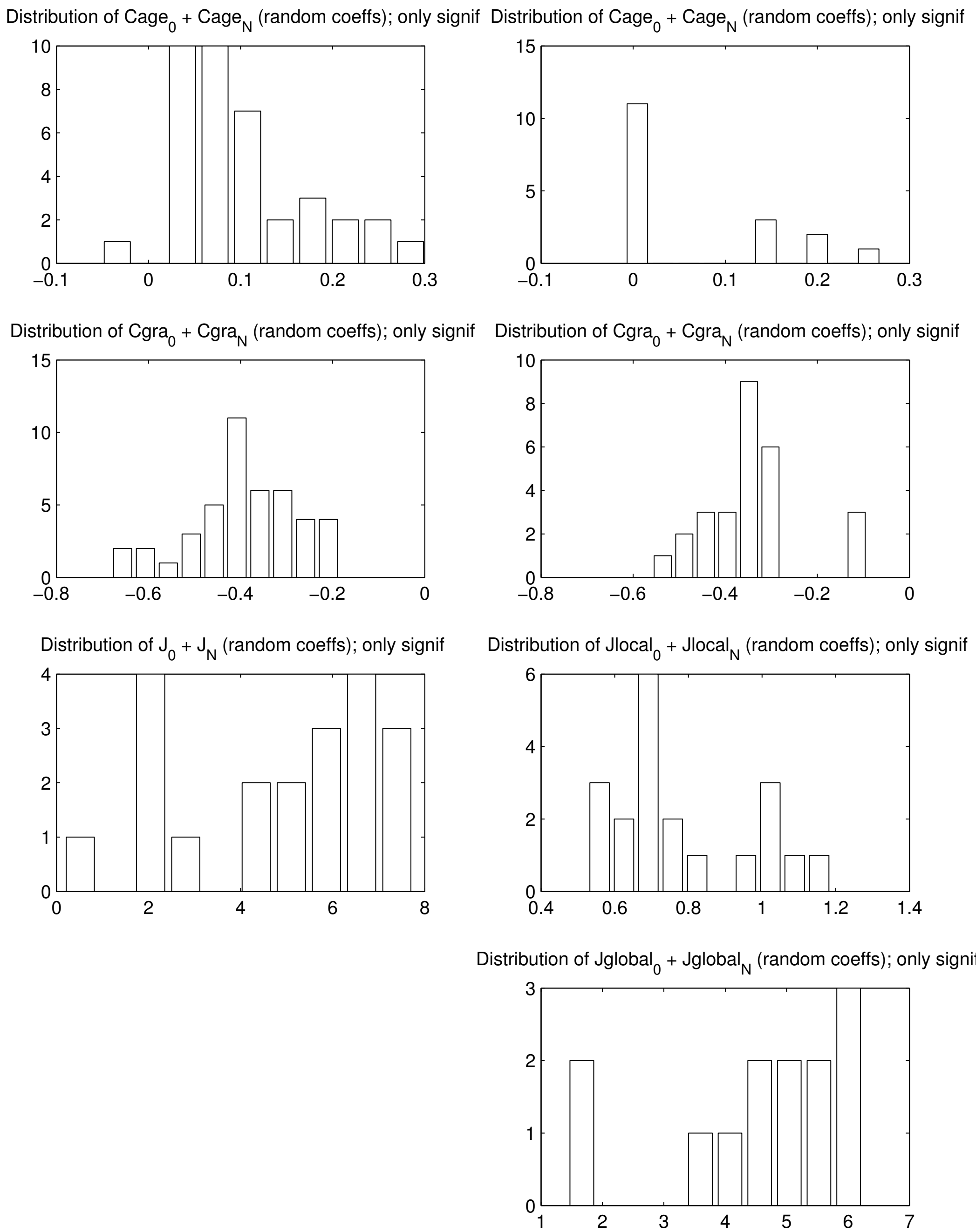

Figure 4: Distributions of parameter values, random coefficients. Left three panes: global interactions only; right four panes: both local and global interactions

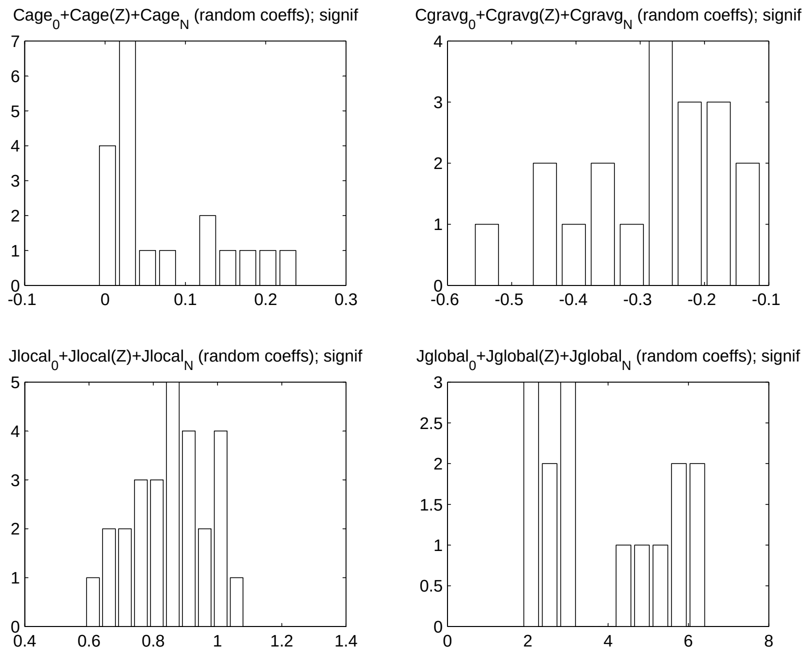

Figure 5: Distributions of parameter values, random coefficients and school characteristics

stratification and segregation along racial, ethnic, or socio-economic status lines within each school. 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[30].

Figure 5 and Table 8 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 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 5 are remarkably similar to those estimated in the unrestricted case, with smoking being positively associated with age and negatively associated with academic achievement. We also find significantly positive local interaction effects and a bimodal distribution for school-wide effects, again with a large dispersion in the magnitude of the

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

	Age	GPA	Local	Global
Intercept	0.0134 (0.0058)	-0.4033 (0.0284)	1.2017 (0.1173)	2.8264 (0.1628)
% Urban	0.0357 (0.0080)	-0.1293 (0.0417)	-0.1000 (0.1252)	0.4317 (0.1884)
% Poverty	-0.0415 (0.0405)	0.6534 (0.1972)	1.1833 (0.6447)	-1.2664 (1.0062)
Female LFPR	0.0068 (0.0126)	-0.0341 (0.0634)	1.0298 (0.2635)	1.6282 (0.3240)
Tbc. training	0.0208 (0.0065)	-0.0153 (0.0330)	-0.7897 (0.1197)	0.4142 (0.1641)
Tbc. stud. pol.	0.0147 (0.0067)	0.1088 (0.0359)	-0.0039 (0.1233)	0.1736 (0.1720)
Tbc. staff pol.	-0.0089 (0.0069)	0.0252 (0.0346)	-0.4069 (0.1366)	-1.0318 (0.1702)

	Constant
Intercept	1.8125 (0.0846)

Table 8: All schools, random coefficients (standard errors in parenthesis, log likelihood -11033.074)

latter.

Table 8 shows somewhat similar patterns to the deterministic specification in Table 7. Again, the presence of tobacco training programs is associated with stronger school-wide but weaker local interactions. Tobacco policies for the staff are associated with weaker social interactions both within personal friendship networks and school-wide. As we discuss in the next Section however, such policy effects do not necessarily imply less smoking in a given school.

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 that give rise to multiplicity of smoking equilibria. As a consequence of this analysis 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 values 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 into the possible effects of such policies.

