

NBER WORKING PAPER SERIES

THE SOCIAL MULTIPLIER

Edward L. Glaeser
Bruce I. Sacerdote
Jose A. Scheinkman

Working Paper 9153
<http://www.nber.org/papers/w9153>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2002

The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

© 2002 by Edward L. Glaeser, Bruce I. Sacerdote, and Jose A. Scheinkman. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Social Multiplier
Edward L. Glaeser, Bruce I. Sacerdote, and Jose A. Scheinkman
NBER Working Paper No. 9153
September 2002

ABSTRACT

In many cases, aggregate data is used to make inferences about individual level behavior. If there are social interactions in which one person's actions influence his neighbor's incentives or information, then these inferences are inappropriate. The presence of positive social interactions, or strategic complementarities, implies the existence of a social multiplier where aggregate relationships will overstate individual elasticities. We present a brief model and then estimate the size of the social multiplier in three areas: the impact of education on wages, the impact of demographics on crime and group membership among Dartmouth roommates. In all three areas there appears to be a significant social multiplier.

Edward L. Glaeser
Department of Economics
315 A Littauer Center
Harvard University
Cambridge, MA 02138
and NBER
eglaeser@harvard.edu

Bruce I. Sacerdote
6106 Rockefeller Hall
Department of Economics
Dartmouth College
Hanover, NH 03755
and NBER
bruce.i.sacerdote@dartmouth.edu

Jose A. Scheinkman
Princeton University

I. Introduction

Empirical work in the social sciences frequently attempts to infer individual behavior from statistical work on aggregates. Individual labor supply is inferred from changes in the tax schedule. Crime deterrence elasticities are inferred from changes in policing or punishment. Changes in policies are often seen as our best means of inferring underlying economic behavior because variation in these policies is, in some cases, orthogonal to individual-specific error terms.

However, using aggregate variation to infer individual-level parameters is problematic when there are positive (or negative) social interactions. If one person's proclivity towards crime influences his neighbor's criminal behavior, then a change in policing will have both a direct effect on crime and an indirect effect through social influence. The presence of positive spillovers or strategic complementarities creates a "social multiplier" where aggregate coefficients will be greater than individual coefficients (as described by Becker and Murphy, 2000). A large body of recent work (including Katz, Kling and Liebman, 2001, and Ludwig, Hirschfeld and Duncan, 2001) seems to confirm the existence of these spillovers in a number of areas. As such, an estimated aggregate elasticity incorporates both the true individual level response and effects stemming from social interactions.¹

For many purposes, particularly policy-related ones, researchers actually want the aggregate coefficient that includes both the individual level response and the social multiplier. In that case, aggregate empirical work is appropriate. Still, it is crucial that the empirical work is done at the same level of aggregation as the ultimate policy. For example, if we want to know the effect of a national change in crime policy, but we work with city-level data, then we will miss the impact of all cross-city interactions. To adequately infer state-level effects from city-level coefficients, we need to know both the

¹ There is a long literature that discusses the so-called general equilibrium effects which may be missing from some econometric estimates. In a sense, positive externalities are just one type of general equilibrium effect.

power of social interactions and the degree to which those interactions decay across jurisdictions.

We refer to the estimated ratio of aggregate coefficients to individual coefficients as “the social multiplier.”² So, if wages are regressed on years of schooling at the individual and at the state level, the ratio of these two coefficients is the social multiplier. It is also true that the same social multiplier can be estimated by regressing aggregate outcomes on aggregate predicted outcomes, where the predictions are based on individual level regressions. In this paper, we present a theoretical framework which maps this estimated social multiplier with underlying social influence variables.

Our theoretical framework tells us that if an individual’s outcome rises “x” percent as his neighbor’s average outcome, then the social multiplier roughly equals $1/(1-x)$ for large enough groups. As such, big social multipliers do not tend to occur unless the value of “x” is .33 or higher. If the spillover works through the neighbors’ exogenous characteristics, not through their outcomes (i.e. your propensity for crime is influenced by your neighbors’ parents’ characteristics, not by their crime level), then typically the social multiplier is smaller.

