Measuring Social Interactions

by

Edward L. Glaeser

Harvard University and NBER

And

José A. Scheinkman*

Princeton University

August 30, 1999

Abstract

This paper presents on overview of the economics that lies behind social interaction models and briefly discusses the empirical approaches to social interactions. We present a simple model with local interactions, similar to Glaeser, Sacerdote and Scheinkman (1996) but using a continuous action space and starting with optimizing behavior. We then extend the model to include both global and local interactions. We suggest and use a methodology for using variation of intra-city aggregates to identify the relative sizes of local and global interactions. We also present a model with endogenous location choice and use the predictions of that model to identify the sources of cross-city variance that are due to sorting and interaction. Finally, we present a brief discussion of using time-series to estimate the social interactions in broad aggregates.

I. Introduction

A growing literature has argued that many economic actions -- crime, education choice, labor force participation, out-of-wedlock births-- are marked by social interactions (see e.g. Akerlof, 1997, Becker, 1997, Bernheim, 1994, Young, 1997).[1] These social interactions imply that the net private benefits from pursuing a particular activity rise as others also pursue this activity. For example, working hard in school might be less painful for a young student if his friends are also studying, both because his friends can help him learn and because his friends are not available for other leisure activities. These interactions can take many forms, ranging from pure physical externalities (while one person is being arrested, the police find it harder to arrest someone else), to learning from one's neighbors, to stigma (the more people who are committing a particular crime--the less likely is that crime to be a negative signal) to pure taste externalities (individuals just enjoy imitating others). We will discuss the many forms of these externalities at length in Section II, but the primary focus of this essay is on measuring the extent of social interactions, not on determining which mechanisms are most important in generating them.

Social interactions are particularly important because they can help explain striking shifts in aggregate outcomes over time and space. There are a large number of variables where shifts over time and space seem far too large to be explainable with standard economic forces. For example, Levitt (1997) shows that only 25 percent of the massive crime increase from 1960 to 1975 can be explained by demographic shifts. Mulligan (1995) argues that massive female labor force participation in World War II cannot be explained by changes in either wages or the opportunity cost of time.

Likewise the dramatic change in divorce rates or the rise in out-of-wedlock births (see Akerlof, Katz and Yellen, 1996) all seem to be only partially connected to visible shifts in observable variables.[2]

Social interactions help to explain these changes, because of the strategic complementarities inherent in social interactions. These strategic complementarities imply that even if changes in fundamentals create only a small change in the level of activity for each individual, each individual's small change will then raise the benefits for everyone else pursuing the activity. The society-wide effect of a small change in fundamentals, because of these ripple effects, may therefore be quite large. Small changes in fundamental variables can set off a cascade in individual behavior so that large shifts in outcomes may result from tiny changes in fundamentals.

The rapid shifts in the variables that we mentioned earlier are of prime policy interest. The rise in female labor force participation is probably the most important single shift in the post-war labor market. The rise in crime over the 1960-75 period led to a ten-fold increase in reported crime in many areas. The rise in out-of-wedlock births and the rise of divorce appear to have caused deep changes in our society. To the extent that theory and measurement of social interactions enables us to understand these massive changes, the study of social interactions has potentially major policy relevance. Furthermore since social interactions usually imply the existence of externalities, the presence of these interactions often suggests some scope for government action.

Indeed, we believe that non-market interactions between people represent most of the human experience. These interactions play a critical role in determining behavior, preferences and utility. Social interactions models of the type discussed in this paper and in this volume are one way of understanding the features of non-market interactions that make them different from more standard interactions that work through market transactions. This paper focuses on one empirical approach to these interactions.

Measuring Social Interactions-- A Brief Literature Review

This paper focuses on a narrow set of issues in empirically measuring the size and nature of social interactions. There are several empirical approaches to understanding these interactions. There is a literature that includes Crane (1991), Case and Katz (1991), Evans, Oates and Schwab (1992), Rauch (1994), Borjas (1995), O'Regan and Quigley (1997) and many others that uses micro-data to examine these connections. The basic structure of this research often involves regressing an action of a person on the average action of a person's "neighbors," where neighbors can mean members of the individuals census tract or some self-reported social group.

There are three problems with this methodology, which are discussed at length by Manski (1993). First, if a person is affected by his neighbors, he also affects his neighbors. As such the supposedly independent variable (the neighbors' actions) is a function of the dependent variable (the individual's actions). Most recent research in this area (see Case and Katz, 1991) addresses this problem by instrumenting for the independent variable using the average levels of other neighbors' characteristics which are supposedly exogenous (such as neighbor's parents characteristics). Second, there may be omitted variables in a particular area which affect the returns to the activity in that area and which would induce a spurious correlation between individuals and neighbor's actions, even if all individuals are immobile. This problem is also potentially treatable using exogenous neighbor's characteristics as instruments for neighbor's actions.

Third, individuals choose their neighborhoods and individuals who are likely to do the same things may choose to live close one another, perhaps because of social interactions. Evans, Oates and Schwab (1992) address this problem by modeling the choice of peer group as an endogenous variable. They argue that standard peer group effects disappear once the endogeneity of peer groups have been properly treated.

General solutions to all of these problems are enormously difficult to find in the absence of controlled experiments, such as Gautreaux or Moving-to-Opportunity, where individuals are actually randomized across neighborhoods. Even these experiments often suffer from the fact that we only observe individuals who chose to join in the experiment or who decide not to turn down the opportunity to move to a new neighborhood. When individuals are selected based on moving, the results are clearly biased because only persons who benefit from moving will choose to move.

Even clever solutions to this bias that use only the randomized part of the experiment are problematic. For example consider an experiment where randomized individuals (perhaps those who draw an even number) are given the opportunity to move to a new neighborhood and others aren't given that opportunity (perhaps those who draw an odd number). By using the number that the individual is given as an instrument for neighborhood movement (thus not using whether or not the individual actually moved), some of the worst part of the bias is eliminated. Nevertheless, since the only people who move are those who benefit from moving, the experiment never tracks the full sample of possible movers. Even the randomized treatment effect must be interpreted as estimating the benefit of having an option to move, not the benefit of actually moving.

Brock and Durlauf (1997) represents a particularly comprehensive and careful discussion of the use of micro-data to estimate social interactions. In particular, they focus on discrete choice problems often in a panel setting. A major contribution is their presenting a thorough discussion of when discrete choice models with social interactions are actually identifiable. Again, though, identification is shown by them to be extremely difficult in many cases, especially when unobserved heterogeneity is particularly important.

Another empirical approach to measuring social interactions relies on using only aggregate information (see Brock and Durlauf, 1995, Glaeser, Sacerdote and Scheinkman, 1996, Gaviria, 1997, Topa, 1997). The intuition of this approach is that since social interactions create high levels of variance across space and time, by using the variance of aggregates, one can measure the extent of these interactions. This approach is free of the most basic endogeneity or reflection problem, because the approach explicitly acknowledges the fact that all individuals effect each other. However, it is free of neither the problem of omitted variables which vary across space, nor of the problem of selection of different types of people into different areas.

Alternative approaches have been proposed by Glaeser, Sacerdote and Scheinkman (1996) to address these problems. We implicitly control for an area specific fixed effects which eliminates some or most of the omitted variables problem. We examine groups which are more or less mobile to see if there appears to be a connection between mobility and measured social interactions, which there would be if measured social interaction just reflected location choice. We use "scaling" rules predicted by the theory that should allow us to differentiate between sorting and direct interaction. Finally, we use the variance of observables to determine the range of reasonable estimates for the importance of unobservables. While these corrections are far from perfect, they do suggest that there are ways that this methodology can be made useful. We believe strongly that given the importance of estimating social interactions, all possible methodologies should be used. Even if the classic approach discussed first was better in 90 percent of the cases (which we do not believe), there is still significant value in using alternative methodologies which do not share exactly the same set of problems (although they have problems of their own).

Topa (1997) also uses aggregate-level variables to study social spillovers in employment status. Formally he writes down a non-homogeneous version of a contact process in which the probability of becoming employed, depends on both individual characteristics and the number of one's neighbors who are employed. The probability of becoming unemployed depends only on individual characteristics. The non-homogeneity allows Topa (1997) to differentiate spatial sorting from spillovers, but it also stops him from explicit derivations of the stationary distribution of the employment process. Instead, he uses the process of indirect inference where parameters are estimated by minimizing a distance between actual data and simulations of the structural model for different parameter values. A principle feature of the Topa-model is that the covariance between individuals--the degree of social interaction-- is determined by spatial distance. He estimates large quantities of spillovers using Chicago Census Tract level information for 1980 and 1990. He finds that spillovers are strongest for minorities and individuals with less education.

Brock and Durlauf (1995) do not present estimation based on aggregates, but rather present a variety of theoretical results which are presented as a first step towards empirical work. They focus on a global interactions model and produce results on the existence of multiple equilibria and the existence of threshold effects.

Outline of this Paper

Our focus is one measuring the size and nature of social interactions. Our particular interest is in interactions where one person's taking a particular action increases the likelihood of another person also taking the same action. We will generally mean the term positive social interactions to refer to just this type of situation. Most of the peer effects and interaction models discussed above (and discussed below) can be said to have this basic structure.

Our primary interest is in empirically determining the size and nature of these positive social interactions. We are interested in the extent to which one person's action will effect his neighbor's action. We are interested in the extent to which this sort of influence decays with geographic and social distance. We are interested in the extent to which individual interactions are increased and reduced as individuals choose the social milieu in which they exist. In principle, if social interactions are to be a major piece of positive economics or policy prescriptions, this type of information is crucial.

This paper extends our previous methodology is four ways, starting with Section III. First, we introduce a social interactions model with a continuous rather than a discrete one-zero choice variable. This change is useful for considering many variables where outcomes are continuous, rather than discrete. If we believe that the action is continuous but that the econometrician only observes a discrete outcome, then this continuous interaction model can be used for thinking about discrete variables. We present a new set of empirical results measuring the extent of social interactions for these continuous variables. One primary difference between continuous and discrete variables is that to use continuous variables it is necessary to have a separate estimate of the population variance of outcomes from micro data (in the case of discrete variables with known mean level p, the population variance is always p(1-p)).

Our second section extends our previous work to include both local and global interactions. A local interaction occurs across neighbors. A global interaction occurs through an aggregate. Classic examples of local interactions may include learning from neighbors (as in Ellison and Fudenberg, 1995) or joint neighbor production of non-work related activities. Global interactions may include community-wide norms or effects that work through the price mechanism. Like local interactions, global interactions produce high variances. Unlike local interaction models, global interactions also naturally produce multiple equilibria, which local interactions do not as long as the interaction from neighbor-to-neighbor decays sufficiently quickly. We demonstrate an empirical methodology for considering multiple equilibria and other social interactions jointly. This methodology finds the existence of multiple equilibria for many variables, but that the bulk of the variance across areas remains even after we have allowed for the existence of multiple equilibria. Actually separating local from global interactions requires sub-area aggregates or micro data where individuals are matched to a peer group below the global level.

Our third section presents a version of the model with both local interactions and locational selection. Individuals choose their areas to maximize utility based on possibly limited information about their own tastes. The variance across areas is then based on the combination of locational decisions and social interactions (of course, the local decisions are also based on the existence of social interactions). The identification of selection vs. social interaction hinges again on a scaling rule. In other words, if we know that people are selecting between sets of areas with different population sizes then it is possible to differentiate between the two sources of cross-area variation.

Finally, we examine local and global interaction models in a dynamic context. Following a large body of work on technology adoption, we note that the level of social interaction determines the extent to which adoption is linear or S-shaped. We present a simple means of testing for the extent of social interaction in dynamic processes, but we do not show how to determine between local and global interactions outside of using simultaneously cross-sectional and time-series information. We present a ranking across a number of dynamic processes of which appear to be the most interactive. In general local interactions seem to generate somewhat slower dynamic change, and in principle it may be possible to differentiate between the two theories just using time series information given sufficient assumptions on functional form.

Our overall conclusion is that it still appears that there is substantial social interaction in a large number of variables. Some of this interaction creates multiple equilibria, but most of the variance that social interaction creates occurs beyond these equilibria. Differential selection into different areas is clearly particularly important, but there is still variance beyond that caused by selective migration.

II. Discussion of Interactive Mechanisms

There is no shortage of the mechanisms that may generate social interactions of either the local or the global variety. Furthermore, while we will stress "positive" social interactions, i.e. interactions where an individual’s action positively influences his neighbors' actions, there are also many well-known cases of negative social interaction. For example, competition for scarce resources is a form of a global negative interaction which operates through the price system. As one individual decides to consume more of a particular commodity, that individual drives up the price and drives down consumption of all others who also face that price. Because of this force, we generally expect to find positive social interactions in actions where there are not scarce resources for which individuals are competing. We loosely divide the mechanisms that generate social interactions into four primary categories: physical, learning, stigma and taste-related interactions.