School ID	N	Global int. (J)	Percentage of smokers			
			Data	Equil. 1	Equil. 2	Equil. 3
41	1056	2.67	5.9	5.9	17.2	99.8
56	1093	6.50	18.7	0.2	18.4	100.0
72	899	0.14	24.4	24.8	–	–

Table 9: Equilibrium smoking levels in a sample of schools

We carry out this analysis both in the case of school-wide interactions only, and in the case of local as well as school-wide interactions. In the former case, we can compute the set of equilibria directly from the equilibrium mapping, since it has a closed form solution. We can then see how the mapping moves as we vary the strength of interactions. In the case with both local and school-wide interactions, on the other hand, we rely on simulations of a Markov process that captures the Brock-Durlauf model, and characterize how its stationary distribution changes as we change its parameters.

6.1 The prevalence of multiplicity

To perform our first set of simulations we use the model specification with random effects (no school covariates Z_n) and school-wide interactions only whose results are reported in the left three panes of Figure 4. 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 (5)). At the same time, we have seen that this specification seems to capture well the variability of parameter estimates found in the unrestricted case. 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 9 reports the equilibria computed for a sample of schools, as an illustration (results for all schools used in this exercise are in Appendix **??**). The first and second columns report school ID and number of observations; the third column contains the estimated school-wide interactions parameter; average smoking in each school in the data is reported in column 4; columns 5-7 report the computed equilibria in the case of multiple equilibria, or the single equilibrium that arises for that school.

Actual smoking averages are very close to smoking level in one of the simulated equilibria in all schools; small differences are due to the fact that we had to discretize the support of the GPA variable for computational reasons. Therefore, this model specification captures the large variation in smoking behavior across schools very well. Multiple equilibria are present in 40 out of 41 cases, given our parameter estimates (see, e.g, schools #41 and 56 in the table).[31] This validates our approach as it shows that in this application and given the data multiple equilibria are prevalent. Finally, in 35 out of the 40 schools that exhibit multiple equilibria, the actual fraction of smokers is consistent with the intermediate equilibrium (this is the case, for instance, for school #56). Thus in most of the 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.

To give an idea of the magnitude of the changes in smoking level across equilibria, we computed the average smoking that would arise if all schools that are estimated to be in the middle equilibrium were instead in the lowest smoking equilibrium (details for each school are reported in the last column of Table **??** in the appendix). The differences are economically significant, ranging from 6.3 to 31.6 percentage points. The average difference in smoking across schools in our sample (weighted by the number of observations in each school) that would arise if all schools were in the lowest equilibrium is 13.9 percentage points. This is quite a large difference, illustrating the importance of equilibrium selection in determining aggregate smoking behavior.

Before turning to the counter-factual exercises with school-wide only interactions, we illustrate the effects of changes in the strength of social interactions J_n in Figure 6. 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

[31] We only report results for 41 of the 45 schools used in the estimation. The remaining four schools also exhibit multiple equilibria but for these the equilibrium mapping lies at the knife-edge case in which the low and intermediate equilibrium coincide. To capture this, our simulation algorithm requires a much finer discretization than that needed for the other schools, therefore for convenience we drop them from this exercise.

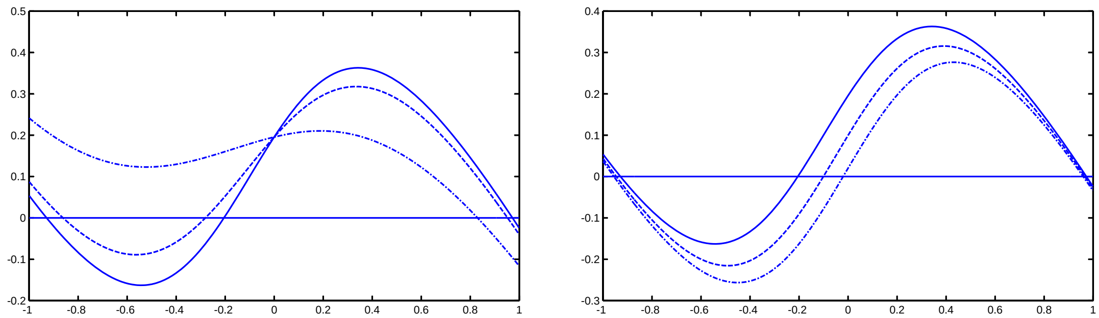

Figure 6: Equilibrium mapping as a function of global interactions J (left pane) and direct utility from smoking C_n (right pane). The mapping becomes flatter' as J decreases and moves down as C_n decreases

School ID	Interactions	Percentage of smokers			
		Data	Equil 1	Equil 2	Equil 3
	Baseline	5.9	5.9	17.2	99.8
41	J reduced 3%		9.1	11.8	99.8
	J reduced 5%		99.8	–	–
	J reduced 10%		99.8	–	–
	Baseline	18.7	0.2	18.4	100.0
56	J reduced 3%		0.3	17.0	100.0
	J reduced 5%		0.4	16.0	100.0
	J reduced 10%		0.9	13.0	100.0

Table 10: Simulations with global interaction estimates

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). This is in contrast to simple regression analyses that did not consider the existence of multiple equilibria: here a change in the strength of one estimated parameter would have a univocal effect on the outcome variable. 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 10 reports two examples to illustrate our findings (the full results of our exercise are available from the authors). The top panel of Table 10 focuses on school #41. Here the prevalence of smokers in our sample is 5.9%, and the school is in the low equilibrium. If the strength of social interactions is reduced by 3% the fraction of smokers in the low equilibrium actually rises to 9.1% (consistently with Figure 6, smoking prevalence increases in the low equilibrium). When the school-wide interaction parameter is reduced even further, by 5% and 10%, a large discontinuity occurs: The equilibrium mapping flattens to the point that it only crosses the 45 degree line only once, at the high equilibrium. Thus smoking prevalence jumps to a very high level, where nearly everyone smokes.

The bottom panel, representing school #56, tells a different story: here, because the school finds itself in the intermediate equilibrium, a reduction in J_n is accompanied by a reduction in smoking prevalence (from 18.4% to 13.0% for a 10% decrease in the interaction parameter). Again, if social interactions are reduced even further (e.g., reducing J_n by 20%), the multiplicity of equilibria disappears leaving only the high smoking equilibrium, in which almost everyone smokes.

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. It would be impossible a priori to predict the direction of a change in smoking behavior resulting from various policies without an estimation approach that fully incorporated multiple equilibria.

Finally, we study the effect of shutting down the social interactions channel completely. As expected, the multiplicity of equilibria disappears in all schools, with only the high equilibrium surviving everywhere (see Table **??**). What is interesting is that, at our parameter estimates, the high equilibrium at $J_n = 0$ is associated with a large fraction of smokers in most schools. The last column in Table **??** reports the difference in smoking between the estimated prevalence and the level of smoking that would prevail in this counter-factual. The average increase in smoking across all schools is about 70 percentage points. Thus, according to our estimates, the presence of social interactions actually reduces equilibrium smoking in our sample of schools. This is an important point since in the context of social pathologies such as smoking, drug and alcohol use, teenage pregnancies, crime or dropping out of school, network interactions (or peer effects) are often viewed as one of the channels that can help sustain and propagate such pathologies.[32] Here instead, shutting down the social interactions channel would lead to a large increase in the prevalence of smoking.

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 Accounting for the variation of smoking across schools

Results presented in the previous subsection show that most schools choose the intermediate-level of smoking equilibrium. Moreover, the correlation between the percentage of smokers and the global interaction coefficient is small, -0.04. Hence, even though equilibrium selection and preferences for interaction play an important role in determining the *level* of smoking, they do not seem to account for most of the variation in smoking across schools. The correlation between smoking and the coefficient relating the utility from smoking and the student's GPA, c_{GPA} is 0.01, whereas the correlaction between smoking adn c_{Age} is -0.16. Therefore, most of the variation seems to be driven by the different interaction of age and smoking across schools. We investigated this possibility by running various simulations.

First, we simulated equilibrium smoking when the preference parameters (c, J) are the same for all schools, and equal to the mean of their estimated values. Results are reported in the second row of Table 11.[33] We find that the standard deviation of smoking across schools in each type of equilibrium is smaller when compared to the standard deviation computed in the original parameter estimates. We interpret this result as indicating that heterogeneity in the distribution of students is not driving much of the smoking variation in the data, and that school heterogeneity is more important.