The presence of sorting will also impact the measured social multiplier. If there is sorting on observables and positive social interactions, then the individual level coefficient will overstate the true individual level relationship. The intuition of this claim is that with sorting, one person’s education will be correlated with his neighbor’s education and the effect of my education (in an individual-level regression) will overstate the true impact of education because it includes spillovers. The presence of this bias will mean that the measured social multiplier will tend to underestimate the true level of social interactions. On the other hand, correlation between aggregate observables and aggregate unobservables will cause the measured social multiplier to overstate the true level of social interactions.

We also introduce a model where social influence exponentially decays with social distance. This introduces a two-parameter model which can capture both the level of social interactions and the degree to which social interactions become less important with social distance. The downside of this exponential social influence model is that it is not good for dealing with very large social groupings, such as counties and states: exponential decay generally will mean that people do not significantly interact with people who are outside of their county.

We then apply our framework to three different contexts. First, we follow Sacerdote (2001) and examine the magnitude of these interactions among Dartmouth college roommates. The Dartmouth roommates data have the advantage of little unobserved heterogeneity and randomized social groups. In this case, we find weak social interactions in academic achievement, but strong interactions in social group membership. We estimate a social multiplier of 1.4 in groups of eight (floors) and 2.2 in groups of 28 (dorms). These estimates are compatible with a weak degree of social influence that then decays very slowly.

Second, we follow Levitt (1999) and look at the influence of demographics on the crime rate. Levitt (1999) argues that the coefficients on age from individual level regressions are far too small to suggest large swings in crime that are related to aggregate changes in the demographic structure. For example, these coefficients tell us that the baby boom can at best explain one-fifth of the rise in crime between 1960 and 1975. While this argument is correct, social interactions may help us to understand why demographics appear to be related to crime in time series regressions. Using crime data we find significant evidence of a social multiplier of 1.7 at the county level, 2.8 at the state level and 8.2 at the nation level. These are extremely high estimates and we don't necessarily believe in the level of social interactions that they imply. Still, they certainly suggest that the aggregation level is crucial.

² In fact, there is a slight difference between the Becker and Murphy definition of social multiplier and our own, although our definition represents a monotonic transformation of the social multiplier as they define

Finally, we follow Rauch (1993) and Acemoglu and Angrist (1999) and turn to the issue of human capital spillovers. In this case, we use individual coefficients to create a predicted wage for an aggregate (such as a state or Public Use Microsample Area, or PUMA) and then regress actual wage on predicted wage. Using this approach, we estimate a social multiplier of 1.67 at the PUMA level and 2.17 at the state level.

We agree with Manski's (1993) generally pessimistic view of the ability to identify social interaction parameters, at least in the absence of true randomization. However, our results suggest that social interactions may be large, and that coefficients at different levels of aggregation differ significantly, either because of social interactions or because of non-random sorting across different areas. While we remain cautious in interpreting our parameter estimates, we do believe that our evidence casts doubt on the use of aggregate changes to make inferences about individual level parameters.

II. A Framework

We follow Glaeser and Scheinkman (2002), and present a simple framework which will connect coefficients from regressions run at different levels of aggregation with underlying social interaction parameters. The simplest algebraic representation of the social multiplier can be shown with a global interaction model, where one person's action depends on the average action in a group.

One particularly simple model of this type is that $A_i = \theta_i + \frac{\gamma}{N-1} \sum_{j \in G(i), j \neq i} A_j$, where A_i is the action of person i , $G(i)$ refers to person i 's group which is of size N , γ is the social interaction parameter and θ_i reflects the exogenous forces increases the level of the

it. Goldin and Katz (2002) use the term in a way that could encompass either definition.

action.³ We assume that groups represent a non-overlapping partition of the entire economy, so that if $i \in G(j)$ then $j \in G(i)$. In general, we will assume that $\theta_i = \sum_k \beta_k X_k^i + \varepsilon_i$, where X_k^i is the value of attribute k for person i and β_k is the direct impact of attribute k and ε_i is a person-specific random effect.