One reason to care about the different reasons why social interactions occur is that there are different policy implications associated with different interactions. For example, if one person's level of education increases his neighbor's education through dissemination of learning then it makes sense to subsidize education. There is a socially desirable spillover that should be subsidized. However, different policy implications appear if one person's level of education increases his neighbor's education for signaling reasons, i.e. as one person gets more education the other person must also get more education or be thought inferior. In that case, there is a socially undesirable spillover that should not be subsidized. While we will not be able to delve into methods of differentiating the sources of spillovers in this paper, this section stresses the wide range of possible mechanisms and the extreme policy importance of recognizing the different ways in which positive interactions might occur.

Physical and Learning Interactions

There are many forms of physical social interactions, even just within a single activity. For example, social interactions may occur in crime because of congestion in law enforcement (as in Sah, 1991). This force surely plays a significant role in riots, where the large number of rioters lowers the probability of arrest (see DiPasquale and Glaeser, 1997). Increases in crime may lower the opportunity cost of legal activity (because legal actors are being robbed) and may therefore lower the opportunity cost of time and raise further the amount of criminal activity (as in Murphy et al., 1991). These interactions may either be local or global depending on the range of criminals and police. For example if criminals attack legal businesses throughout the area, then this interaction is global. If criminals only attack very close legal operations then the interaction is local.

Network externalities are a classic physical interaction. In these externalities, it is more valuable to use a technology when others are using it as well. For example, telephones and e-mail become more valuable when others also have these communication devices. Cities themselves are networks and the existence, growth and decline of urban agglomerations depend to a large extent of these interactions.

The presence of investment also can generate these physical interactions. Investing in learning the QWERTY keyboard may only make sense when a large fraction of keyboards follow this configuration. Investing in an IBM versus a Macintosh or a Betamax versus a VHS video recorder depends on the presence of complements to use such as software or videocassettes. These complements are much more likely to abound when others are also using the technology. As a result there is a positive, global interaction that moves the nation to the extreme of using one or the other technology (as in the case of VHS vs. Beta, see Arthur, 1989, for a discussion of "historical lock-in") or an uneasy co-existence between two technologies (as in the case of IBM and Macintosh). In these cases, it has often been argued that suboptimal equilibria often continue to exist supported by social interactions.

Other social interactions based on learning may occur if individuals actually help each other learn (as in Benabou, 1993). In Young (1993), individuals learn by observing past actions and learning produces convergence of strategies to a Nash equilibria. Having neighbors who are taking an action makes it easier to learn about this action. This learning may take the form of just learning that a new technology exists (as in Griliches, 1958) or learning how to operate a technology correctly or learning the returns of this technology. Again, depending on how the technology operates, the interaction may either be local or global. Ellison and Fudenberg (1993) explicitly consider global learning where people interact with random members of a broad population. Ellison and Fudenberg (1995) examine local learning where people interact with their near neighbors. Fads and herding are other examples of behavior where learning-related externalities can create social interactions (Banerjee, 1992, Bikhchandani et. al. 1993).

Signaling and Taste Interactions

Interactions can also be generated through the desire to resemble outwardly the group that is taking a particular action. When actions are signals, then there is a natural interaction that comes about because the value of a signal is a function of who else is taking that signal. For example, Rasmussen (1996) develops a model of stigma and criminal behavior where more criminality tends to lower the stigma associated with criminality. As a result, more people become criminals. Glaeser (1992) argues for positive social interactions in labor market mobility, where more people changing firms in high mobility countries (such as the U.S.) eliminates the stigma associated with rapid mobility in low mobility countries (such as Japan).

Of course, the presence of signaling doesn't necessarily yield positive interactions. For some actions (particularly snob goods), greater participation necessarily means that the action goes from being a positive signal to being a negative signal (see e.g. Pesandorfer, 1996). As more people perform the action, or consume the snob good, there is less of a signaling demand for the product. There is an inherent asymmetry between actions which are demanded because they are positive signals and actions which are avoided because they are negative signals. As more individuals perform actions which are positive signals, the signal dissipates and the value of the action disappears. As more individuals perform actions which were once negative signals, again the signal dissipates, but in this case the demand for the action will rise with the disappearance of the signal.

To make this point clearly consider the following simple model where individuals choose a discrete one-zero action. There is a distribution of "quality" across people, denoted θ, and individuals want to resemble high quality individuals. The value of the action is a function (denoted *W*(.,.))of the average quality of people consuming the action (denoted $\hat{q}$) and the quality of the individual (denoted q_i). This value function is a reduced form that is meant to capture the signaling value of the action. Assuming that *W*(.,.) is monotonic with respect to individual quality, equilibria will be defined with a marginal individual, denoted with $q*$, who is indifferent over taking the action, i.e. $W(q*,\hat{q}) = 0$.

We can discuss two possible equilibria. First, if ?W/?q_i is always greater than zero, then only individuals with quality greater than $q*$ will take this action. In this case, an increase in the number of individuals who are taking the action (i.e.. a reduction in $q*$) will lead to a reduction in the average quality level and an overall reduction of demand for the action. In this case, social interactions will lower variation in levels of the action over space. If ?*W*/?q_i is everywhere negative, then only individuals with quality less than $q*$ will take the action. An increase in the number of people taking the action will raise $q*$ and increase demand for the action. In this case, social interactions are positive. The implication is that social interactions should be particularly important in generating large variances across time and space for actions which stigmatize rather than elevate.[3]

A second type of stigma model involves a community norm of behavior where deviations are punished by the community (the rationality of this punishment strategy is generated by repeated game or Folk theorem like arguments). This community norm may serve to eliminate negative externalities from particular types of behavior. In this case, as more people participate in the action, fewer people become available to participate in the punishment and the costs of deviation become smaller. Again, a positive social interaction occurs because costs decline with the number of individuals taking the action.

A variety of literature has also argued for the possibility that interactions enter directly into the utility function. Bernheim (1995) argued for a taste for conformity where individuals experience a loss in utility just for deviating from the norms of the crowd. Akerlof (1997) examines a more general set of preferences where social choices enter into the utility function. Clearly, if the number of users of a commodity enters directly into one's taste for a commodity, then there will be social interactions.

Much of the more casual discussion of these taste based preferences often hinges upon people adopting the norms of behavior from others. For example, individuals think that certain types of behavior are "acceptable" because they see others also following these forms of behavior. By and large these stories can often be well captured with learning models where agents learn optimal behavior from their neighbors or with community-punishment types models. However, some observers tend to think that there is too much adherence to learned community norms to be justified by this type of model.

One alternative model assumes that individuals maximize a utility function which is the sum of utility from standard consumption and from one's living up to one's ideal self (this follows a long Freudian literature and is close to Akerlof and Kranton, 1997), or $Utility = U(X) + V(Z, \overline{Z})$, where X represents standard consumption variables and Z is a stock variable that captures one's identity (i.e. Z could include years of education or not being a liar or being thin).

There is a vast variety of things that individuals could care about being like (i.e. in principle anything could influence Z), so in practice parental and community norms must then determine which norms matter. In the utility function, this is accomplished with the $\overline{Z}$ term which is meant to capture the inputs from outside sources that determine which actions individuals should base their self-image upon (i.e. how important is it to be hard-working or clever or attractive or decent). Social interactions occur because through learning this ideal behavior, individuals influence each other. Natural examples of this type of effect occurs in crowd behavior where individuals seem to completely forego what is commonly thought of as civilized behavior because they are sanctioned by the crowd (see for example the extensive literature on the motivation of Nazis).

Of course, in any of these taste based theories we must try to understand what function these tastes would be playing in an evolutionary setup. Evolution should optimally just give individuals the actual evolutionary utility function (maximize DNA propagation) and enough intelligence to do this well. Clearly evolution isn't able to do this exactly and every set of combinations of tastes and computing ability is some solution to a second best problem. Interdependent preferences, if they exist, are surely solving some evolutionary problem. In particular, they may be acting to help get the optimal degree of social learning. However, without a better idea of the costs that stymie evolution just making people optimal social learners, we cannot tell why this particular form of utility would have evolved.

III. A Simple Model of Local Interactions

The following model description somewhat generalizes the model in the text of Glaeser, Sacerdote and Scheinkman (1996) in allowing for a richer action space on the part of individuals. Individual *i* now chooses an action *A(i)* from a subset of the line. Individuals are arranged on a one-dimensional lattice (a circle or line) and the choice of an individual's action is based entirely on his own taste for the action and his taste for imitating his predecessor on the line. More precisely, a fraction of individuals *(1-p)* receive sufficient utility from copying their predecessor that they will exactly imitate their predecessor's actions. The remaining individuals will choose their action independently; we will refer to these individuals as fixed agents. The mean action taken by these fixed agents is m_A and the variance is s_A^2.

If the probability of being a fixed agent is i.i.d. over the lattice, then in the equilibrium of this model two agents who are separated by K other agents, will either do exactly the same thing if there are no fixed agents between them (which occurs with probability $(1-p)^K$) or they will choose their actions independently if there is a fixed agent between them (which occurs with probability $1-(1-p)^K$). Thus, the covariance between two such agents equals $(1-p)^K$ times s_A^2. Using this fact, elementary algebra reveals that the sum of the city's actions, when divided by the square root of the city size, satisfies:

(1) $$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(A(i)-m_A\right)\xrightarrow{N\to\infty}N\left(0,\frac{2-p}{p}\,s_A^2\right).$$

This implies that the variance of normalized city averages will go to s_A^2 in the case where there are only fixed agents or go to infinity in the case where there are no fixed agents.

An Alternate Model

An alternative and equivalently simple model, which is somewhat more appealing in its assumptions about individual behavior, but is somewhat less appealing in its restrictions on the action space assumes that the actions space is the real line. In this case, we can assume that individuals' utility is a function of their own tastes, their actions and their predecessors actions:

(2) $$U\left(A_i, A_{i-1}, \Theta_i\right) = \Theta_i A_i - \frac{1-a}{2}A_i^2 - \frac{a}{2}\left(A_i - A_{i-1}\right)^2,$$

so that the marginal utility of the action for individual i is directly influenced by an idiosyncratic taste shock Θ_i, and by his neighbors' action, A_{i-1}. In order to incorporate observable individual characteristics into the formula, we define $\Theta_i = q_i + f(X_i)$, where q_i has mean 0 and variance s_q^2 (which is constant across cities), and X_i is the individual's set of observable characteristics which may include individual level characteristics (e.g. age and gender) and city-level characteristics (e.g. spending on welfare). We write $\overline{f(X)}$ for the mean level of the function *f(.)* and $\overline{f(X)}_j$ for the mean level of the function *f(.)* in city j.

In this case, the individual's action is defined by $A_i = q_i + f(X_i) + aA_{i-1}$, or equivalently:

$$(3) \qquad A_i - \overline{A}_j = q_i + f(X_i) - \overline{f(X)}_j + a\left(A_{i-1} - \overline{A}_j\right),$$

where $\overline{A}_j$ is the mean action level in city j. The variance of an individuals' actions can be found by noting that equation (3) and the fact that conditional on city j, $f(X_i)$ is independent of A_{i-1} (this uses our assumption that there is no sorting across neighborhoods within cities):

$$(4) \quad Var(A_i - \overline{A}_j) = s_q^2 + Var_j^{f(X)} + a^2 Var(A_{i-1} - \overline{A}_j) = \frac{s_q^2 + Var_j^{f(X)}}{1 - a^2},$$

since in equilibrium $Var(A_i - \overline{A}_j) = Var(A_{i-1} - \overline{A}_j)$, and where $Var_j^{f(X)}$ refers to the variance of *f(X)* within city j. As long as the *X* variables are independently distributed, then the correlation coefficient between individuals who are separated by *K* other individuals is now a^K. As *N* grows large, a version of equation (1) again holds:

$$(5) \qquad Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(A(i) - \frac{\overline{f(X)}_j}{1-a}\right)\right) \xrightarrow{N\to\infty} \frac{s_q^2 + Var_j^{f(X)}}{(1-a)^2}.$$

In general, we will assume that $Var_j^{f(X)}$ is constant across cities.

In order to determine the underlying parameters, if the econometrician observed the variance of $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(A(i) - \frac{\overline{f(X)}_j}{1-a}\right)$ (denoted Var_{agg}) and the population variance of A within cities (denoted Var_{ind} -- which is assumed to be constant across cities), then it is clear that in the limit $a = \frac{Var_{agg} - Var_{ind}}{Var_{agg} + Var_{ind}}$, and given our estimate of α and $Var_j^{f(X)}$ it is possible to estimate s_q^2.

This model requires more modification for discrete action spaces. One interpretation is to assume that there is a latent continuous variable that expresses only in measurable discrete units. For example, individuals may choose a continuous quantity of criminality that displays itself in a discrete value, whether or not the individual was arrested or individuals choose a continuous level of sexual behavior that displays itself in the number of out-of-wedlock births. For this model to be technically correct, it must be true that neighbors observe and make their decisions based on the actual continuous variable, not the discrete outcome.