[32]See for instance Crane (1991) or Patacchini and Zenou (2009).

[33]We excluded from these simulations all schools with less than three equilibria.

Simulation	Eq. 1	Eq. 2	Eq. 3
Original estimates	2.7	4.8	0.1
All schools same c, J	0.6	3.0	0.0
All schools same set of students as School #56	2.2	6.0	0.2

Table 11: Standard deviation of smoking in various simulations

To confirm this intuition, we simulated smoking behavior using the estimated values of (c, J) in each school, but assumed that all schools have the same student body, that of of school #56. In the low level equilibrium, the variance of smoking in this simulation is very similar to the one computed from the estimates (see the third row of Table 11).[34] When schools have the same set of students, the variance of smoking in Equilibrium 2 is larger than in the original estimates, confirming that smoking variation is not determined by differences between the students' distributions across schools.

6.3 Counterfactuals 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; changing the direct (personal) utility of smoking (this experiment can also be interpreted as introducing a tax on cigarettes; changing the number of friends in students' friendship networks. Second, again given our estimated parameter values, we simulate the smoking process for an artificial school, to study the impact of different network structures (in the presence of student heterogeneity) on outcomes. In particular, we focus on two polar cases, 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

[34]Simulations performed using the student-body of different schools provide the same qualitative results.

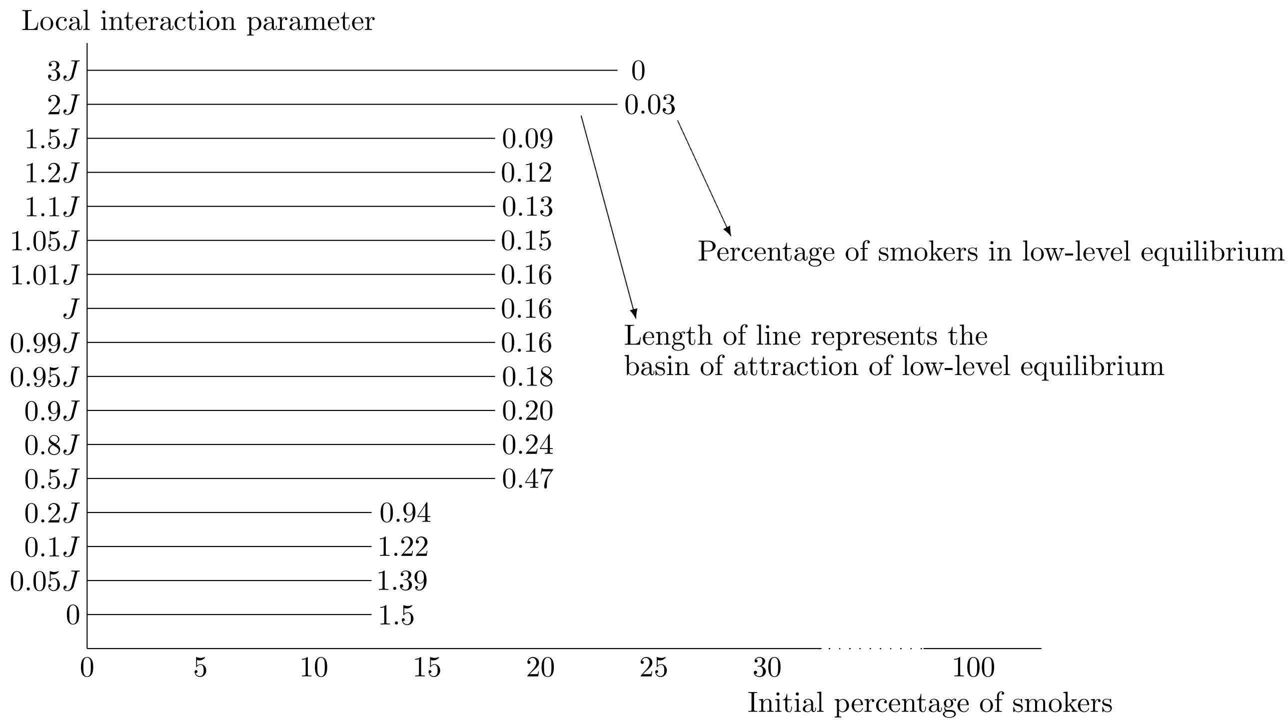

Figure 7: Percentage of smokers and basins of attraction of low-level equilibrium in a representative school for different values of the local interaction parameter (High level equilibrium at 100% smokers in every simulation)

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 1,000 iterations (we have also let the process run for up to 10,000 periods as a robustness check; the reported results do not change). We use the parameter estimates from the model specification with random effects (no school covariates Z_n) and both local and school-wide interactions, as a parallel to the specification we use for the counter-factuals with school-wide interactions only.

6.3.1 Results for a representative school

We choose school #56 as a representative school, both in terms of demographics and in terms of its parameter estimates. Figures 7-11 report the results of our counterfactual experiments to show the importance of multiple equilibria and non-linearities. Figure 7 reports the equilibria we find for several values of the local interactions parameter (displayed

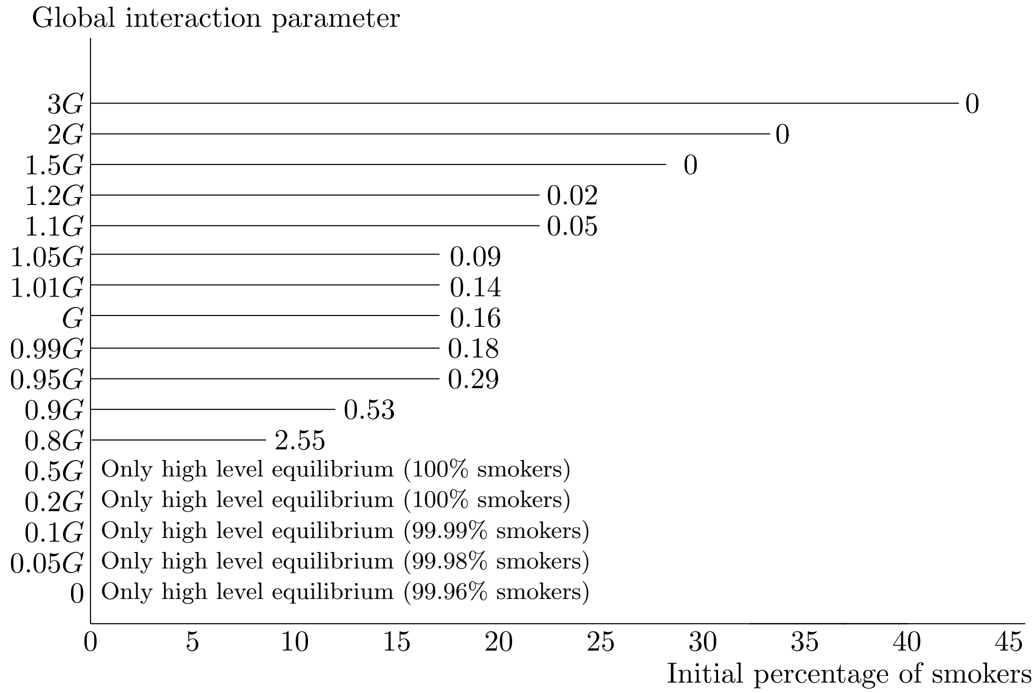

Figure 8: Percentage of smokers and basins of attraction of low-level equilibrium in a representative school for different values of the global interaction parameter. High level equilibrium at 100% smokers unless indicated

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 about 17%, the procedure converges to a low-level equilibrium with almost no students smoking. For all other initial levels of smoking, the procedure converges to the high level of smoking, with 100% of students smoking. Figure 8 repeats the exercise by varying the school-wide interaction parameter, Figure 9 by changing the number of friends, Figure 10 by changing both local and global interactions, and Figure 11 by changing the direct utility of smoking.