This model implies that $\frac{1}{N} \sum_{j \in G(i)} A_j = \frac{1}{1-\gamma} \frac{1}{N} \sum_{j \in G(i)} \theta_j$ and

$$A_i = \theta_i \left(1 + \frac{\gamma^2}{(1-\gamma)(N-1+\gamma)} \right) + \frac{\gamma}{N-1+\gamma} \sum_{j \in G(i), j \neq i} \theta_j. \quad \text{The term } \frac{\gamma^2}{(1-\gamma)(N-1+\gamma)} \text{ captures}$$

the fact that if person A has an intrinsically higher propensity to do an activity this will both have a direct impact on his activity level, but will also indirectly impact his activity through its influence on the other individuals in the group. When N is large, this term will be negligible (as long as γ is bounded away from one). When N=2, this term becomes $\frac{\gamma}{1-\gamma}$, which is greater than one whenever $\gamma > .5$.

The simplest case occurs when the values of θ_i are independent within a group, and where the included X_k^i regressors are independent of the error term. In that case, an individual-level regression where A_i is regressed on an exogenous variable X_k^i yields a coefficient estimate of $\beta_k \left(1 + \frac{\gamma^2}{(1-\gamma)(N-1+\gamma)} \right)$. An aggregate regression where group level average outcomes are regressed on group level average X_k^i characteristics yields a coefficient estimate of $\frac{\beta_k}{1-\gamma}$.

We define the social multiplier as the ratio of the group level coefficient to the individual level coefficient, or the amount that the coefficient rises as we move from individual to

³ This equation can be justified by assuming that people maximize $A_i \left(\theta_i + \frac{\gamma}{N-1} \sum_{j \in G(i), j \neq i} A_j \right) - \frac{A_i^2}{2}$.

group level regressions. In this case, the social multiplier will equal $\frac{1}{1-\gamma + \frac{\gamma^2}{(N-1+\gamma)}}$

or $\frac{N-1+\gamma}{(1-\gamma)(N-1)+\gamma}$. As N grows large, this value approaches $\frac{1}{1-\gamma}$. When γ is small, this coefficient will be close to one and as γ approaches one, the social multiplier approaches infinity.

The social multiplier can also be estimated by regressing group level outcomes on the group level outcome that would be predicted using individual coefficients. If this procedure is followed, the coefficient estimated by regressing average community

outcome on $\sum_k \beta_k \left(1 + \frac{\gamma^2}{(1-\gamma)(N-1+\gamma)} \right) \sum_{j \in G(i)} X_j^k / N$, the outcome predicted by

coefficients estimated using individual level regressions, again equals $\frac{1}{1-\gamma + \frac{\gamma^2}{(N-1+\gamma)}}$

One generalization of our assumption is to assume that there is sorting across groups, at least on the basis of observables. To include sorting within the framework, we $X_k^i = \bar{X}_k + \mu_k^i$, where \bar{X}_k represents a group level average and μ_k^i represents an individual specific component which is independent across people. We maintain the assumption that the values of X_k^i are independent across characteristics (although not across people) and independent of the error term. We use the notation $\sigma = \frac{Var(\bar{X}_k)}{Var(X_k^i)}$,

where σ represents the share of the variation in observable characteristics which is due to the group level component.

This new assumption does not change the group level regression coefficient, which remains $\frac{\beta_k}{1-\gamma}$. However, allowing a group-specific correlation of characteristics does

influence the individual level coefficients because now the individual-level coefficient includes the impact of a correlation between the individual's X_k^i value and the X_k^j values of his neighbors. If we see a correlation between the school outcomes of children and the schooling of their parents, this may all occur because parental schooling influences children's outcomes directly. Alternatively, it may in part reflect the fact that well-schooled parents live together and as a result, the children of the more educated have more successful peers.

The estimated coefficient from an individual level regression now equals $\beta_k \left(1 + \frac{\gamma^2 + \sigma\gamma(1-\gamma)(N-1)}{(1-\gamma)(N-1+\gamma)} \right)$. The bias due to the cross-person correlation roughly equals $\sigma\gamma$ --the product of the degree of sorting and the amount of social influence. If either sorting or social influence is unimportant, then this term is small and can be ignored, but we think in many cases, both of these terms will be big.