Methodological Discussion

The fundamental empirical idea of this methodology is to use the relationship between the variance of community level aggregates and the variance of individual data to estimate the size of the social interactions. Our first step is to estimate the variance of action levels within cities. To do this, we allow for city-specific means and just estimate a common variance of the action around these city-specific means.

Next we assume the $f(X) = b'X$ and we estimate $\frac{b'}{1-a}$ by regressing sample average action outcomes on sample average city level characteristics (including state effects which should eliminate the effect of state level laws and regulations.):

(6) $$\hat{A}_j = \frac{b'\hat{X}_j}{1-a} + e_j$$

Using the predicted value from this regression, we obtain a value of $\frac{b'\hat{X}_j}{1-a}$. With these estimates, the predicted levels of outcomes across cities based on city level variables, we can estimate the variance of $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(A(i) - \frac{\overline{f(X)}_j}{1-a}\right)$. This aggregate variance and the individual variance are sufficient to estimate α.

If we are interested in differentiating between variance caused by observables ($Var_j^{f(X)}$) and variance caused by idiosyncratic tastes (s_q^2), we must then assume that there is no sorting within the city. Then we can estimate the regression using the 1990 Census Public Use Micro Sample to estimate a regression of the form $A_i = City\ \ Fixed\ \ Effect + b'X_i + e_i$. The city fixed effect will eliminate any bias that comes from differential sorting of individuals across cities.[4] With this regression, we have now estimated the coefficients β, on individual level characteristics. Given these coefficients, we form a value of $b'X$ for the all individuals and we can calculate the value of $Var_j^{f(X)}$, by calculating the variance of this predicted action level within cities.

Some Results

Table 4.1 presents our first set of results for female headed household rates. This variable represents the share of all families that are headed by a woman. This can be thought of as roughly the share of women who "choose" to have a family without a husband, conditional upon choosing to have a family (there are very few male headed households without a woman). All of our data comes from the 1990 census summary tape files. Our unit of observation is the metropolitan statistical area (MSA or when applicable the primary metropolitan statistical area).

Our goal is to estimate a -- the parameter which captures the degree of social interaction. This basic formula for this parameter is that $a = \frac{Var_{agg} - Var_{ind}}{Var_{agg} + Var_{ind}}$. where Var_{agg} and Var_{ind} are the aggregate and individual level variances described above.

The first row in Table 4.1 shows the individual and aggregate level variances when no observables are allowed as control variables. The individual level variance is the variance from a national mean. In this case we find an a value of .998, which is extremely close to one and quite far from zero.

The second row shows results where we have controlled for city level variables in calculating the city aggregate variance and calculated the individual level variances from a city-level mean. A wide battery of city level variables have been included which are described in the Table. While many of these variables may be endogenous, our goal is to control for as much as possible rather than to include only exogenous variables. Both the individual variance and the aggregate variance decline, but the aggregate variance declines by much more. The estimate of a thus falls to .995, which still represents quite sizable levels of social interaction. This level of a implies that the actions of individuals who are separated by 100 other individuals have a correlation coefficient of .606.

The third row gives results where we calculate the aggregate variance controlling for city level variables and state level fixed effects. These state level fixed effects should control for any omitted state level legal variables effecting this outcome variable. As expected the aggregate variance declines substantially, beyond our controls for city level variables. The individual level variance is still estimated around city level means and the overall value of a declines to .987, which implies that the correlation coefficient of actions of individuals who are separated by 100 other individuals is .27.

An Aside on Multi-Dimensional Interaction Models

Glaeser, Sacerdote and Scheinkman (1996) presents a variant on the voter model of physics (e.g Kindermann and Snell, 1980). Agents were located at points on a one-dimensional lattice and chose one of two possible actions. In our model, there were fixed agents who choose their actions at random. The other type, imitative agents, copied the action of one of their two neighbors with equal probability. Without fixed agents, the voter model in one (or two) dimensions produces unanimity in the long run and this unanimity is clearly inaccurate empirically for many variables. More precisely, suppose the agents are in Z_d, the set of points in R^d with integer coordinates, and that each agent chooses at time t, an action $a_i^t \in \{0,1\}$. At time zero, each agent chooses an action that is independent of other agents with $\Pr ob\{a_i^t = 1\} = p$. The neighbors of an agent $i \in Z_d$ are given by $N(i) = \{j \in Z_d : \max_{l=1,\ldots,d} |i^l - j^l| = 1\}$. For each $i \in Z_d$ there exists a Poisson process P_i with rate γ with P_i independent of P_j and such that at each epoch τ, the agent revises his action. Assume in addition that if agent i revises his action then $a_i^t = a_j^t$ with probability $\frac{1}{\# N(i)}$ for each $j \in N(i)$. That is, i copies the action of one of his neighbors at each epoch τ. If d=1,2 for any $i \in Z_d$, $j \in Z_d$ there exists an ε>0, there exists T such that if $t > T$, $\text{Prob}\{a_i^t = a_j^t\} > 1 - e$. In other words, in one or two dimensions, agents behavior will eventually be unanimous.

However, if d=3, unanimity no longer holds and there exists a stationary measure μ(p) and if $S_n = \sum_{|i|\le n} \frac{a_i - p}{\left((2n+1)^d\right)^{1/2+1/d}}$ then $S_n \to N[0, s^2]$. This formula suggests that empirically, one could in principle estimate the number of dimensions that explain the observed variance of group-level average actions. The larger the dimension of the interactions, the lower is the exponent *(1/2+1/d)* that must be used to normalize to get a normal distribution. In the limit as d grows, the exponent approaches *1/2*. Thus, the higher dimension lattices increasingly resemble the case where decisions are independent. Intuitively, sufficiently large amounts of interaction eliminate the tendency of interaction to produce all-or-nothing outcomes. Another way of generating scaling rules other than scaling by *1/2* involves models with long spatial dependency (Glaeser and Scheinkman, 1997).

III. Combining Local and Global Interactions

In this case, we assume that utility depends both upon the actions of a neighbor and of the community as a whole. The community average can either increase or decrease the incentives to engage in the particular level of behavior. A globally high level of crime may mean that many voters are criminals who do not want to spend on police expenditures. Alternatively, a high community-wide level of crime may reduce the incentives to engage in crime. Also perhaps, as more people are criminals, there may be fewer potential victims, so the returns to crime in the community may fall (again this is only a global interaction if criminals choose their victims from a global rather than a local pool).

To formally treat global interactions, we again assume that action levels are continuous and that individuals choose their actions treating the global levels as exogenous to maximize:

(7)

$$U\left(A_i, A_{i-1}, \frac{\sum_{\ell \neq i} A_\ell}{n}, \Theta_i\right) = \Theta_i A_i - \frac{1-a}{2} A_i^2 - \frac{a\,j}{2}\left(A_i - A_{i-1}\right)^2 - \frac{a\,(1-j\,)}{2}\left(A_i - g\left(\frac{\sum_{\ell \neq i} A_\ell}{n}\right)\right)^2$$

which implies:

$$A_i = \Theta_i + a\,jA_{i-1} + a\,(1-j\,)g\left(\frac{\sum_{\ell \neq i} A_\ell}{n}\right) \tag{8}$$

Both a and j are strictly less than one and greater than zero (when $j = 1$ this is the pure local interactions model discussed above). As before to make the system symmetric $A_0 = A_n$. We will also treat two separate assumptions about the taste shocks. First, we assume that $\Theta_i = q_i$ where q_i is i.i.d., with mean that we normalize to zero and variance s_q^2. Second, we assume that $\Theta_i = q_i + f(X_i)$, where q_i is again i.i.d. with mean zero and variance s_q^2, and X refers to observable characteristics of the individual. In both cases, the variable Θ_i is assumed to have a bounded support. The function *g(.)* is bounded and continuously differentiable with a bounded derivative.

For any given sequence $\{\Theta_i\}_{i=1}^{n}$, we can define a function $F: R^n \to R^n$ that maps the vector $(A_0,\ldots,A_{n-1})$ by:

(9)

$$F(A_0,.., A_{n-1}) = \left(\Theta_1 + a\, jA_0 + a\,(1-j\,)g\left(\frac{\sum_{\ell \neq 1} A_\ell}{n}\right),..,\Theta_n + a\, jA_{n-1} + a\,(1-j\,)g\left(\frac{\sum_{\ell \neq n} A_\ell}{n}\right)\right)$$

.

Here $A_0 = A_n$. As $a < 1$, φ<1 and *g(.)* is bounded, the function will have at least one fixed point which will solve equation (5). In general, however, there is no guarantee that this fixed point is unique. It is entirely possible that there exist multiple solutions to equation (8). Further, the optimal action of agent i depends on the total population size n and we denote this dependence by writing A_i^n for the action taken by individual *i*.

Summing equation (8), and writing $\hat{A}_n = \frac{\sum_{i=1}^{n} A_i^n}{n}$ we find:

(10) $$(1 - a\, f)\hat{A}_n - \frac{a\,(1-\, f)}{n}\sum_{i=1}^{n}[g(\frac{\sum_{k \neq i} A_k^n}{n-1})] = \frac{\sum_{i=1}^{n} \Theta_i}{n}$$

Further iteration of equation (8) yields:

(11) $$A_i^n = a^{\,i} f^{\,i} A_0 + \sum_{\ell=1}^{i} a^{\,i-\ell} f^{\,i-\ell}[\Theta_j + a\,(1-f\,)g(\frac{\sum_{k \neq \ell} A_k^n}{n})]$$

At this point, we will separate our discussion into two sections, based on our two assumptions about Θ.

Case 1: $\Theta_i = q_i$

We will assume that there are a finite number of solutions to the equation $g(x) = \frac{1 - a j}{a(1 - j)} x$ and that at each such solution $g'(x) \neq \frac{1 - a j}{a(1 - j)}$. This is a “generic” assumption.

In the appendix, we show that the sequence $\hat{A}_n$, of average actions in a population of size *n*, converges, as $n \to \infty$, to a solution of the equation:

(12) $$g(\overline{A}) = \frac{1 - a f}{a(1 - f)} \overline{A}.$$

We denote $a_i^n = A_i^n - \overline{A}$, and we also show in the appendix that:

(13) $$\sum_{i=1}^{n} \frac{a_i^n}{\sqrt{n}} \to N\left(0, \frac{s_q^2}{\left(1 - a j - a(1 - j) g'(\overline{A})\right)^2}\right).$$

Within city variance (using equation (8)) is $\frac{s_q^2}{1 - a^2 j^2}$. It is somewhat meaningless to try and determine between the effects of a, j, and $g'(\overline{A})$, but even if we attempt to distinguish between s_q^2, $a j$, $a(1 - j) g'(\overline{A})$, it is impossible without more information. Essentially we have three variables and only two equations.

If we had an additional variable, for example the covariance of actions of individual *i* and *i*-1 within a city, then we could back out these three variables. Within a given city, $\operatorname{cov}(A_i, A_{i-1}) = a j Var(A_{i-1}) = \frac{a j s_q^2}{1 - a^2 j^2}$. This covariance can be either found directly (if micro-data is available) or found by city-subaggregates (i.e. neighborhood level averages). In this case, $a j$ equals the correlation coefficient of two neighbor's actions, and by using the variance, s_q^2 can be found. With these two parameters, it is possible to determine the size of the global interaction by looking at variances across cities.

Alternatively, one could identify the model by examining the variance of neighborhood level averages within a single city. If a neighborhood has size *h*, then conditioning on the city level mean, $Var\left(\frac{\sum_{j=i}^{i+h} A_j}{\sqrt{h}}\right) = \frac{s_q^2}{(1-a j)^2}$. The only difference between this expression and equation (13) is that all neighborhoods within a city are affected by the same global interaction term, so there are no terms involving *g'(A)*.

If it is desirable to control for observables and still use this simpler framework, a simple assumption is that observed action *Y=A+f(X)*, where *Y* is the observed action and *f(X)* is a function of observables. Thus A can be inferred by subtracting *f(X)* from Y, if *f(X)* is known (and given our assumption, there is no reason why it cannot be estimated from either micro-level or aggregate-level regressions). All of the statements about A are unchanged with this assumption. Empirically, it is necessary to work with *Y-f(X)*, the residuals from a first stage regression. This framework allows for a simple manner of controlling for observables. However, it is not satisfying in that we are assuming that one's influence on one's neighbors is only a function of unobservable factors. The next section introduces a more complicated setup, where we allow observables to influence neighbors.