Overall, multiple equilibria are pervasive. In our simulations, only the high and low equilibria appear in the long run. Because we are using a dynamic process in which agents are myopic, 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

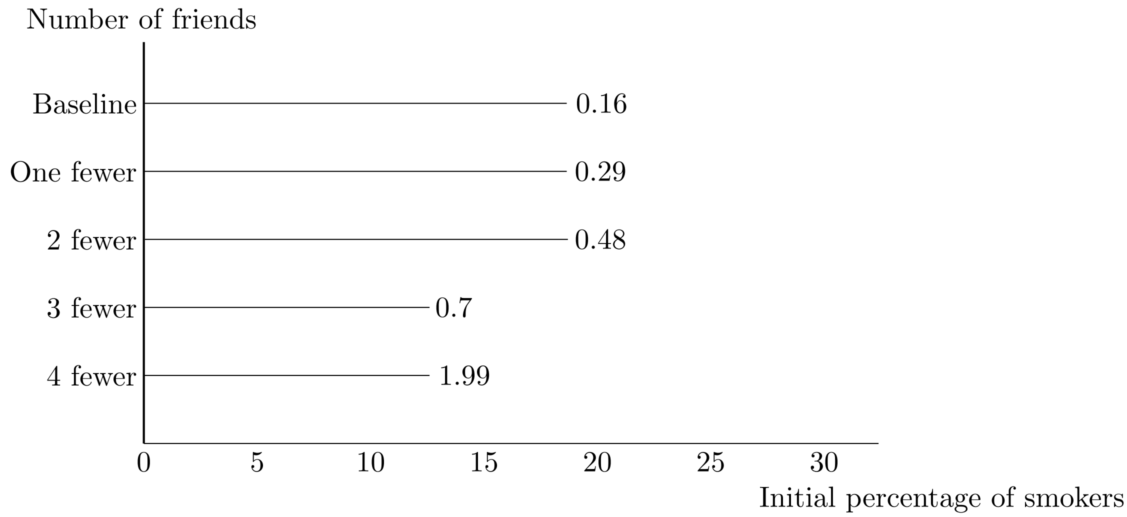

Figure 9: Percentage of smokers and basins of attraction of low-level equilibrium in a representative school for different number of friends

friends in students' personal networks, all go in the same direction (see Figures 7-9). 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 disappears; this is consistent with the left panel of Figure 6 (which illustrates movements in the equilibrium mapping following a change in G). The fraction of smokers increases slightly 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 10): 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.

Therefore, as is the case for school-wide only interactions, a decrease in the magnitude of social interactions leads, in a very non-linear fashion, to a sharp increase in the prevalence of smoking in the representative school. For low levels of social interactions, only the high equilibrium survives, in which almost every student smokes. This confirms the observation made earlier, that the presence of social interactions may actually lead to lower equilibrium smoking than would be the case in their absence.

Figure 11 reports the effect of changes in the individual cost of cigarettes (measured in "utils"): specifically, a decrease in the cost of smoking is modeled as an increase 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

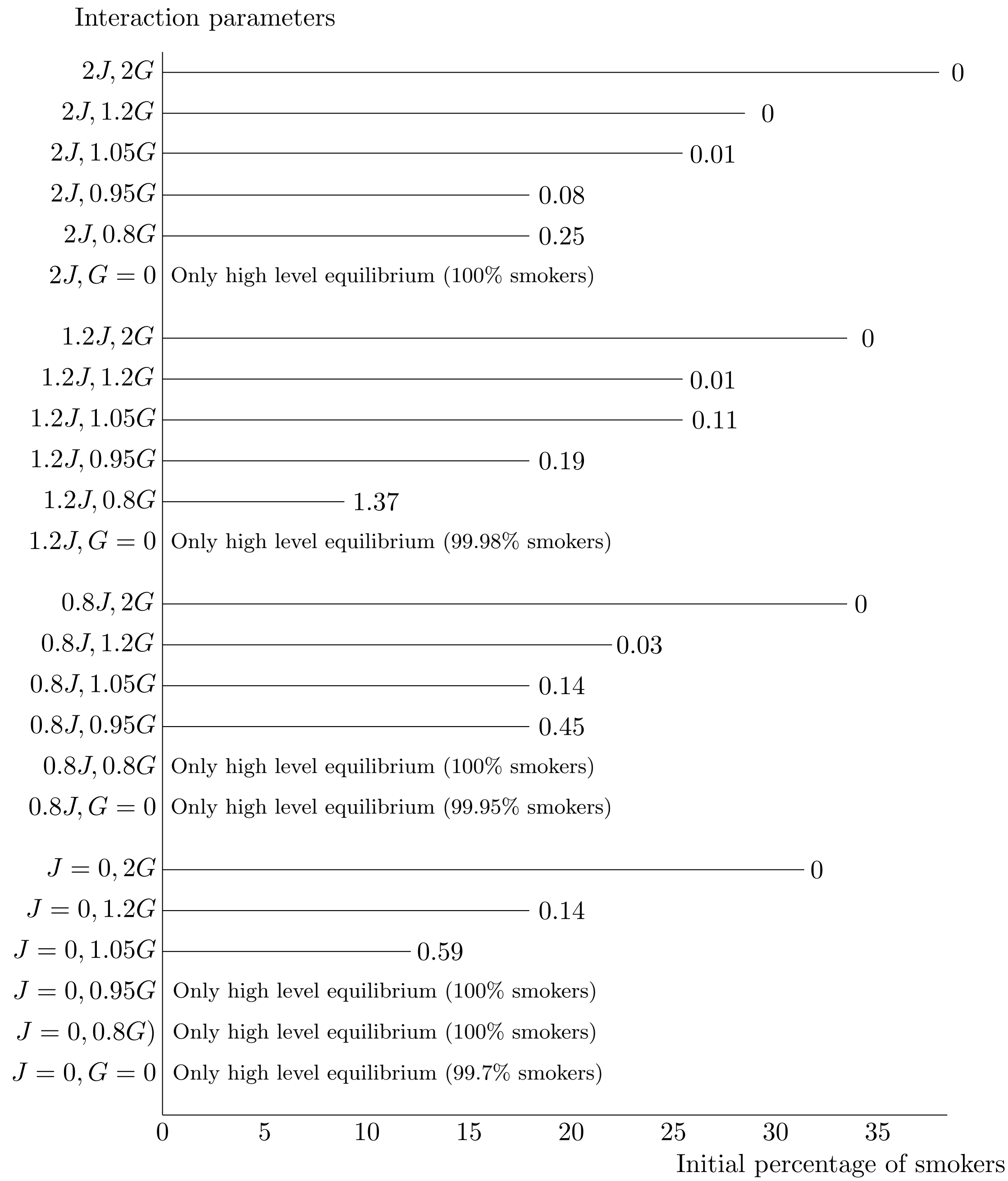

Figure 10: Percentage of smokers and basins of attraction of low-level equilibrium in a representative school for different values of local and global interactions. High level equilibrium always 100% smokers unless indicated

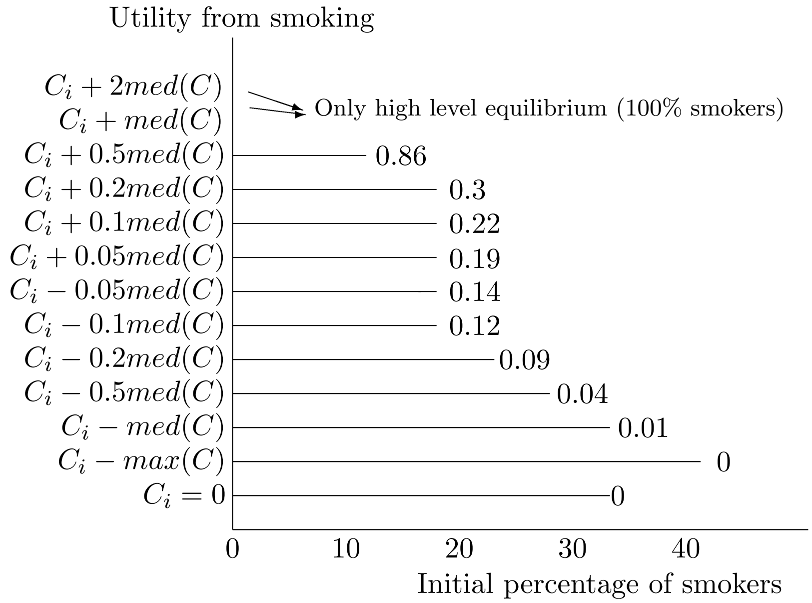

Figure 11: Percentage of smokers and basins of attraction of low-level equilibrium in a representative school for different values of direct utility from smoking

Figure 6. 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 everyone smokes.