The social multiplier now equals $\frac{N-1+\gamma}{(1-\gamma)(N-1)(1+\sigma\gamma)+\gamma}$, which will approach $\frac{1}{(1-\gamma)(1+\sigma\gamma)}$ as N gets large. If we know the value of σ , then we can infer the size of the social influence parameter γ . This formula implies that for high levels of N, the presence of sorting on observables will always cause the social multiplier to decline. As such, the measured social multiplier will tend to understate the true level of social interactions, primarily because the individual level coefficient is biased upwards.⁴

We now consider the case where there is sorting across neighborhoods on the basis of unobservable characteristics. To formalize this, we assume that $\varepsilon_i = \bar{\varepsilon} + \nu_i$, where $\bar{\varepsilon}$

⁴ If the social multiplier is estimated by regressing an aggregate outcome on a predicted aggregate outcome, where the predicted value is formed using individual level coefficients then the estimated multiplier will again equal $\frac{N-1+\gamma}{(1-\gamma)(N-1)(1+\sigma\gamma)+\gamma}$, but in this case it is necessary that $\sigma = \frac{Var(\bar{X}_k)}{Var(X_k^i)}$ for each of the observable variables.

represents a community specific average level of the unobserved shock. Furthermore, we assume that $\bar{\varepsilon}$ and the \bar{X}_k in question covary. In particular, we assume that $\lambda = \frac{Cov(\bar{X}_k, \bar{\varepsilon})}{Var(\bar{X}_k)}$. The person-specific shocks are still assumed to be independent of each other.

In this case, the person specific coefficient will equal $\beta_k \left(1 + \frac{\gamma^2 + \sigma\gamma(1-\gamma)(N-1)}{(1-\gamma)(N-1+\gamma)} \right) + \lambda\sigma\gamma \frac{N-1}{N-1+\gamma}$. The bias created by sorting on observables will approximately equal $\lambda\sigma\gamma$ as N grows large. The aggregate coefficient now becomes $\frac{1}{1-\gamma} \left(\beta_k + \lambda \frac{\sigma N}{\sigma N + 1 - \sigma} \right)$. The social multiplier, in this case, equals:

$$\frac{(N-1) \left(1 + \frac{\lambda}{\beta_k} \frac{\sigma N}{\sigma N + 1 - \sigma} \right)}{(N-1) \left(1 - \gamma + \sigma\gamma \left(1 - \gamma + \frac{\lambda}{\beta_k} \right) \right) + \gamma}, \text{ or } \frac{1 + \frac{\lambda}{\beta_k}}{(1-\gamma)(1+\sigma\gamma) + \frac{\sigma\gamma\lambda}{\beta_k}} \text{ as N gets large.}$$

The social multiplier will rise with λ if and only if $1 > \gamma(1 + \sigma\gamma)$. The reason that the social multiplier can either rise or fall with this form of sorting is that sorting impacts both the micro and macro coefficients. If $1 > \gamma(1 + \sigma\gamma)$ then the macro-coefficient will increase with sorting more than the micro-coefficient and an increase in sorting causes the social multiplier to rise. If this condition does not hold, which really only occurs when the form of social interactions are very intense (and the degree of sorting on observables is quite high), then increases in sorting mean the micro-coefficient increases significantly through the sorting related bias and this causes the coefficient to increase.

In general, we will rarely know the value of λ . In some cases, such as the Dartmouth College Roommates data set described below, we know that roommates are randomly assigned and in that case λ equal zero. However, in other cases, it might be quite high

and we can only guess about the extent that this influences the measured social multiplier.