Case 2: $\Theta_i = q_i + f(X_i)$

In this case, we assume that taste shocks contain both an individual specific, i.i.d. component, and also a component that is based on an individual's observable characteristics and a component based on city level characteristics (which may include both individual-specific and city-specific attributes). We assume that Θ_i has a second moment. Using a similar logic to the one used in Case 1, and making an analogous assumption concerning the finiteness of the set of solutions to the equation $(1-a\,j\,)x - a\,(1-j\,)g(x) = \overline{f\,(X\,)}_j$ one can show that, in the limit, the mean level of the action in city j must satisfy:

(14) $$(1-a\,j\,)\overline{A}_j - a\,(1-j\,)g(\overline{A}_j) = \overline{f\,(X)}_j$$

Equation (14) typically has many solutions for each value of $\overline{f\,(X)}_j$. From now on, we condition on a "branch" of the solution and note that except for a finite set of values of $\overline{f\,(X)}_j$, the solutions will vary smoothly with $\overline{f\,(X)}_j$ within each branch. Importantly, for each value of $\overline{A}_j$ there exists at most one $\overline{f\,(X)}_j$ that solves the equation. Thus, in principle one can estimate $\overline{f\,(X)}_j$ as a function of $\overline{A}_j$.

Using the implicit function theorem and differentiating (14) implies:

(15) $$\frac{\P\,\overline{f(X)}_j}{\P\overline{A}_j} = (1-a\,j\,) - a(1-j\,)g'(\overline{A}_j)$$

Thus the derivative of predicted value with respect to outcome level will yield an estimate of $(1-a\,j\,) - a(1-j\,)g'(\overline{A})$. The connection between realized outcome, and predicted outcome based on micro-level variation, gives us an estimate of the extent to which there are spillovers.

If we assume that the distribution of *X* is constant across neighborhoods such that the average level of *f(X)* in each neighborhood within city j is $\overline{f(X)}_j$. Then

$$\text{(16)} \qquad Var\left(\sum_{i=1}^{h}\frac{A_i-\overline{A}_j}{\sqrt{h}}\right)\to\frac{s_q^2+Var_j^{f(x)}}{(1-a\,j)^2},$$

where h again indices the members of the neighborhood. Furthermore, the variance of action levels within the city will again equal $\frac{s_q^2+Var_j^{f(x)}}{1-(a\,j)^2}$. The variance of neighborhoods, and the variance of individual level actions within cities allows us to identify $a\,j$. Thus, as we learn $(1-a\,j)-a(1-j)g'\left(\overline{A}_j\right)$ from the aggregate regressions we are able to separate the extent to which spillovers come from local and global sources, after we condition in an equilibrium.

Unfortunately this approach requires us to assume that there is no sorting by observables across neighborhoods. If we actually were able to run individual level regressions within cities with neighborhood level fixed effects, we could then drop this assumption. Then we could note that $(1-a\,j)\left(\overline{A}_h-\overline{A}_j\right)=\overline{f(X)}_h-\overline{f(X)}_j$, where quantities with the h subscript indicate neighborhood level outcomes. The relationship between predicted outcomes and actual outcomes then provide a separate estimate of $(1-a\,j)$.

A Discrete Version

Since many of our variables are discrete, it makes sense to consider an analogous model where only two actions {0, 1} are possible. For simplicity we only describe the model in the case without observables. In this case, assume that a city has n agents on a circle. With probability α, agent *i* bases his actions exclusively on the actions of agent *i*-1, and we again identify agent 0 with agent n. With probability 1-α, agent *i* bases his action on the global average. In this case, the probability that agent *i* chooses action 1 is given by $g\left(\frac{\Sigma_{j\neq i} A_j}{n-1}\right)$ where *g(.)* is a continuously differentiable function defined for $x \in [0,1]$ with 1>*g(*x)>0. Following a similar reasoning to that of the previous model, we may conclude that the average action in a city must converge as $n \to \infty$ to some solution of the equation $\bar{A} = g(\bar{A})$. Furthermore the variance of normalized city-level averages satisfies $Var\left(\frac{\sum_{j=i}^{i+n} a_j}{\sqrt{n}}\right) = \frac{1+a}{1-a} \times \frac{\bar{A}(1-\bar{A})}{1-g'(\bar{A})}$. The variance of normalized neighborhood-level averages $Var\left(\frac{\sum_{j=i}^{i+h} a_j}{\sqrt{h}}\right) = \frac{1+a}{1-a}\bar{A}(1-\bar{A})$.

Of course the variance of any one individuals action is $\bar{A}(1-\bar{A})$. These three equations allow us to empirically identify the model. We will estimate the parameters for discrete variables as if they were continuous variables in the next section, but more properly discrete variables need to be treated differently using this particular formulation.

Empirical Implementation

This section employs two distinct methodologies. The first methodology assumes that observable variables do not create spillovers and can just be controlled for and then ignored. The second methodology assumes that observable variables create their own spillovers. For both methodologies, we can estimate the value of $(1-a\,j)$ by using the micro-level variance and the variance of neighborhood averages. If we assume that the observables can just be subtracted, we begin by regressing outcomes on observables (in a micro-level regression) and then using those coefficients to subtract the effect of observables from any aggregate.

Methodology 1-- Social Influence comes only from Unpredictable Elements of Decisions

The first methodology relies upon the assumption that we can ignore the effect of observables on social interactions. In this case, we first regress our micro-outcome variable (does the family have a single head) on a battery of family level characteristics including city level fixed effects. This regression furnishes us with estimates of the effect of observable characteristics on the outcome variable and using these estimates we correct tract-level and city-level outcome variables for observable characteristics. As discussed earlier, because of sorting across cities, observable characteristics may be correlated with the action of one's neighbors and as such either city level regressions or micro-level regressions that do not control for city level fixed effects may well be biased. Of course, we are, unrealistically, forced to assume that there is no sorting within cities.

Thus, all further procedures within this methodology are done using corrected female headed household rates where the effect of observable characteristics have been eliminated (except for the row marked "raw female headship rate in families" and Table 4.3a). Using the corrected city level female headship rates, we then determine how much of the variance across cities can be plausibly explained by the existence of multiple equilibria (of normalized city averages) and how much can be determined by the variance of cities within each of these equilibria. While the variance created by the multiple equilibria is not the only variance due to global interactions, in our model it represents one component of the global interactions.

Of course, this result is due to the assumption that the global interaction may be non-linear while the local interaction is linear. If the global interaction was linear, then it could not generate multiple equilibria. If the local interaction was non-linear, then even in the absence of global interactions, multiple equilibria could still exist. Perhaps it is therefore wiser to interpret the amount of variance created by multiple equilibria as the variance associated with non-linearities in the interaction process rather than as the outcome of global interactions.

We allow for the presence of multiple equilibria by using the EM algorithm to fit a mixture of normal distributions to the observed distribution of corrected city level headship rates. Tables 4.3a, 4.3b and 4.3c show the results of estimating multiple equilibria via the EM algorithm for female headed household rates. This algorithm allows us to estimate that each city is drawn from up to five distributions with different means and standard deviations. Since the data is always fit better by more distributions, a loss function must be specified so that we allow more distributions only if a sufficiently large amount of explanatory power is generated by allowing for an extra distribution. We implement this loss function using the Akaike Information Criterion which allows us to compare across numbers of distributions to determine which one gives us the most explanatory power relative to its criterion.

In Table 4.3a, we estimate the number of distributions for percent female headed household without any additional controls. In this case, the Akaike Information Criterion is minimized with three component distributions. The first distributions, which contains 55 percent of the cities, has a low mean and a variance roughly comparable to the aggregate variance. The second distribution has 42 of the distribution and a much lower variance, which leads to a correspondingly lower value of social interactions for this group. The third distribution has only 3 percent of the cities, but it also has an extremely high mean and variance.

In Table 4.3b we again control for city level variables in estimating the aggregate variance term. The EM algorithm is used on the distribution of female headed household rates across cities after we have first orthogonalized these rates to a battery of city level characteristics. In this case, the three component distribution again minimizes the Akaike Information criterion. In Table 4.3c, we orthogonalize city level female headed household rates with respect to city level variables and with respect to state level fixed effects. The average aggregate variance is reduced much less by allowing for the presence of multiple equilibria. In this case the Akaike Information Criterion is minimized with two component distributions. The first distribution has 78 percent of the cities and the second distribution has 22 percent of the cities.

Once we have estimated the number of distributions that best fit the data, we use the variance of the city level aggregates around each distribution to estimate the degree of local and global interactions. Notice that the reduction in variance created by allowing the presence of multiple equilibria is already one sign that global equilibria matter. Table 4.2 shows the results from this procedure for female headship rates. We use the average city-level variance rather than the variance for each one of the component distributions in order to produce a single set of results. We use the three formulas $Var_{tract} = \frac{s_q^2}{(1 - aj)^2}$, $Var_{ind} = \frac{s_q^2}{1 - a^2 j^2}$, and

$Var_{city} = \frac{s_q^2}{\left(1 - a j - a(1 - j)g'\left(\overline{A}\right)\right)^2}$ to estimate the key parameters of the model: $a j$ (which captures the importance of local interactions) and $a(1 - j)g'\left(\overline{A}\right)$ (which captures the importance of global interactions, after we condition in an equilibrium).

Our findings, shown in Table 4.2, are that allowing for multiple equilibria and global interactions substantially reduces the importance of local interactions in female headship rates. The value of $a j$ is comparable to the value of a in Table 4.1, with only local interactions, and it is clear that including global interactions has lessened the importance of social interactions. However, after conditioning in an equilibrium, the importance of local interactions is much higher than the importance of global interactions. We find that an increase in your neighbor's action is more than twelve times more important than an increase in the city level average (which is found by comparing $a j$ -- the effect of the neighbor-- with $a(1 - j)g'\left(\overline{A}\right)$ the effect of the city level average). Of course, we are really differentiating between city-level and tract-level interactions, are we are referring to tract-level interactions as local interactions. It might be that tract-level interactions are not actually local as described by our model, i.e. neighbor-to-neighbor. Instead, the tract-level interactions that we identify as local might occur equally across the tract, i.e. individuals are influenced by the average level of behavior in their tract. Table 4.3 exhibits parameter estimates for global vs. local interaction models for several other variables.

Methodology 2-- Allowing Control Variables to Influence Interactions

The previous methodology assumes that only the unpredictable component of actions creates social interactions. In this section, we rely upon the fact that the component of individuals' actions that are attributable to observable characteristics will have exactly the same social interaction effect as the components of individuals' actions that are not attributable to any observable characteristics. If observable characteristics influence neighbors, but not as strongly as unobservable characteristics then neither one of these procedures is correct and some mixture of the two procedures is best.

Table 4.4 presents an estimation of the strength of global interactions for female headship rates. To estimate $a j$ we use exactly the same procedure as we did above. First we regress micro-outcomes on observables and then we use the parameter estimates from this regression to correct for observable characteristics. We then estimate the individual level variance and the tract level variance and using $Var_{tract} = \frac{s_q^2}{(1 - a j)^2}$, and $Var_{ind} = \frac{s_q^2}{1 - a^2 j^2}$ we form an estimate of $a j$.

Our estimate of global interactions is more difficult. As discussed above, we use the parameter estimates from the micro regression to create a predicted outcome level for each city based on the city level observables and the parameter estimates from the micro level regression. We then regress this city-wide predicted action level on the action. If there were substantial non-monotonicities in this function, then multiple equilibria would be a possibility. As it is, the function is completely monotonic and thus this procedure does not confirm the existence of multiple equilibria at the city level.

Intuitively, the result of regressing predicted outcomes on actual outcomes can be best thought by considering the null hypothesis of no social interactions. In that case, we would expect the predicted outcome level to move one for one with the actual outcome level (on average). However, as we see we find that the predicted outcome level moves less than one-for-one, which means that large changes in the actual outcome level are associated with smaller changes in the predicted outcome level. This finding is quite supportive of the existence of social interactions.

However, our formula is that $\frac{\partial \overline{f(X)}}{\partial \bar{A}} = 1 - \alpha\phi - \alpha(1-\phi)g'(\bar{A})$. Our point estimates of $\frac{\partial \overline{f(X)}}{\partial \bar{A}}$ range from .386 to .792, but our estimate of $\alpha\phi$ (from the micro-level and tract-level variance) is .927. As such, the global interaction terms must be negative (although in one case, the global interaction term is not statistically different from zero). There are two possible ways of interpreting this result. First, it is possible that the global interaction terms are negative.

Second, and we think more realistically, there are problems associated with the fact that we are cobbling together two different procedures to estimate the global and local interaction. In principle it would be possible to estimate $\alpha\phi$ by regressing average tract level predicted outcome on average tract level outcome. If there is any tendency of the unobservable causes of actions to create more social interaction than the observable components then our current estimate of $\alpha\phi$ will be much higher than it would be using this alternative method. Also, if unobservable causes of actions matter more than our using of two procedures will lead to many more estimation problems than using a common procedure to estimate both components of local and global interactions.