6.3.2 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: 16 years old and has a 3.0 GPA. Subsequently we allow for different types of students, with varying propensity to smoke.

Figures 12-14 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 representative 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

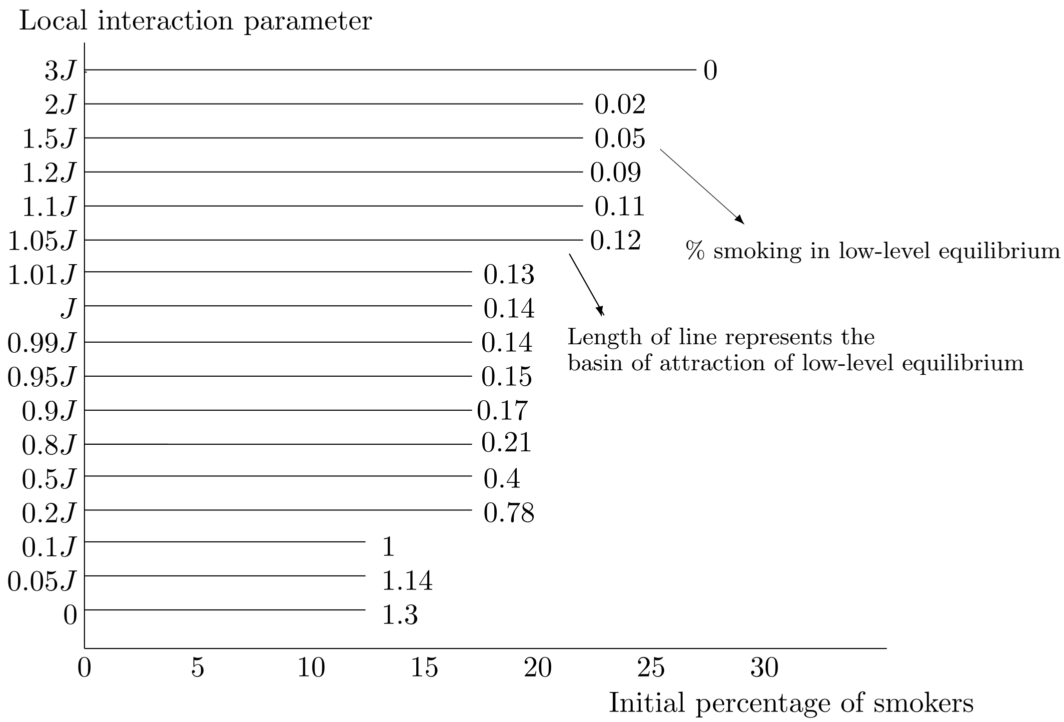

Figure 12: Percentage of smokers and basins of attraction of low-level equilibrium in artificial school for different values of the local interaction parameter. High level equilibrium at 100% smokers in all simulations

equilibrium to become smaller and eventually disappear. While the low equilibrium persists, 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 qualitative effect on the basin of attraction of the low equilibrium and on the fraction of smokers in this equilibrium while it persists. Thus, as in the representative school case, stronger social interactions and/or a larger number of friends tend to reduce smoking according to our estimates.

Figure **??** 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 younger students, or those with a higher GPA, the intercept shifts down relative to our baseline student: this induces the basin of attraction of the low smoking equilibrium to widen. Changing our baseline artificial student to an older student, or to a student with a lower GPA, induces the opposite effect.

Figures **??** and **??** 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

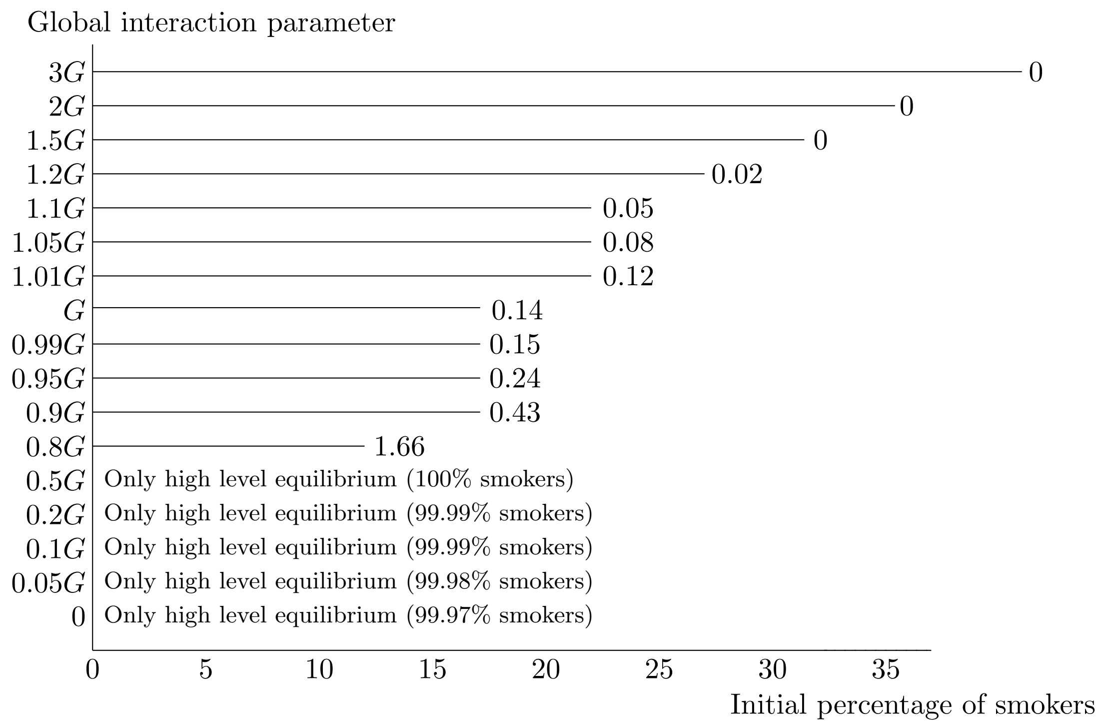

Figure 13: Percentage of smokers and basins of attraction of low-level equilibrium in artificial school for different values of the global interaction parameter. High level equilibrium at 100% smokers unless indicated

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), focusing on two polar cases. In Figure **??** we vary the relative fraction of the two types, but maintain an arrangement that is "perfectly integrated": $LMLM...$, or $LLLMLLLM...$, or $LLLLLLLMLLLLLLLM....$ In Figure **??** 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).