We now turn to three variations on this model. First, we consider the case of a continuous outcome where the externality depends on the innate characteristics of the individuals and not their actions (or outcomes). This is particularly useful in the case of human capital spillovers where we think that wages (and productivity) may well be a function of human capital in the area. In this case the model becomes

$$A_i = \theta_i + \frac{\gamma}{N-1} \sum_{j \in G(i), j \neq i} \theta_j.$$

The expected value of the individual level coefficient on X_k will equal β_k if there is no correlation across people in the value of X_k . When there is correlation, the individual level coefficient equals $(1 + \sigma\gamma)\beta_k$. When there is correlation both across people in the value of X_k and sorting on the basis of unobservables, the micro-level coefficient equals $(1 + \sigma\gamma)\beta_k + \sigma\gamma\lambda$.

When there is no sorting on unobservables (i.e. whether there is correlation in observables or not), the expected value of the group-level coefficient equals $(1 + \gamma)\beta_k$. Thus, without sorting the social multiplier equals $1 + \gamma$ and with sorting the social multiplier equals $\frac{1 + \gamma}{1 + \sigma\gamma}$. Notice that pure input externalities significantly decrease the possibility of very large multipliers. This occurs because the feedback effects, which are the key to large multipliers in output-based externality models, are absent in this case.

When there is sorting on unobservables, the expectation of the aggregate coefficient equals $(1 + \gamma) \left(\beta_k + \lambda \frac{N(1 - \sigma)}{N(1 - \sigma) + \sigma} \right)$, and thus the social multiplier equals

$$\frac{1 + \gamma}{1 + \sigma\gamma} \left(1 + \frac{\lambda}{\beta_k} \frac{N\sigma}{N\sigma + 1 - \sigma} \right) \text{ which approaches } \frac{1 + \gamma}{1 + \sigma\gamma} \left(1 + \frac{\lambda}{\beta_k} \right) \text{ as } N \text{ gets large.}$$

Just as before, the measured social multiplier has the potential to look very large if there is significant sorting on the basis of unobservable characteristics.

In the case of the Dartmouth roommates, the key dependent variables are discrete and this requires further assumptions. Following Glaeser and Scheinkman (2002) (but not Brock and Durlauf, 2001), we will assume that observed outcomes are discrete, but that individual actions are still continuous. As such, each individual still chooses a level of “A”, but then this “A” is translated into a zero-one outcome. As such, each individual still chooses a level of “A”, and this continuous “A” is what influences neighbors, but the observable outcome takes on a value of one, if and only if $A > k$ for some fixed cutoff k .

The correct approach to this problem would be to postulate a normal or logistic distribution for ε and then to estimate the parameters using maximum likelihood. However, the purpose of this short paper is to give an easily usable method for calculating the relative size of social multipliers with simple calculation. As such, we proceed with the approximation that the probability of taking the action equals $\bar{p} + (A - \bar{A})$, where \bar{A} is the nationwide average level of A. Essentially, we are assuming that the distribution of the error term is approximately uniform, with density one, in the relevant region of estimation.⁵ In that case, the algebra describing the estimated coefficients is the same as in the continuous case, and we can use the previous discussion without alteration.

The Depreciation of Social Influence over Social Distance

This global interactions approach helps to make the point that a macro-coefficient does not necessarily imply much about a micro-relationship. Unfortunately, this simple approach does not help us to understand the degree to which interactions depreciate over space. As such, it gives us no guidance about what the impact of a policy evaluated on city-level data will have on the country as a whole. In order to have a framework which helps us understand the relationship between effects at different levels of aggregation, we

⁵ Assuming uniformity is, of course, a very strong assumption, but assuming that the density equals one conditional upon normality is innocuous, since A can always be rescaled so that this is true.

will need a model where the level of social influence changes with the degree of social proximity.

In order to present a simple model which captures the depreciation of social influence, we assume that $A_i = \theta_i + \gamma \sum_{d=1}^{\infty} \delta^{d-1} A_d^i$, where A_d^i is the action taken by the person who is exactly “d” units of social distance from the actor. We will think of people as being organized on a line, and a group of size N as including N people who are closest to one another on the line. Individuals could be located on a multidimensional lattice and could interact with any number of neighbors.⁶ We consider the simplest case of a line with unidirectional social influences.⁷ In other words, we assume that people are only influenced by people who are behind them in this line, i.e. person 2 follows person 1 and person 0, but person 1 only follows person zero.