This section has been highly exploratory and we hope that future work will extend this approach. However, we have argued that there are two distinct methodologies for estimating local and global interactions. First, it is possible to use aggregate variances and compare these variances with micro-level variances. Second, it is possible to use the connection between average predicted level outcome and average outcome. While this second point is similar to micro-level estimation techniques (see, e.g. Case and Katz, 1991) which look for social interactions by using neighbors characteristics as instruments, our procedure is essentially novel in many ways and needs more development before it can be counted on to produce reliable results.

IV. Selection into Locations

We now revert back to our simpler local interactions model and assume

(2') $$U(A_i, A_{i-1}, q_i, P_z) = q_i A_i - \frac{1-a}{2} A_i^2 - \frac{a}{2}(A_i - A_{i-1})^2 - P_z$$

where P_z represents the cost of living in city z. As before, individuals choose their actions so that $A_i = q_i + aA_{i-1}$. In this model, however, individuals choose their city as well as their action. Furthermore, they will choose their city before observing who their neighbor will be or even exactly what their value of q_i will be. One justification for this is that individual tastes will change over time (so individuals are unsure as to what their tastes will be). We further assume that no one knows who his neighbor will be. However, individuals will have a guess as to what sort of people are selecting into the city.

To implement the idea that individuals have some imperfect knowledge about their own tastes, we assume $q_i = \hat{q}_i + e_i$, where $\hat{q}_i$ and e_i are both mean zero, i.i.d. random variables with variances $s_{\hat{q}}^2$ and s_e^2, and suppose γ solves $(1-g)s_{\hat{q}}^2 = gs_e^2$. The term $\hat{q}_i$ represents tastes that are known ex ante and may include the effect of observable individual characteristics. The term e_i represents individual tastes that are only known after migration which are assumed to be independent of all other individuals' taste shocks. There are a fixed number of C cities labeled 1 to C and all of these cities are ex ante identical (this represents a simplification over the previous models). Furthermore, while it is possible to endogenize the size of communities (as long as cost of living rises sufficiently quickly with population size there will always be interior solutions for city size), we will assume that the fractions of the population $(k_1,...,k_C)$ are exogenous.

Using the fact that individual i knows that he will choose his actions optimally so that $A_i = q_i + aA_{i-1}$, the expected utility of individual i who chooses city after observing $\hat{q}_i$ will be:

(17) $$E(U_i|z) = \frac{\hat{q}_i^2 + s_e^2}{2} + a\hat{q}_i E(A_{i-1}|z) - \frac{(1-a)a}{2} E(A_{i-1}^2|z) - P_z$$

The key point in this equation is that there is a strategic complementarity between an individual's expected proclivity towards the action ($\hat{q}_i$) and the tendency in the city to follow the action. For example if agent i weakly prefers city z to city z', where $E(A_{i-1}|z) > E(A_{i-1}|z')$ then any agent i' for whom $\hat{q}_{i'} > \hat{q}_i$ will prefer z strictly to z'. This fact implies that there will be strict sorting of individuals across cities, unless the cities are exactly identical (as in Benabou, 1993). Without getting into more detailed dynamic issues, usual ad hoc notions of stability assure us that these symmetric equilibria will be unstable. The intuition of this is that if one of two initially symmetric cities becomes slightly higher in expected action levels, then all of the individuals with higher $\hat{q}_i$s will tend towards that city. Because of this instability of the symmetric equilibria cases, we will focus exclusively on the stable complete sorting equilibria.

The determination of which cities will be high action and which will be low action is not determined by the model. We will normalize and order the cities so that expected action levels rise monotonically with *z*. We let l_z denote the highest skill level in city *z*. Naturally, $l_z < l_{z+1}$ and in equilibrium if $\hat{q}_i \in (l_{z-1}, l_z)$ then individual i lives in city *z*. For simplicity we will assume that the distribution $\hat{q}_i$ has a density function *f(.)*, so we will not worry about individuals at the boundary. Since individuals at the boundary need to be indifferent between the two cities, if mass points are present we may assign a convenient fraction of the agents of that type to each city.

Formally, an equilibrium is a vector $Inf(\hat{q}) = l_0, l_1, l_2, \ldots, l_{C-1}, l_C = Sup(\hat{q})$ and a vector $p_1, p_2, \ldots, p_{C-1}, p_C$ such that (1) if $\hat{q}_i \in [l_{z-1}, l_z]$ then $E(U_i|z) \geq E(U_i|z')$ for each *z*'=1, ... *C*, and (2) $Probability(\hat{q}_i \in [l_{z-1}, l_z]) = k_z$ for each *z*=1,..., *C*. The existence of an equilibrium is easy to show. The model does not pin down the level of property values, only the difference of housing costs across cities, so we will normalize $p_1 = 0$.

We denote $\bar{q}_z = E(\hat{q}_i|z) = E(q_i|z)$ and $\bar{A}_z = E(A_i|z)$. Using the first order condition $A_i = q_i + aA_{i-1}$, since $E(A_i|z) = E(A_{i-1}|z)$, it follows that $\bar{q}_z = (1-a)\bar{A}_z$. We further let A_i^z (q_i^z) denote the action (taste shock) associated with individual *i* in city *z* and note that city *z* has $n_z = k_z T$ inhabitants, where *T* denotes the total population of the country. Since $\hat{q}_i$ has a finite variance, it follows that:

(18)
$$\sum_{i=1}^{n_j} \frac{A_i^z - \bar{A}_z}{\sqrt{n_z}} - \frac{1}{1-a} \sum_{i=1}^{n_z} \frac{q_i^z - \bar{q}_z}{\sqrt{n_z}} \to N\left(0, \frac{s_z^2}{(1-a)^2}\right).$$

where s_z^2 is the overall variance of q_i^z. We know that $s_z^2 = \hat{s}_z^2 + s_e^2$, where $\hat{s}_z^2$ is the variance of the signal q_i^z in city *z*, which must satisfy

$$\hat{s}_z^2 = \int_{l_{z-1}}^{l_z} \left(\hat{q} - \bar{q}_z\right)^2 \frac{f\left(\hat{q}\right)}{k_z} d\hat{q} = \int_{l_{z-1}}^{l_z} \hat{q}^2 \frac{f\left(\hat{q}\right)}{k_z} d\hat{q} - \bar{q}_z^2.$$

We now let $w_z = \frac{1}{1-a} \sum_{i=1}^{n_z} \frac{q_i^z}{\sqrt{n_z}} = \frac{1}{1-a} \sum_{i=1}^{n_z} \frac{q_i^z - \bar{q}_z}{\sqrt{n_z}} + \frac{\bar{q}_z \sqrt{n_z}}{1-a}$. This term reflects the weighted average of taste shocks in city z. By analogy to equation (18), the variance of w_z is equal asymptotically to the variance of the normalized sum of the deviations of actions of individuals in city z from the average action of the population. We assume that the ratio of any one's city population to the average city population (denoted $\hat{n}_z$) is bounded from above and below by two constants K_1 and K_2, so that $K_1 < \frac{n_z}{\hat{n}_z} < K_2$.

The variance of the random variable $\sqrt{n_z}\, w_z = \sqrt{k_z T} w_z$. the weighted variance of the w_z terms, equals:

(19) $$CVar\left(\sqrt{k_z} w_z\right) = C \frac{TVar\left(k_z \bar{q}_z\right)}{(1-a)^2} + CE\left(\frac{k_z s_z^2}{(1-a)^2} \right).$$

In this equation, the moments can be computed against the measure that attributes probability 1/*C* to each of the *C* cities.

The second term of the right hand side of equation (19), equals $\frac{s_e^2 + E\left(\hat{q}^2\right) - \sum_{z=1}^{C} k_z \bar{q}_z^2}{(1-a)^2}$. It then follows that:

(20) $$Var\left(C\sqrt{k_z} w_z\right) = \frac{1}{(1-a)^2} \left[s_e^2 + E\left(\hat{q}^2\right) + \sum_{z=1}^{C} (n_z - 1) k_z \bar{q}_z^2 \right].$$

Intuitively, this equation makes it clear that the variance of weighted city level averages is determined by the overall level of interaction, the variance of taste shocks, and the amount of sorting across cities.

The first two terms in brackets are generally invariant with respect to increases in C. The third term is generally of order *T/C* (which is average city size). If the support of θ is bounded, then the last term is at most of order *T/C*. Further, as $K_1 < \frac{n_z}{\hat{n}_z} < K_2$ implies that $K_1 / N < k_z$ we can place a lower bound on $\bar{q}_z$ provided that $j = N/4$ or $j = 3N/4$. Hence the last term in brackets is at most of order *T/C* .

Empirical Approaches

While in principle there could be many different approaches to estimating the share of the cross-city variance that comes from sorting and the share that comes from local interaction, we focus on the last implication of the model. The taste and local interaction related variance terms are not of order *T/C* (i.e. they do not change with average city size), while the sorting source of variance is of order *T/C* . In principle, then if we compared across sets of locations, where there is no migration between each set of cities but there is migration within each set of cities, where the average city size differs then we could determine the extent to which sorting determines the variance across cities. For example, if we found that the variance was much higher in areas with larger city sizes then we would attribute much of the variance to sorting and less to local interaction.

While in practice it is impossible to perform this sort of exercise perfectly, we will present a crude facsimile using U.S. states. Of course, for this estimation to be perfectly correct we would need to assume that all sorting occurs within states not across states. However, the estimation procedure would still be basically unbiased if the means of the underlying taste distribution differed across states but the variance of the taste distribution stayed constant or at least did not change in a way that was systematically related to average city size.

We estimate the variance of weighted outcome variables $Var\left(C\sqrt{k_z}w_z\right)$ within each state, using the relationship between this variance and the variance of the normalized sum of the deviations of actions of individuals in city z from the average action of the population, and then regress this variance on the average city size within the state. The amount that remains in the intercept can be interpreted as the amount of variance that can properly be attributed to social interaction. Table 4.5 shows our estimates for three variables. As a test case, in the third row of Table 4.5 we used percent non-white, which should reveal variance only due to sorting. The intercept in this case was negative and statistically insignificant, which means that the methodology is not inappropriately identifying social interaction in this case.

In the case of female-headed households and the crime rate, we find positive levels of social interaction and significant levels of sorting. In both cases, the sorting effect is significantly positive. In the case of the crime rate, the social interaction effect is significantly positive. In the case of female headed households, the social interaction effect is not statistically significant, but it is economically sizable. As a result, we must conclude that this procedure shows promise but is far from precise.

Of course, it is worthwhile stressing that sorting itself only occurs because of social interactions. Similar individuals would not choose to locate near each other if there weren't social interactions. Hence, our results should be seen as estimating the extent to which social interactions operate through sorting or through interaction after sorting.

V. Dynamic Models

Many of the sorting problems just discussed disappear when considering time series variation. While we lose sources of variation, we do eliminate some of the hardest problems of estimating social interactions. In this case, we consider a simple class of models in which all individuals start in a particular state and then may choose to switch to another state (or action). The switch is assumed to be irreversible, which admittedly it will not be in many cases. We have in mind choosing a particular technology or moving to a new country or perhaps women entering the labor market (although in this case, the decision is clearly reversible).

Descriptively, we will focus on the last example, despite the reversibility issue. There are many reasons to suspect that there are substantial social interactions involved in women entering the labor market. As more women entered they lowered the stigma of work, reduced the discrimination against women in the workplace and eliminated the social network that facilitated not working in the formal labor market.

We model a single location with population n. Each agent is indexed by an integer i=1, 2, 3, ..., n. At time t an individual i is in one of two states. The state s_t^i=0, if the agent has not entered the labor market, otherwise s_t^i=1. We will assume that entering the labor market is irreversible i.e. if s_t^i=1, and $t'>t$, then $s_{t'}^i=1$.

Agent *i*'s flow of utility per period depends on his type $\tau \in \{0, 1, 2\}$, and on his own state. In addition if s_t^i=1 the utility also depends on the states of agents in a set *N(i)* of "friends" of *i*, at the last time t_1 such that $s_{t_1}^i$ =0. In order to simplify the forecasting problem of agents, we assumed that although new workers benefit from the presence of older entrants into the labor market that are their friends, the reverse is not true.[5] We will begin by assuming that *N(i)={i-1, i+1}*. For symmetry we identify 0 with *n* and *n*+1 with 1, i.e. we set *N*(1)=*{n*, 2*}* and, *N(n)={n*-1,1*}*. At each time t if s_t^i=0, we assume that agent *i* will be given, with probability p a choice to enter the labor market. She then must compare the value of staying with the value of working that is a function of her type and the states of her friends. Let $u_{t+1} = s_t^{i-1} + s_t^{i+1}$ and $V(t,u)$ denote the value of entering the labor marke, as a function of the type τ and the value υ of the sum of the states in *N(i)* as of the preceding period. Since it is only the value υ, in the period before entering the labor market that matters, if δ is the discount factor per period then $V(t,u) = \frac{d}{1-d} h(t,u)$, where here $h(t,u)$ is the per-period utility of an individual that works, as a function of her type τ, and the value υ of the sum of the states in *N(i),* in the period before she worked. Similarly, let $U(t,u)$ denote the value of staying, as a function of the type τ and the value υ of the sum of the states in *N(i),* in the preceding period. Elementary dynamic programming implies that:

$$(21) U(t,u_{t-1}) = g(t) + d[(1-p)E(U(t,u_t) + pE(\max\{U(t,u_t),V(t,u_t)\})],$$

where *g(*τ) is the per-period utility of an individual that stays outside the labor market as a function of his type τ and E denotes the expected value over the value of u_t conditional on u_{t-1}. Individuals work whenever $U(t,u_t) < V(t,u_t)$. Since an agent has a choice of never working, we known that $U(t,u_t) \geq \frac{1}{1-d} g(t)$.