The results can be summarized as follows. In both the perfectly integrated and the perfectly segregated school, the basin of attraction of the low smoking equilibrium widens as local or school-wide interactions increase, as the number of friends increases, and as the fraction of "high smokers" decreases in the school. In addition, in the segregated school there are some combinations of parameter values for which a new intermediate equilibrium appears. This equilibrium is characterized by a fraction of smokers that is equal to the

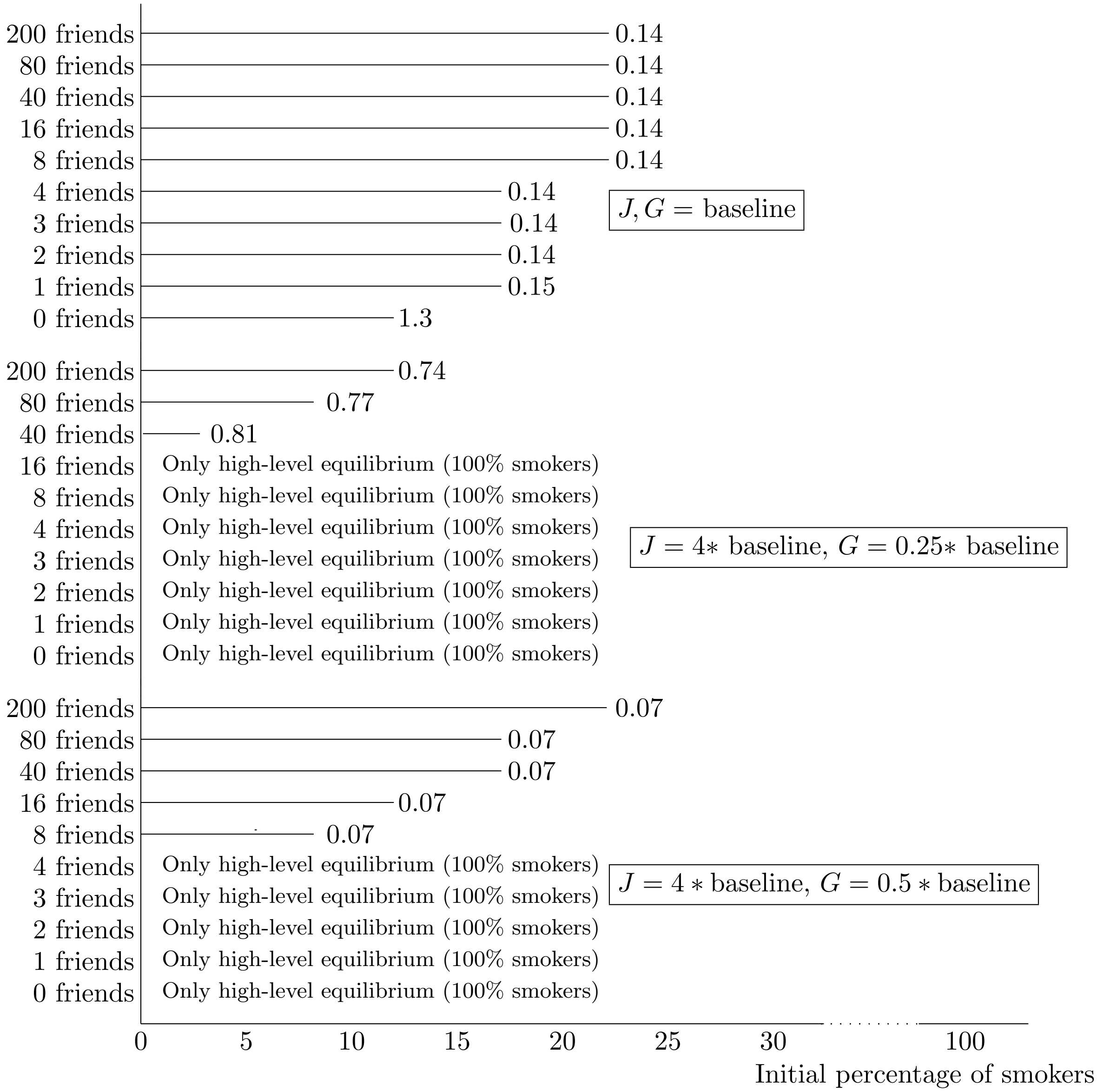

Figure 14: Percentage of smokers and basins of attraction of low-level equilibrium in artificial school for different number of friends and interaction parameters. High level equilibrium at 100% smokers unless indicated (continues on next page)

proportion of high smokers in the school. Thus, when students are heterogeneous, the arrangement of students within the school and the composition of students' personal networks play a role in the determination of equilibrium outcomes.

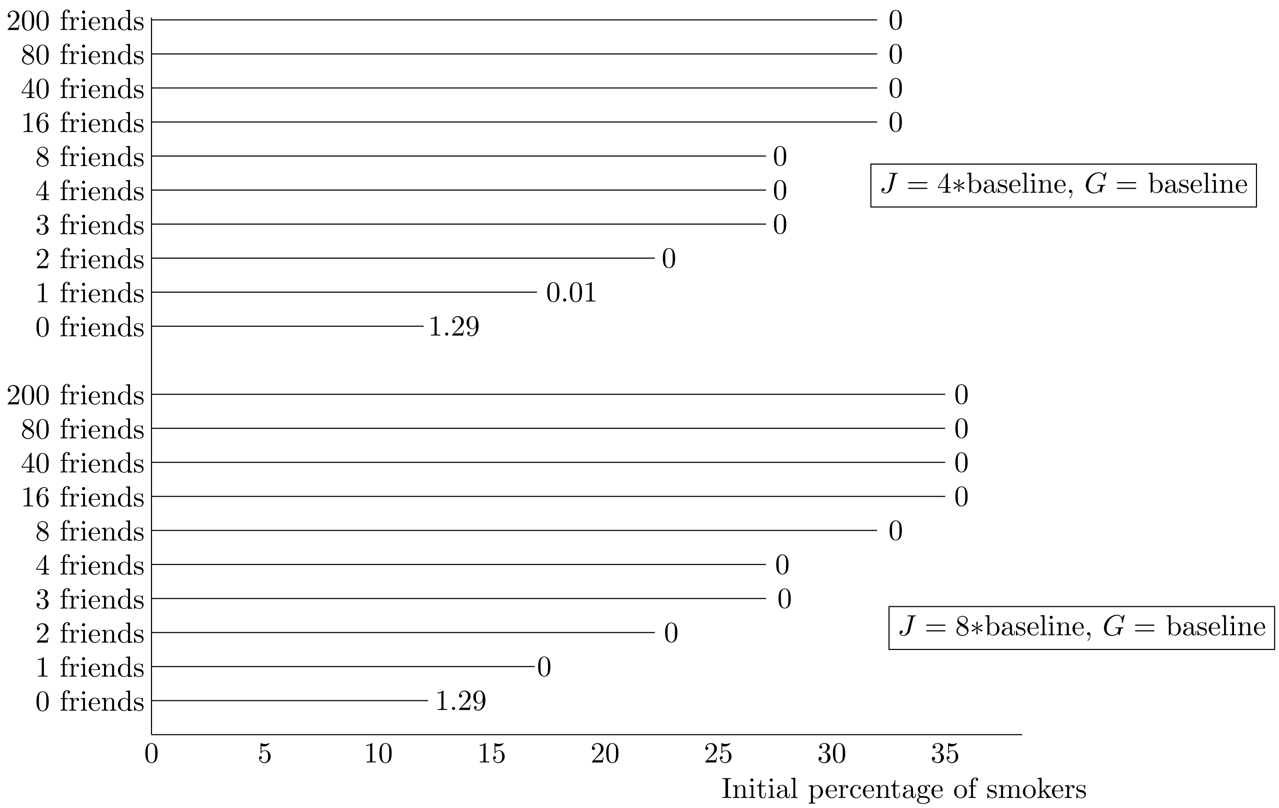

Figure 14: (continued) Percentage of smokers and basins of attraction of low-level equilibrium in artificial school for different number of friends and interaction parameters. High level equilibrium at 100% smokers unless indicated