This one-sided feedback ensures that the individual level regression yields an unbiased estimate of β_k (as long as there is no sorting). Aggregating yields the formula:

$$\frac{\sum_{i=1}^N A_i}{N} = \frac{((\gamma + \delta)A_0 - \delta\theta_0)}{N} \frac{1 - (\gamma + \delta)^N}{1 - \gamma - \delta} + \frac{1}{N} \sum_{i=1}^N \left(1 + \gamma \frac{1 - (\gamma + \delta)^{N-i}}{1 - \gamma - \delta} \right) \theta_i. \quad \text{We assume}$$

that $1 > \gamma + \delta$.⁸ The coefficient from an aggregate regression equals

$$\beta_k \left(1 + \gamma \left(\frac{1}{1 - \gamma - \delta} - \frac{1 - (\gamma + \delta)^N}{N(1 - \gamma - \delta)^2} \right) \right) \quad \text{and thus the social multiplier is}$$

$$1 + \gamma \left(\frac{1}{1 - \gamma - \delta} - \frac{1 - (\gamma + \delta)^N}{N(1 - \gamma - \delta)^2} \right), \quad \text{which converges to } \frac{1 - \delta}{1 - \gamma - \delta} \quad \text{as N gets large and}$$

equals $1 + \gamma/2$ when N equals 2.

III. Example # 1: Dartmouth Roommates

⁶ We have discussed social structure of this kind in our previous work, e.g. Glaeser and Scheinkman (2002).

⁷ If the structure is a bi-directional circle, then even determining $\partial A_i / \partial X_j^i$ is not straightforward—higher values of any individual “X” variable will have both a direct effect and an indirect effect through the influence of this X on the peers who then in turn influence the individual in question.

⁸ This assumption is necessary to guarantee that actions have finite variance.

In this section, we follow Sacerdote (2001) and look for the presence of social interactions among Dartmouth College roommates. The advantage of Dartmouth roommates is that they are essentially randomly assigned.⁹ As such, it provides one example of a situation where social connection is random and not the result of sorting. Thus, the social multiplier methodology seems most likely to be cleanly applicable in this case.

There are three natural units of aggregation within Dartmouth College: the room, the floor and the dormitory. The average room contains 2.3 students. The average floor contains eight students and the average dormitory contains twenty-seven students. Our “exogenous” variables are gender, verbal Scholastic Aptitude Test (SAT) score, math SAT score, high school grade point average (GPA), family income and a dummy variable that takes on a value of one if the individual drank beer in high school. The data on GPA, SAT scores and family income comes from the Dartmouth admissions department. The data on beer consumption comes from the Survey of Incoming Freshmen (sponsored by UCLA) which is filled out by thousands of entering college students.

We first examined the determinants of college GPA. We found no evidence of a social multiplier in this case (results not shown). The coefficients on individual level regressions were the same as the coefficients on aggregate regressions. For example, a 100 point increase in math SAT score raised freshman year GPA by .13 points in an individual level regression, .12 points in a room level regression and .10 points in a floor level regression. These results are not a surprise—Sacerdote (2001) also found that no influence of roommate background characteristics on freshman year grades.¹⁰ In this case, there is little evidence for social interactions or a social multiplier.

⁹ In fact, the assignment is only conditional within blocking group, where blocking groups are defined by answers to a pre-college survey. Sacerdote (2001) controls for this non-random element of assignment, but finds no effect of this control. For simplicity, therefore, we will ignore this minor element of selection.

¹⁰ Sacerdote (2001) does, however, find a correlation between the grade of two roommates which is some evidence for spillovers. As such, the actual magnitude of intra-room spillovers remains something of a puzzle.