We make three assumptions:(A1) $h(0,u) < g(0)$, for any 0=υ=2,(A2) *h(*1,0)=*h(*1,1)=*h(*1,2) >*g(*1), and (A3) *h(*2,0)<*g(*2)<*h(*2,1)=*h(*2,2). Assumption A1 states that agents of type 0 get a higher flow of utility by not working. Assumption A2 states that agents of type 1, do not care about the number of friends that work, but always get a higher utility by working. Assumption A3 states the additional benefit to a type two agents from the previous working of a second friend. In addition, the utility of a type two agent not working is larger than that of working by herself, but lower than that of working, if at least one fried works. These assumptions makes the solution of Problem 1.1 quite simple. Type 0 will never work. Type 1 will always work when she is given a chance. Type 2 will not work if none of her friends work. Obviously we can weaken the qualities assumed in the assumptions, and still retain the solution to problem 1.1 by combining hypotheses about the function *h* and *g* with hypotheses about the discount factor δ and the probability *p*. Each agent is type τ with probability q_τ, independent of the type of all other agents.

Given any *i*, let *i*- be the largest integer less than *i* such that τ_{i-}?2, and let i+ be the smallest integer greater than *i* such that τ_{i+}?2. Here, again we identify *n* as the predecessor of 1 and 1 as the successor of *n*. As $t \to \infty$, $s_t^i \to 1$ unless $\tau_{i-}=\tau_{i+}=0$ in which case $s_t^i = 0$ for all *t*. We write s^i for the asymptotic value of s_t^i. The asymptotic distribution of states can be derived in a manner similar to the derivation of the steady state distribution in the models of Glaeser, Sacerdote and Scheinkman (1996). In particular, the expected fraction of workers converges to

$m \equiv 1 - q_0 - (1 - q_0 - q_1)\dfrac{q_0^2}{(q_0 + q_1)^2}$, where q_0 and q_1 are the fractions of type zero and type one respectively. Furthermore, under the asymptotic distribution, if m_n denotes the fraction of workers in population of size n the $(m_n - m)\sqrt{n} \to N(0, s^2)$.

To establish this central limit behavior it is enough to observe that if j>j' and $A_{j,j'}$ is the event that at least for two values of j>i>j', $\tau_i \in \{0,1\}$ then conditional on

$A_{j,j'}$, s^j is independent of $s^{j'}$. Since the probability of the complement of $A_{j,j'}$ goes to zero exponentially as j-j' does to infinity, we know that m_n displays central limit behavior. Furthermore the variance of the limit random variable s^2 can be computed in a standard way by calculating the covariance between s^j and $s^{j'}$ on the complement of $A_{j,j'}$. The variance, s^2, can be made arbitrarily large if we let the fraction of type 2 individuals converge to one-- the presence of individuals who are sensitive to social interactions increases the variance across populations.

The dynamics are also possible to compute. The expected number of workers in the first period is npq_1; the expected number of entrants in the second period is $np(1-p)q_1 + 2np^2q_1q_2$. Hence if $q_2 > 1/2$ the expected number of entrants in the second period is larger than in the first period. This fact means that if q_2 is large then the expected cumulative migration starts as a convex function of time. Eventually, the expected number of workers converges to $n\left(1-\frac{q_0^2}{(q_0+q_1)^2}\right)$. This indicates that an S-shaped cumulative entrant curve is to be expected.[6]

Unfortunately we do not have closed form solutions that would allow us to estimate these curves at this time. Instead, we will present some results based on a simpler method of estimating social interactions in a dynamic setting.

While this dynamic local interactions model needs further investigation, using simulations it appears to be quite close to a dynamic global interactions model, many of which have been studied extensively theoretically and empirically as well (as pioneered by Griliches, 1958, see also Besley and Case, 1991). While global interactions models have appeared regularly in the literature on technological adoption so our presentation is in no sense novel, in the spirit of collecting together a wide number of approaches to measuring social interactions, we present a particularly simple model here, without any claims to innovation. We assume that in each time period a fraction of individuals (which is denoted β-*X*, where *X(t)* is the state variable for the number of workers) receiving exogenous shocks inducing them to work. Likewise all workers also interact with another individual, who is drawn randomly from the pool of individuals. If a non-worker interacts with a worker, then the non-worker will begin working. Thus, there are two sources of growth in the working population-- an exogenous rise due to idiosyncratic shocks and a rise due to interactions, which will be global since individuals meet with each other randomly (if individuals always only met their neighbors then this would be a local interactions model). Given these two processes, at each point in time the fraction of non-workers who begin to work is $a_0+a_1X(t)$.

While this differential equation is not that actually suggested by the previous model, we know that we can fit the simulations of the model quite well (r-squareds typically over 99.9%) with a differential equation of this form, so we believe that this functional form is both reasonable and provides us with a convenient measure of the degree of social interaction. We thus have a differential equation of the form:

$$\dot{X}(t) = (a_0 + a_1 X(t))(b - X(t)) = a + bX(t) + cX(t)^2 \tag{22}$$

This equation is meant to be flexible. Simulations showed that the time series predicted by the previous model is well captured by a differential equation of this form. The relative importance of the α_1 term gives us the importance of social interactions (or contagion) in the process; the relative importance of the α_0 dictates the importance of non-social related forces. One interpretation of this equation, in the context of technology adoption, is that a fraction of those individuals who haven't adopted (but will eventually adopt), adopts each period and a fraction adopts if and only if they meet someone who has already adopted. The solution for this equation (conditional on knowing the initial value) is:

$$(22)\qquad X(t) = \frac{a_1 b - a_0 - (a_1 b + a_0) Tanh\left[-\frac{1}{2} Log\left[\frac{a_0 + a_1 X(0)}{a_1 b - a_1 X(0)} \right] - \frac{a_1 b + a_0}{2} t \right]}{2 a_1}$$

where *X(0)* is the initial value. This equation can itself be fit using maximum likelihood.

The estimate of β describes the final level of the action. The α_0 term tells us about flat growth; the α_1 term tells us about interactive growth. When comparing dynamic processes, if we compare a process which takes 50 years and a process which takes 1 year to get close to β, both α terms will be much bigger in the faster process. To avoid these issues, we normalize assuming a common *T*, for $X(T)=Zb$, where *Z* is a parameter fixed by the econometrician (perhaps .95). This normalization essentially means that each process is normalized so that it takes exactly the same amount of time to run its course. The normalization also means that only one free parameter (chosen to be α_1) other than β remains. Different values of α_1=0. Figure 4.1 shows how different values of This free parameter influences the S-shaped form of the process. When α_1 ˜0 the curve is concave. As there is more interaction, (α_1 increases) the adoption curve becomes more S-shaped.[7]

To show the efficacy of this estimation procedure, we estimate curves for three time processes which seem to be one-sided and social. The first variable of interest is to consider urbanization in the United States, which moved from 5.1% to 75% between 1790 and 1980. Taking 1790 as year zero, and estimating .83 for β, and normalizing the period of urbanization to twenty years (which will be our standard normalized period), our estimates become α_0 =-.007[.001] and α_1 =.33[.01].[8] The share of 17-year olds who graduate from high school rises from 2% in 1870 to approximately 75% in 1970, again following an S-shaped curve. With 1870 as the base year, normalizing and estimating .79 for β, we find that we estimate that α_0 =-.077[.005] and α_1 =.49[.01]. Using the third variable, the ratio of phone to households, we normalize and find that we estimate that α_0 =-.0005[.00003] and α_1 =.56[.01]. The normalized rankings suggest that phones are more interactive than schooling which is more interactive than urbanization. The following graph shows the results in the raw data. The curve closest to the y-axis shows the results for urbanization. The second curve shows the schooling results and the final curve shows the results for telephones.

FIGURE 4.1 goes here

VI. Conclusion

This paper has presented a tour of primary issues in estimating social interactions. A first issue is estimating the extent to which the high variance of different processes should be thought of as the result of multiple equilibria or high variances around those equilibria. Following our first estimation technique, which essentially asks whether the distribution of city variances is best fit by one or more distributions, we found that multiple distributions fit the data better. Our second estimation technique, which involves examining the connection between prediction and actual city level outcomes, reveals no evidence of multiple equilibria. In both cases, we found that there was usually a large component of the variance that was not explained by the existence of multiple equilibria.

A second issue is the extent to which interactions are due to local or global interaction processes. We show that the key to estimating which processes operate are to have data at the sub-city level. Using sub-city data tends to support the importance of local (tract or sub-tract level) interactions relative to city level interactions.

A third issue is the extent to which interactions reflect sorting on tastes and the extent to which they reflect social interactions after sorting occurs. Of course, sorting itself reflects the presence of some social interaction that induces like individuals to be with each other. We found large evidence of sorting behavior.

Finally, we examined a simple dynamic model and used a simple methodology that lets us compare the degree of social interactions across different dynamic processes.

Appendix

In this appendix we establish for the model in section III that (i) The average action in a population of size n, $\hat{A}_n$, converges, as $n \to \infty$, to a solution of the equation:

(A1) $$g(\bar{A}) = \frac{1 - a f}{a(1 - f)} \bar{A} \quad ,$$

and (ii) if $a_i^n = A_i^n - \bar{A}$, then

(A2) $$\sum_{i=1}^{n} \frac{a_i^n}{\sqrt{n}} \to N\left(0, \frac{s_q^2}{\left(1 - a j - a(1 - j) g'(\bar{A})\right)^2}\right)$$

Define B by:

(A3) $$B = \left(\frac{1}{1 - a f}\right)\left(\frac{\sup(|q|)}{1 - a f} + \frac{\sup(|g|)}{a f}\right).$$

It follows from equation (8) in section III that, since $A_0^n = A_n^n$, then $A_0^n \le B$ and if $|A_i^n| \le B$ then $|A_{i+1}^n| \le B$. Hence for each n, i, $|A_i^n| \le B$, and $|\hat{A}_n| \le B$. Also for each $i \le n$,

(A4) $$\left|\frac{\sum_{k \neq i} A_k^n}{n - 1} - \hat{A}_n\right| \le \left|\frac{\hat{A}_n}{n - 1} - \frac{A_i^n}{n - 1}\right| \le \frac{2B}{n - 1}$$

Since $|g'(A)| \le K$ for some K, we have that:

(A5) $$\left|\frac{1}{n}\sum_{i=1}^{n} g\left(\frac{\sum_{j\neq i} A_j^n}{n-1}\right) - g(\hat{A}_n)\right| \leq \frac{2BK}{n-1}$$

Using equation (10) from section III and the strong law of large numbers

$$(1-af)\hat{A}_n - \frac{a(1-f)}{n}\sum_{i=1}^{n} g\left(\frac{\sum_{k\neq i} A_k^n}{n-1}\right) \rightarrow 0 \text{ with probability one.}$$

Hence using (A5) we have that with probability one:

(A6) $$(1-af)\hat{A}_n - a(1-f)g(\hat{A}_n) \rightarrow 0 \quad .$$

Since $\hat{A}_n$ is a bounded sequence, it must have limit points. From equation (A6) we have that any such limit point $\bar{A}$ must satisfy $g(\bar{A}) = \frac{1-af}{a(1-f)}\bar{A}$, *i.e.* (A1).

Since $\left|\hat{A}_n - \hat{A}_{n-1}\right| \leq \frac{2B}{n}$ and as equation (A1) has a finite number of fixed points, all the limit points of a given sequence must coincide. Hence $\hat{A}_n$ must in fact converge to some $\bar{A}$ that solves (A1), which establishes the first claim of this appendix.

Combining equations (A1) and (10) from section III and multiplying by $\sqrt{n}$, we obtain:

(A7) $$\sqrt{n}(1-af)(\hat{A}_n - \bar{A}) - \sqrt{n}a(1-f)[g(\hat{A}_n) - g(\bar{A})] = \frac{\sum_{i=1}^{n} q_i}{\sqrt{n}} + \sqrt{n}a(1-f)[g(\hat{A}_n) - \frac{\sum_{i=1}^{n} g\left(\frac{\sum_{k\neq i} A_k^n}{n-1}\right)}{n}].$$

Equation (A5) implies that the second term in the right hand side of (A7) converges to 0. Hence the Central Limit Theorem guarantees that the right hand side of (A7) converges to a normal random variable with mean zero and variance s_q^2.