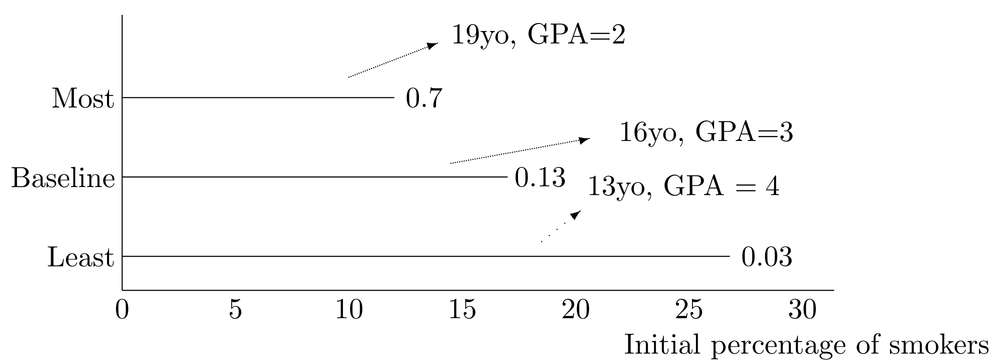

Figure 15: Percentage of smokers and basins of attraction of low-level equilibrium in artificial school for different artificial (homogenous) student. High level equilibrium at 100% smokers unless indicated

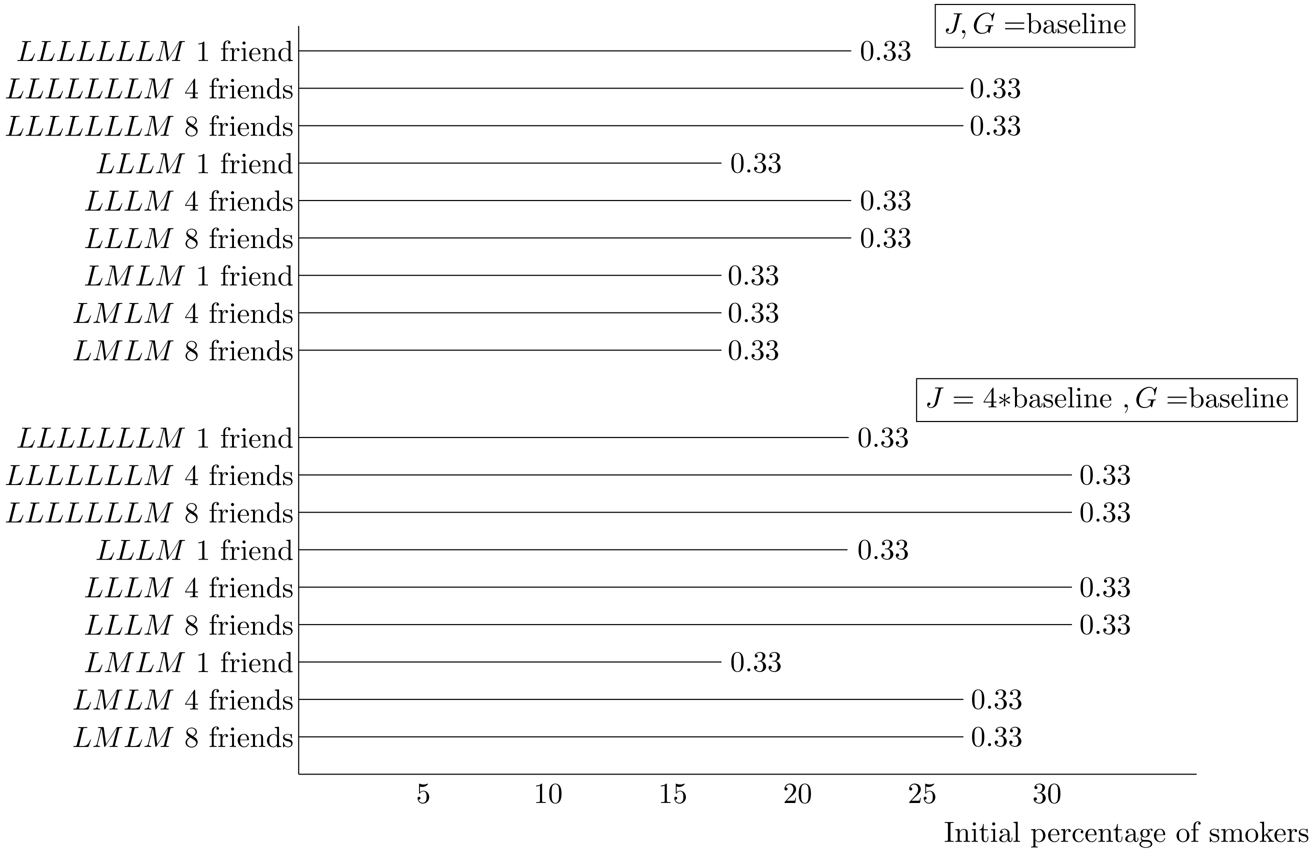

Figure 16: Percentage of smokers and basins of attraction of low-level equilibrium in perfectly integrated artificial school for different arrangements of two types of students: L: least smoker; M: most smoker. High level equilibrium at 100% smokers in all simulations

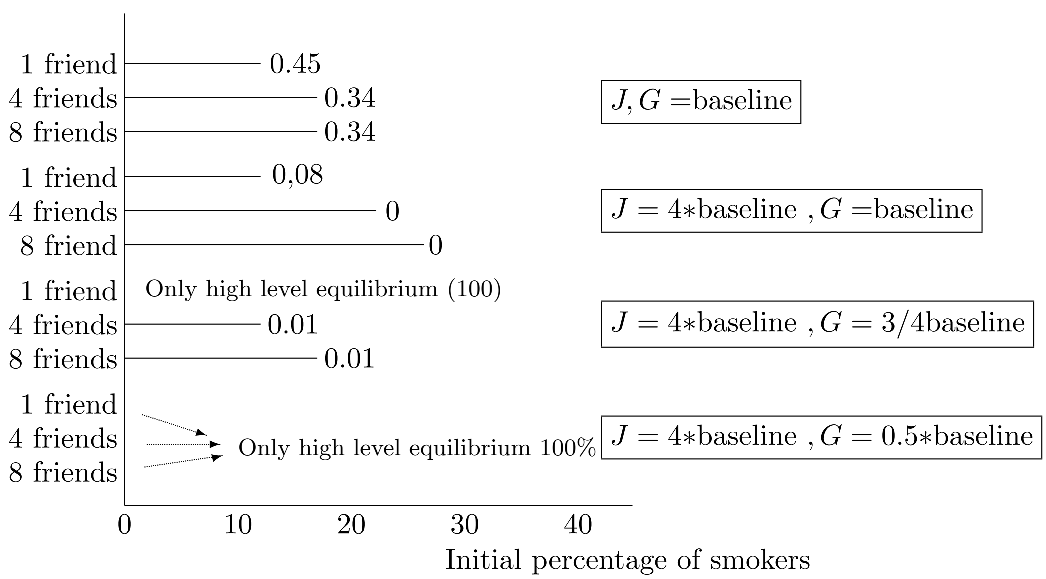

Figure 17: Percentage of smokers and basins of attraction of low-level equilibrium in perfectly segregated artificial school for different groupings of students. Type of students: L=50%, M=50%. High level equilibrium at 100% smokers unless indicated

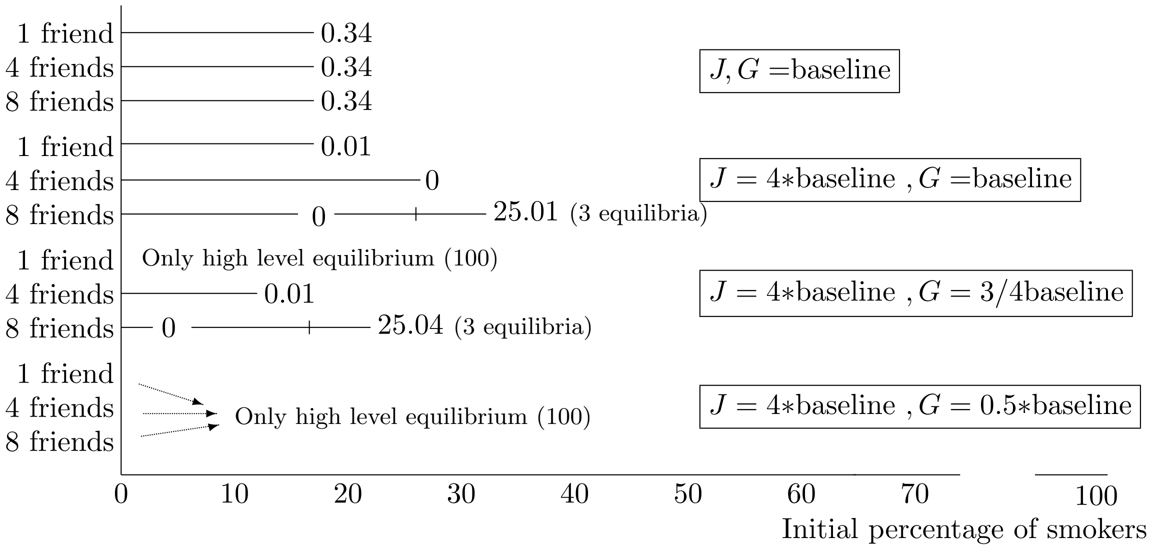

Figure 17: (continued) Percentage of smokers and basins of attraction of low-level equilibrium in perfectly segregated artificial school for different groupings of students. Type of students: L=75%, M=25%. High level equilibrium at 100% smokers unless indicated