We then turn to the area of fraternity or sorority membership. Fifty-one percent of Dartmouth undergraduates join a fraternity or sorority. Table 1 shows the results from estimating linear probability models in the case of fraternity membership. At the individual level, individuals who drank beer in high school are 10.4 percent more likely to join a fraternity or sorority. There are also more surprising individual level coefficients. Higher math scores increase fraternity membership—a 100 point increase in math SAT score leads to a 5 percent greater likelihood of joining a fraternity. Higher high school GPAs also increases the likelihood of joining a fraternity. People from richer families are also more likely to join a fraternity.¹¹ Men are more likely to join fraternities than women are to join sororities.

When we aggregate to the room level, the impact of drinking beer, gender and family income both increase slightly. The impact of GPA and math SAT score decline. None of these changes are statistically significant. Aggregating to the floor and then dormitory level causes beer drinking to become even more important (to a statistically and economically significant degree), but the other variables become insignificant. These regressions illustrate both the potential and the problems with social multiplier analysis. The coefficient on beer rises with the level of aggregation, just as the model predicts. The other coefficients just bounce around. As such, we will focus our analysis on the changes in the coefficient on past beer drinking.

If we use the global interaction model, we estimate different values of γ for each of the different levels of aggregation. The formula for the social multiplier is $\frac{N-1+\gamma}{(1-\gamma)(N-1)+\gamma}$. At the room level, the estimated social multiplier is less than one (although we can't reject small positive multipliers). At the floor level, the social multiplier 1.4 and the group size is eight. Together these imply that γ equals .38, which strikes us as a reasonable number. At the dormitory level, the estimated social multiplier is 2.23 and the group size is 57. These imply that the social multiplier equals .56. In fact,

¹¹ The magnitude of these effects are almost the same if we use a probit rather than a linear probability model.

our estimates are sufficiently imprecise that we cannot reject the null hypothesis that these two values are the same. Still, we find the general pattern of coefficients increasing with the level of aggregation.

Of course, logically, we expect the social multiplier to increase with the size of the group. Presumably, the bigger the group, the greater the share of social influences being included. Still, the global interactions model gives us little ability to actually interpret the extent to which the social multiplier changes with the level of aggregation. The local interactions model is meant to remedy this lack.

Using the local interactions formula for the social multiplier, and given an average floor

size of eight, this implies that $\gamma \left(\frac{1}{1-\gamma-\delta} - \frac{1-(\gamma+\delta)^8}{8(1-\gamma-\delta)^2} \right) = .44$ and the dormitory level

regressions tell us that $\gamma \left(\frac{1}{1-\gamma-\delta} - \frac{1-(\gamma+\delta)^{28}}{28(1-\gamma-\delta)^2} \right) = 1.27$. Together these equations

imply that: $\frac{20/28 - (\delta + \gamma)^8 + (8/28)(\delta + \gamma)^{28}}{7 - 8(\delta + \gamma) + (\delta + \gamma)^8} = 1.88$, which implies that $\delta + \gamma = .95$,

and using and hence $\gamma = .14$ and $\delta = .81$.

These numbers imply that each individual has only a small influence on his neighbor, but this influence depreciates quite slowly over time. The overall social multiplier is quite high, and as N gets large, it approaches 2.8. While the standard error bands surrounding our estimates are sufficiently large to make us quite cautious about accepting these numbers, they still suggest that the methodology does provide estimates that are at least plausible.

IV. Example # 2: Crime

In our previous work (Glaeser, Sacerdote and Scheinkman, 1996), we have focused on social interactions in criminal behavior. There is a large amount of anecdotal behavior supporting the existence of these interactions, and it seems reasonable to expect to find a

social multiplier in the level of crime. As we have no data set featuring randomized interactions in this context, we will have to use existing data on the level of crime to produce preliminary estimates of the social multiplier in criminal behavior.

Individual level crime rates do not exactly exist. There are data on people who are arrested and people who go to prison. And there is self-reported data on criminal behavior. Self-reported data is problematic for two reasons. First, people do not always report their illegal activities honestly. Second, standard self-reported information on criminal behavioral (e.g. the National Longitudinal Survey of Youth) does not contain crime data that is closely comparable to information about crime rates.