We can now establish:

Proposition 1: $\sqrt{n}(\hat{A}_n - \bar{A})$ is bounded with probability 1.

Proof: The right hand side of equation (A7) is bounded with probability 1. Suppose that a subsequence n_k has the property that $\sqrt{n_k}\left|\hat{A}_{n_k} - \bar{A}\right| \to \infty$. Dividing both sides of equation (A7) by $\sqrt{n_k}\left|\hat{A}_{n_k} - \bar{A}\right|$ and taking the limit as $n_k \to \infty$ we establish that $g'(\bar{A}) = \dfrac{1 - af}{a(1-f)}$, what is a contradiction. We denote $a_i^n = A_i^n - \bar{A}$. We know that:

(A8)

$$\sqrt{n}(1 - af)(\hat{A}_n - \bar{A}) - \sqrt{n}a(1-f)g'(\bar{A})(\hat{A}_n - \bar{A}) + \sqrt{n}a(1-f)o(\hat{A}_n - \bar{A}) \to N(0, s_q^2)$$

.

Proposition 1 implies that $\sqrt{n}o(\hat{A}_n - \bar{A}) \to 0$. Hence the second claim of this appendix is established.

References

Akerlof, G. (1997) "Social Distance and Social Decisions," *Econometrica* 65(5) 1005-1028.

Akerlof, G., Katz and J. Yellen (1996) "An Analysis of Out-of-Wedlock Childbearing in the United States," *Quarterly Journal of Economics* CXI (2): 277-318.

Akerlof, G. and R. Kranton (1997) "The Economics of Identity," Brookings mimeograph.

Arthur, W. B. (1989) "Increasing Returns, Competing Technologies and Lock-in by Historical Small Events: The Dynamics of Allocation under Increasing Returns to Scale," *Economics Journal* 99(1): 116-131.

Banerjee, A. (1992) "A Simple Model of Herd Behavior," *Quarterly Journal of Economics* 107: 797-818.

Becker, G. (1997) "Social Economics," University of Chicago mimeograph.

Benabou, R. (1993) "Workings of a City: Location, Education and Production," *Quarterly Journal of Economics* CVIII: 619-652.

Bernheim, D. (1994) "A Theory of Conformity," *Journal of Political Economy* 102 (5): 841-877.

Besley, T. and A. Case (1994) "Diffusion as a Learning Process: Evidence from HYV Cotton," mimeographed.

Bikhchandani, S., D. Hirshleifer and I . Welch (1992) "A Theory of Fads, Fashions, Customs, and Cultural Change as Information Cascades," *Journal of Political Economy* 85: 365-390.

Borjas, G. (1995) "Ethnicity, Neighborhoods and Human Capital Externalities," *American Economic Review* 85: 365-390.

Brock, W. and S. Durlauf (1995) "Discrete Choice with Social Interactions I: Theory" NBER Working Paper # 2591.

Brock, W. and S. Durlauf (1997) "Discrete Choice with Social Interactions II: Econometrics" University of Wisconsin mimeograph.

Case, A. and L. Katz (1991) "The Company You Keep: The Effect of Family and Neighborhood on Disadvantaged Youth," NBER Working Paper # 3708.

Crane, J. (1991) "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out and Teenage Childbearing," *American Journal of Sociology* 96: 1226-1259.

DiPasquale, D. and E. Glaeser (1997) "The L.A. Riot and the Economics of Urban Unrest," *Journal of Urban Economics,* forthcoming.

Duesenberry, J. (1949) *Income, saving and the theory of consumer behavior*. Cambridge: Harvard University Press.

Ellison, G. and D. Fudenberg (1993) "Rules of Thumb for Social Learning," *Journal of Political Economy* CI:612-643.

Ellison, G. and D. Fudenberg (1995) "Word-of-Mouth Communication and Social Learning" *Quarterly Journal of Economics* CX (1): 93-126.

Evans, W., Oates, W. and R. Schwab (1992) "Measuring Peer Group Effects: A Model of Teenage Behavior," *Journal of Political Economy* 100 (5) 966-991.

Gavaria, A. (1997) "Increasing Returns and the Evolution of Violent Crime: the Case of Columbia," UC San Diego Mimeograph.

Glaeser, E. L. and J. Scheinkman (1997) "Social Interactions and Long-Range Dependence,"

Glaeser, E.L., Sacerdote, B. and J. Scheinkman (1996) "Crime and Social Interactions," *Quarterly Journal of Economics* CXI(2): 507-548.

Glaeser, E.L. "Two Essays on Information and Labor Markets," University of Chicago Ph.D. Dissertation.

Griliches, Z. (1958) "Research Costs and Social Returns: Hybrid Corn and Related Innovations," *Journal of Political Economy* 66: 419-431.

Jovanovic, B. (1985) "Aggregate Randomness in Large Noncooperative Games," mimeographed.

Kindermann, R. and J. L. Snell (1980) "On the relationship between markov random fields and social networks," *Journal of Mathematical Sociology* 7.

Levitt, S. (1997) "The Exaggerated Role of the Changing Age Structure in Explaining Aggregate Crime Changes," University of Chicago

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

Mulligan, C. (1995) "Pecuniary and Nonpecuniary Incentives to Work in the U.S. During World War II," University of Chicago, Population Research Center Discussion Paper Series # 95-3.

Murphy, K., A. Shleifer and R. Vishny "Why is rent seeking so costly for growth?" *American Economic Review* LXXXIII (1993), 409-414.

O'Regan, K. and J. Quigley, "Spatial Effects upon Employment Outcomes: The Case of New Jersey Teenagers," *New England Economic Review*, May/June (1996a), 41-57.

Pesandorfer, W. (1996) "Design Innovation and Fashion Cycles," *American Economic Review* 85(4): 771-792.

Rauch, J. (1994) "Productivity Gains from geographic concentration of human capita: evidence from the cities. *Journal of Urban Economics* 34: 380-400.

Rasmussen, E. (1996) "Stigma and Self-Fulfilling Expectations of Criminality," *Journal of Law and Economics*.

Sah, R. (1991) "Social Osmosis and the Patterns of Crime," *Journal of Political Economy* XCIX, 1272-1295.

Schelling, T. (1978) *Micromotives and Macrobehavior*. New York: Norton.

Topa, G. (1997) "Social Interactions, Local Spillovers and Unemployment," NYU Economics Department mimeograph.

Young, P. (1993) "The Evolution of Conventions," *Econometrica* 61: 57-84.

Young, P. (1997) "Social Coordination and Social Change," Johns Hopkins University, mimeograph.

FIGURE 4.1

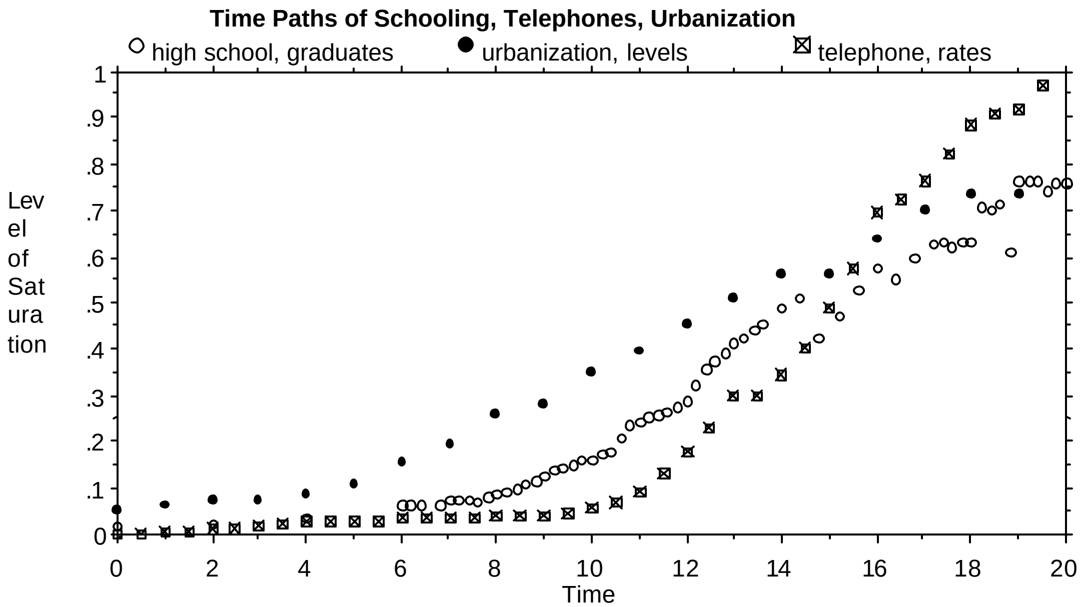
Time Paths of Schooling, Telephones, Urbanization
high school, graduates
urbanization, levels
telephone, rates
Level of Saturation
1
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
0
2
4
6
8
10
12
14
16
18
20
Time

Table 4.1:

Estimation of Strength of Social Interactions Affecting

Female Headship Rates in Families

	Var_{ind}	Var_{agg}	a [1]
Unadjusted Female Headship Rate[2]	.134	171.53	0.998
Rate Controlling for City Level Observables[3]	.132	52.09	0.995
Rate Controlling for State Effects and City Level Observables[4]	.132	19.52	0.987

Source: Individual level data are from the 1990 Census Public Use Micro Sample. Aggregate data are from the 1990 Census Summary Tape Files.

1. $a = \dfrac{Var_{agg} - Var_{ind}}{Var_{agg} + Var_{ind}}$
2. Var_{ind} is the individual-level variance. It is the raw variance of the female headship rate among families in the US. Var_{agg} is the adjusted variance of the city aggregate rate. It is the variance of $\sqrt{N_c}\left(A_c - \overline{A}\right)$ where N_c is the number of families in the city, A_c is the average action in the city, and $\overline{A}$ is the average action in the US.
3. The individual variance is the variance of the female headship rate, controlling for city fixed effects. The aggregate variance is the variance of $\sqrt{N_c}\left(A_c - \dfrac{b'X_c}{1-a}\right)$ where $\dfrac{b_c}{1-a}$ is estimated from a city level regression of A_c on median family income, the number of people aged less than 18 per family, and on the fraction of the population

of families that is black, Hispanic, a high-school dropout, a college graduate, in poverty, and headed by someone aged 15-24, 25-34, 35-44, 55-64 or 65-100. All variables are defined for the population of family heads except the education variables which are defined for the population over 18 years old.

4. The individual variance is the variance of the female headship rate controlling for city fixed effects. The aggregate variance in this row is calculated in the same way as in the second row of the table except that every raw variable is replaced with its deviation from the state mean.

Table 4.2:

Female Headship Rate

Local vs. Global Interactions

	Var_{ind}	Var_{tract}	Var_{city}	$a j$ [1]	s_{q}^{2} [1]	$g'(\overline{A}) a (1 - j)$ [1]
Raw Female Headship Rate in Families[2]	.134	10.74	171.53	0.975	0.0017	.025
Female Headship Rate Controlling for Individual Traits and City Fixed Effects[3]	.111	2.94	83.69	0.927	0.016	0.073
Female Headship Rate Controlling for Individual Traits and State Effects[4]	.111	2.92	36.23	0.927	0.016	0.073

Source: 1990 Census Summary Tape Files and Public Use Microsample

1. $a j = \frac{1 - Var_{ind} / Var_{tract}}{1 + Var_{ind} / Var_{tract}}$ $\quad s_{q}^{2} = Var_{ind}\left(1 - a^{2} j^{2}\right)$ $\quad g'(\overline{A}) a (1 - j) = 1 - a j - s_{q}^{2} / Var_{city}$

2. Var_{ind} is the individual variance from Table 4.1 row 1. Var_{tract} is the average across cities of the variance of adjusted tract averages: $\sqrt{N_t}(A_t - A_c)$ where N_t is the population of families in the tract, A_t is the female headship rate in the tract, and A_c is the average in the city. Var_{city} is the aggregate variance from Table 4.1 row 1.