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 counter-factual 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 an illustration 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 (Bisin-Moro-Topa (2010)), 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. Taking this endogeneity explicitly into account can help identify the extent of social interactions.

A Simulation results for all schools

School ID	N	Local int. (J)	Percentage of smokers Data	Simulated eq.	Equil. 1	Equil. 2	Equil. 3	Pct. Diff.
13	444	7.78	15.5	15.1	0.2	15.4	100.0	15.3
18	417	3.48	19.7	20.1	2.4	18.9	100.0	16.6
23	427	6.40	22.2	18.3	0.1	22.1	100.0	22.0
29	602	7.28	15.9	13.1	0.2	15.8	100.0	15.7
31	866	4.25	23.7	23.7	0.7	23.3	100.0	22.6
33	479	3.19	23.2	24.4	2.2	22.6	100.0	20.4
34	418	3.48	16.3	13.6	4.3	15.3	100.0	11.1
35	501	3.52	19.2	16.2	2.3	18.8	100.0	16.5
40	458	5.75	22.9	19.7	0.2	22.8	100.0	22.6
41	1056	2.67	5.9	5.3	5.9	17.2	99.9	0.0
42	679	5.01	18.9	18.6	0.6	18.7	100.0	18.2
44	939	2.37	18.2	16.9	7.8	17.3	99.8	9.5
47	452	3.79	24.6	24.3	1.1	24.2	100.0	23.2
49	896	3.97	19.0	17.3	1.6	18.5	100.0	16.9
50	1085	3.63	18.3	19.0	2.1	18.1	100.0	16.0
52	479	2.66	7.1	9.2	8.0	13.5	99.9	0.0
53	935	2.32	19.8	17.2	7.5	19.0	99.7	11.5
54	448	4.05	25.2	26.6	0.6	25.0	100.0	24.4
56	1093	6.50	18.7	17.6	0.2	18.5	100.0	18.3
57	421	2.73	21.6	18.8	4.5	20.8	99.9	16.3
58	578	2.95	33.6	32.5	1.4	33.0	99.9	31.6
60	1105	3.87	21.2	18.6	1.4	20.7	100.0	19.3
62	1510	2.45	22.7	20.5	5.6	22.0	99.7	16.4
65	1225	3.45	6.4	7.0	7.9	100.0	—	0.0
67	449	3.91	24.9	24.3	0.8	24.7	100.0	24.0
71	549	3.24	20.9	22.8	2.5	20.6	100.0	18.2
72	899	0.14	24.4	25.0	24.8	—	—	0.0
73	422	3.41	15.2	17.1	4.1	14.4	100.0	10.3
74	435	3.53	20.7	20.5	2.1	20.1	100.0	18.0
76	662	5.50	16.2	17.4	0.7	16.0	100.0	15.3
77	818	2.44	9.0	8.8	9.3	13.9	99.8	0.0
82	452	6.64	20.6	16.8	0.1	20.3	100.0	20.2
86	911	2.72	19.5	17.1	5.4	18.4	99.9	13.1
87	450	2.88	23.3	21.3	3.9	22.2	99.9	18.4
91	862	2.09	25.3	25.2	7.6	24.4	99.2	16.8
149	484	6.74	9.1	9.7	1.3	8.8	100.0	7.6
259	936	5.11	26.3	24.8	0.2	26.0	100.0	25.8
267	402	2.85	17.9	13.9	5.8	16.6	99.9	10.8
268	792	2.50	9.6	10.0	9.8	99.9	—	0.0
269	711	7.48	26.6	25.5	0.1	26.5	99.9	26.4
271	552	5.67	9.1	7.6	2.3	8.6	100.0	6.3
Simple average								15.0
Weighted average (by n. of students)								14.3

Table 12: Equilibrium smoking, actual smoking, and percent difference with lowest smoking equilibrium in every school

School ID	N	Percentage of smokers Data	Simulated eq.	Equilibrium	Pct. Diff.
13	444	15.5	100.0	99.8	-84.4
18	417	19.7	93.4	93.5	-73.7
23	427	22.2	99.6	99.3	-77.4
29	602	15.9	99.9	100.0	-84.0
31	866	23.7	95.7	96.5	-72.0
33	479	23.2	89.3	88.1	-66.1
34	418	16.3	91.8	90.7	-75.5
35	501	19.2	93.9	90.8	-74.7
40	458	22.9	99.0	99.3	-76.1
41	1056	5.9	85.4	86.8	-79.5
42	679	18.9	99.1	98.5	-80.2
44	939	18.2	81.4	80.8	-63.1
47	452	24.6	91.5	91.2	-66.9
49	896	19.0	96.1	96.1	-77.1
50	1085	18.3	95.5	94.9	-77.1
52	479	7.1	86.2	85.4	-79.1
53	935	19.8	78.6	74.4	-58.8
54	448	25.2	94.6	94.9	-69.4
56	1093	18.7	99.8	99.6	-81.1
57	421	21.6	83.9	80.8	-62.2
58	578	33.6	74.1	72.0	-40.5
60	1105	21.2	94.3	93.7	-73.1
62	1510	22.7	78.3	79.2	-55.5
65	1225	6.4	95.3	96.5	-88.9
67	449	24.9	93.9	92.9	-69.0
71	549	20.9	91.0	91.3	-70.0
72	899	24.4	27.3	27.3	-2.9
73	422	15.2	94.0	94.8	-78.8
74	435	20.7	92.5	93.3	-71.8
76	662	16.2	99.5	99.8	-83.3
77	818	9.0	82.6	82.5	-73.5
82	452	20.6	99.6	99.6	-79.0
86	911	19.5	84.6	83.0	-65.1
87	450	23.3	81.1	80.7	-57.7
91	862	25.3	70.4	70.4	-45.1
149	484	9.1	100.0	99.8	-90.9
259	936	26.3	97.2	97.5	-70.9
267	402	17.9	86.3	86.3	-68.3
268	792	9.6	85.0	85.9	-75.4
269	711	26.6	99.8	99.7	-73.1
271	552	9.1	99.8	99.8	-90.7
Simple average					-70.7
Weighted average (by n. of students)					- 69.8

Table 13: Simulation with global interactions set to zero. Equilibrium smoking, actual smoking, and percent difference with lowest smoking equilibrium in every school

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.

Chiappori, P. A. and Ekeland, I. (2009): “The Micro Economics of Efficient Group Behavior: Identification,” *Econometrica*, 77, 763-99.

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 Disequilibrium 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.

Gilleskie, Donna B. and Koleman S. Strumpf (2005): ”The Behavioral Dynamics of Youth Smoking”, *Journal of Human Resources*, XL(4):822-866.

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.

Gruber, J. and J. Zinman (2000): ”Youth Smoking in the U.S.: Evidence and Implications”, NBER Working Paper No. 7780, July 2000.

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.

Patacchini, E. and Y. Zenou (2009): "Juvenile Delinquency and Conformism", CEPR Discussion Paper 7565.

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.