3. Var_{ind} is the variance of the residual from the following micro-regression: $A_i = b'X_i + e_i$ where A_i is the deviation of the female headship rate from the city mean and X_i is a vector of deviations of individual traits from city means. These traits include income and number of children and a set of dummies indicating whether the family head is: black, Hispanic, a high-school dropout, a college graduate, aged 18-24, 25-34, 35-44, 55-64, 65-100, and in poverty. Var_{tract} is the average across cities of the within-city variance of $\sqrt{N_t}\left(\hat{A}_t - \hat{A}_c\right)$ where $\hat{A}_t = A_t - \hat{b}'X_t$, $\hat{b}$ is the vector of parameter estimates from the micro-level regression, X_t is the vector of tract-level averages of individual traits (not deviated from city means), and $\hat{A}_c$ is the within-city average of $\hat{A}_t$. Finally, Var_{city} is the variance of $\sqrt{N_c}\left(\hat{A}_c - \hat{A}\right)$ where, as for the tract level variables, $\hat{A}_c = A_c - \hat{b}'X_c$ and $\hat{A}$ is the national average of $\hat{A}_c$.
4. The variables in row 3 of the table are calculated in an analogous manner to those in row 2 except that the aggregate variances are calculated controlling for state fixed effects.

TABLE 4.3:

Parameter Estimates for Global vs. Local Interactions Model

	$a j$	s^{2}_{q}	$g'(\overline{A})a(1-j)$
Female Headship Rate in Families	0.905775	0.001737	0.247777
Fraction of Population Over 5 in Same House as 1985	0.991719	0.000017	0.004193
Unemployment Rate	0.959802	0.000093	0.076921
Fraction not in Labor Force	0.983460	0.000061	-0.015368
Fraction on Welfare	0.978419	0.000030	0.061562
Fraction in Poverty	0.992624	0.000011	0.014454
Fraction of Housing Owner Occupied	0.992535	0.000012	0.026431
Number of Cars	-0.921217	16280.685916	-1.091941
Average Rent	0.9999996172960	0.00000000000015	0.000002

Source:

Aggregate Estimates: 1990 Census PMS SA data

Individual Estimates: 1990 Census Public Use Microsample.

Note:

Variable of interest is: sqrt(city population)*(city level rate-country level mean)

Table 4.3a:
Female Headship Rate of Families
Multiple Equilibria Model[1] (3 Component Distribution is Optimal)

	Components of Distribution of Adjusted City Female Headship Rates of Families					Average Aggregate Variance[2]	$g'(\overline{A})a(1-j)^3$
	First	Second	Third	Fourth	Fifth		
Mean	-2.93						
Variance	171.02					171.02	.0250
Weight	1.00						
Means	-4.58	15.740					
Variances	55.50	1097.40				139.89	.0250
Weights	0.92	0.08					
Means	-3.35	-5.17	32.75				
Variances	120.77	12.52	1680.20			126.98	.0250
Weights	0.55	0.42	0.03				
Means	-5.02	-5.04	3.75	142.50			
Variances	77.76	12.17	307.44	0		96.06	.0250
Weights	0.42	0.38	0.19	0			
Means	-8.18	-5.32	-1.96	4.38	142.50		
Variances	71.40	10.22	46.21	306.07	0	89.32	.0250
Weights	0.25	0.31	0.25	0.19	0		

Source: 1990 Census Summary Tape Files

1. The adjusted PMSA female headship rate defined as $\sqrt{N_c}\left(A_c - \overline{A}\right)$ where N_c is the city population of families, A_c is the city female headship rate among families, and $\overline{A}$ is female headship rate in the US, is modeled as a random variable distributed as a mixture of normals. The mixtures are estimated using the em algorithm. The Akaike information criterion is minimized by the three component distribution.

2. The average aggregate variance is the weighted average of the variances of the components of the overall distribution.

3. As in Table 4.2, $g'\left(\overline{A}\right)a\left(1-j\right)=1-aj-s_q^2 / Var_{city}$. aj and s_q^2 are estimated in Table 4.2 row 1 and Var_{city} is the average aggregate variance calculated in this table.

Table 4.3b:

Female Headship Rate of Families

Controlling for Individual Traits and City Fixed Effects

Multiple Equilibria Model[1] (3 component Distribution is Optimal)

	Components of Distribution of Adjusted City Female Headship Rates of Families					Average Aggregate Variance[2]	$g'(\overline{A})a(1-j)$ [3]
	First	Second	Third	Fourth	Fifth		
Mean	-2.72					83.44	.0728
Variance	83.44						
Weight	1.00						
Means	-4.77	5.35				66.49	.0728
Variances	22.60	242.07					
Weights	0.80	0.20					
Means	0.90	-4.99	55.75			56.72	.0727
Variances	126.15	16.98	497.29				
Weights	0.32	0.67	0.01				
Means	-3.33	-4.88	18.65	51.73		42.51	.0726
Variances	63.07	11.00	15.03	541.84			
Weights	0.50	0.45	0.04	0.01			
Means	-3.42	-9.35	-4.44	18.61	51.74	41.88	.0726
Variances	61.33	1.38	8.45	15.12	541.72		
Weights	0.53	0.04	0.39	0.04	0.01		

Source: 1990 Census Summary Tape Files

1. The adjusted PMSA female headship rate controlling for individual traits and city fixed effects is defined as $\sqrt{N_c}\left(\hat{A}_c - \hat{A}\right)$ as in Table 4.2 row 2. This random variable is modeled as a random variable distributed as a mixture of normals. The mixtures are estimated using the em algorithm. The Akaike information criterion is minimized by the three component distribution.

2. The average aggregate variance is the weighted average of the variances of the components of the overall distribution.

3. As in Table 4.2, $g'\left(\overline{A}\right)a\left(1-j\right)=1-aj-s_q^2 \big/ Var_{city}$. aj and s_q^2 are estimated in Table 4.2 row 2 and Var_{city} is the average aggregate variance calculated in this table.

Table 4.3c:

Female Headship Rate of Families

Controlling for Individual Traits and State Effects

Multiple Equilibria Model[1] (2 Component Distribution is Optimal)

	Components of Distribution of Adjusted City Female Headship Rates of Families					Average Aggregate Variance[2]	$g'(\overline{A})a(1-j)$[3]
	First	Second	Third	Fourth	Fifth		
Mean	-1.98						
Variance	36.12					36.12	.0726
Weight	1.00						
Means	-2.81	1.03					
Variances	8.97	121.67				33.76	.0726
Weights	0.78	0.22					
Means	-0.48	-2.92	50.59				
Variances	67.86	7.10	0			26.54	.0724
Weights	0.32	0.68	0				
Means	-2.62	-3.03	-0.46	50.59			
Variances	9.96	6.08	69.34	0		26.62	.0724
Weights	0.24	0.45	0.31	0			
Means	-14.51	-8.71	-2.95	0.41	23.10		
Variances	16.39	0.16	6.57	46.29	347.86	21.57	.0723
Weights	0.02	0.01	0.66	0.29	0.01		

Source: 1990 Census Summary Tape Files

1. The adjusted PMSA female headship rate controlling for individual traits and state effects is defined as $\sqrt{N_c}\left(\hat{A}_c - \hat{A}\right)$ as in Table 4.2 row 3. This random variable is modeled as a random variable distributed as a mixture of normals. The mixtures are estimated using the em algorithm. The Akaike information criterion is minimized by the two component distribution.

2. The average aggregate variance is the weighted average of the variances of the components of the overall distribution.

3. As in Table 4.2, $g'\left(\overline{A}\right)a\left(1-j\right) = 1 - aj - s_q^2 \big/ Var_{city}$. aj and s_q^2 are estimated in Table 4.2 row 3 and Var_{city} is the average aggregate variance calculated in this table.

Table 4.4:

Estimation of Strength of Social Interactions

Global vs. Local Interactions

Using Regression of Predicted City-Level Female Headship Rate

on Actual City-Level Rate

	Var^{1}_{ind}	Var^{2}_{tract}	$\frac{\partial \overline{f(X)}}{\partial \overline{A}}$ [3]	$a j$ [4]	$s_q^2 + Var_j^{f(X)}$ [4]	$g'(\overline{A})a(1-j)$ [4]
Female Headship Rate Controlling for Individual Traits and City Fixed Effects	.111	2.94	.792 (0.052)	0.927	0.016	-0.719
25th percentile[3]			.386 (.176)			-0.178

Source: Individual level data are from the 1990 Census Summary Tape Files. Aggregate level data are from the 1990 Census Public Use Microsample.

1. Var_{ind} is the variance of the residual from the following micro-regression: $A_i = b'X_i + e_i$ where A_i is the deviation of the female headship rate from the city mean and X_i is a vector of deviations of individual traits from city means. These traits include income and number of children and a set of dummies indicating whether the family head is: black, Hispanic, a high-school dropout, a college graduate, aged 18-24, 25-34, 35-44, 55-64, 65-100, and in poverty.
2. Var_{tract} is the average across cities of the within-city variance of $\sqrt{N_t}\left(\hat{A}_t - \hat{A}_c\right)$ where $\hat{A}_t = A_t - \hat{b}'X_t$, $\hat{b}$ is the vector of parameter

estimates from the micro-level regression, X_t is the vector of tract-level averages of individual traits (not deviated from city means), and $\hat{A}_c$ is the within-city average of $\hat{A}_t$.

3. $\frac{\partial \overline{f(X)}}{\partial \bar{A}}$ is the slope estimated from a spline-regression of $\hat{A}_c = \hat{b}'X_c$ on the actual rate, A_c. Row 1 is the estimate of the slope for those cities above the 25th percentile. Row 2 is the estimate of the slope for cities below the 25th percentile.

4. $aj = \frac{1 - Var_{ind}/Var_{tract}}{1 + Var_{ind}/Var_{tract}}$ $\quad s_q^2 = Var_{ind}\left(1 - a^2 j^2\right)$ $\quad g'\left(\bar{A}\right)a\left(1 - j\right) = 1 - aj - \frac{\partial \overline{f(X)}}{\partial \bar{A}}$

TABLE 4.5:

Sorting Equilibria:

Results from Regression Analysis

Equation Estimated is:

$$\text{var}_l\left(N\sqrt{k_j}h_j\right)= b_0 + b_1(n_l / N_l) + e_l$$

	b_0	b_1	R^2	Var_{ind}	a
Female Headship Rate of Families	39.42 (146.76)	0.005 (1.5E-05)	0.169	0.16	0.992
Crime Rate	30.88 (11.44)	2.6E-04 (1.2E-04)	0.095	0.06	0.996
Percent Non-white	-299.39 (508.47)	0.018 (0.005)	0.201	1.88E-06	---

Sources: Regression data and the variances of crime and percent non-white are derived from the County and City Data Book 1994 . The female headship rate data is from the 1990 Census Public Use Microsample.

Note: Standard errors are in parentheses. h was estimated as: $\sqrt{\overline{N}}\left(A_i - \overline{A}\right)$ where $\overline{N}$ is the average city size, A_i is the action, and $\overline{A}$ is the average of A_i. k_j is 1/number of cities in the state. a is calculated using the formula, $a = \frac{Var_{agg} - Var_{ind}}{Var_{agg} + Var_{ind}}$ where the intercept, b_0, is used for Var_{agg} and Var_{ind} is the individual-level variance of the action in the US.

[*] Both authors acknowledge financial support from the National Science Foundation. Glaeser also thanks the Sloan Foundation. Andrea Eisfeldt and Lars Nesheim provided excellent research assistance.

[1]Many of these ideas had antecedents in the classics works on social interactions, such as Schelling (1978) which presents a discussion of a wide range of social interactions, and Duesenberry (1949) who first formalizes interdependent preferences. Jovanovic (1985) is also a particularly prescient formalization of a social interactions model.

[2] Akerlof, Katz and Yellen (1996) actually specifically link the rise in out-of-wedlock births with changes in abortion and birth control, but this link is indirect and works through the stigma associated with being an unwed mother or a delinquent father.

[3]Possible exceptions to this might occur when $?W/?q_i > 0$ if individuals don't want to take the action if no one else is taking the action, because it has no signaling value in that case. In that case, some consumers are necessary for their to exist the positive sorting equilibrium. Of course well established theory about reasonable beliefs when no one is taking an action, argues that people's beliefs about off-the-equilibrium path behavior should ensure that the action still has positive signaling value when no one is taking the action.

[4]Furthermore, it will eliminate the effect of any variables that are city, rather than individual, specific. Of course, it will not eliminate the problems of sorting across neighborhoods. That problem can only be solved with neighborhood fixed effects.

[5]In general this will still leave a forecasting problem for agents since they may be better off waiting for their friends to act. However, in our model we will assume that the gains from acting are such that each type will either never act or act as soon as a certain fraction of their friends have acted.

[6]We can also obtain an S-shaped curve where every agent interacts equally with every other agent. Under a global interactions model, we lose the variance across populations over and above the characteristics of the populations.

[7]There is a literature on this topic which we are not referencing. We apologize at this point for failing to survey the technology adoption literature adequately at this point. This excellent and lengthy literature is, of course, connected to this topic but too far afield from our basic interest to be given significant page space.

[8]These errors are biased because we have treated the observations as independent, further work will deal with the variety of standard time series issues involved in estimating this non-linear trend. The r-squared, which is again somewhat misleading is 99.62%, which is typical for these estimates.