Because of these problems, we used nationwide arrest rates by age to form our basic individual level estimates. These data correspond to arrests, not crimes. In order to make these two sets of numbers comparable, we multiply the age-specific arrest rate by the national ratio of reported crimes to arrests. In other words, we ensure that at the national level, our predicted crime measure is the same as the actual crime level. Nonetheless, our use of arrest rates will be problematic if the ratio of crimes to arrests differs across age categories. Still, because this work is meant to be exploratory, we will go ahead with this information. A further issue is that since our only independent variable is age, we may miss many possible sources of strategic complementarities in the level of crime.

These individual crime rates provide us with a predicted level of crime in each neighborhood. As described above, we use the individual level coefficients to predict an aggregate crime measure, i.e. $\sum_a \pi(a)p(a)$, where $\pi(a)$ represents the arrest rate in each age category “a” and $p(a)$ represents the share of the population in that age category. We ignore any issues that might come from aggregating a discrete variable.

In Table II regression (1), we report the results from regressing actual crime rates on predicted crime rates at the county level. The coefficient is 1.72. In regression (2), we show that the state level social multiplier is 2.8. In principle, both of these numbers might be biased because of sorting on observables or unobservables. In fact, the sorting

by age across counties and states is quite low. Sorting on unobservables (or observable variables that are not included in the regression) is likely to be higher, and as a result it makes sense to take these results warily, as they may well overstate the true social multiplier.

Finally, in regressions (3) and (4), we look at the social multiplier that is estimated using time series data at the nation level. In these regressions, we follow Levitt (1999) closely and in a sense merely duplicate his evidence showing that, if micro-level coefficients are used, then aggregate changes in demographics can explain little of the changes in aggregate crime. In our framework, this observation shows itself in an estimated social multiplier of 8.16 for crime as a whole and 4.47 for homicides. These high social multipliers tell us that crime rates are moving around very quickly, given the fairly modest changes in aggregate demographic compositions. We have our doubts about the interpretation of these estimates. It is at least as likely that these high estimates are due to a correlation between demographics and unobservable elements. Still, their high values continue to provide some evidence that social interactions are important in the level of crime and more generally that social multipliers are worth worrying about.

We have estimated three different social multipliers at different levels of aggregation. The estimated social multiplier rises substantially with the level of aggregation, so one might think that the exponential model could be useful in interpreting this data. However, the exponential model is actually pretty hard to use when addressing such large aggregations. To make the point, the difference between state and county level

coefficients implies that $\frac{.9834 - (\delta + \gamma)^{86,700} + .0166(\delta + \gamma)^{5,207,000}}{86,699(1 - \delta - \gamma) + (\delta + \gamma)^{86,700}} = .622$, which implies

that $\delta + \gamma > .999$. However, if $\delta + \gamma > .999$, then the estimated social multiplier at the county level requires that $\gamma < .001$. The basic problem is that with an exponential model, there should be little social interactions beyond close neighbors and as such, differences between county, state and nation-wide social multipliers are difficult to work with. Future work will hopefully come up with a better model for addressing the depreciation of social influence in large groups.

V. Example # 3: Schooling and Earnings

In the case of schooling and earnings, we turn to the variant of the model where spillovers occur across θ s (i.e. inputs) not A s (i.e. outputs). This framework corresponds more closely to the idea that individual earnings are a function both of their own schooling and of the schooling of their neighbors. In this case, the social multiplier should equal $\frac{1+\gamma}{1+\sigma\gamma}\left(1+\frac{\lambda}{\beta_k}\right)$. We will use the approach of regressing aggregate outcomes on predicted aggregate outcomes. We will use the Individual Public Use Micro-Sample from 1990, and include all adults between 18 and 60 years of age.

Our individual coefficients are found by regressing individual wages (in levels) on gender and race dummies, marital status, a third order polynomial in education, and a fourth order polynomial in age. All of our coefficients in this first stage regression looked quite standard and we don't report them to save space.

In Table III, we report our results from aggregating wages and predicted wages up to the Public Use Microsample Area (PUMA) and State level. These are the two levels of geography that are available in the 1990 census. PUMAs on average have 82,800 members. The average state has 2.8 million members.

In regression (1), we find a PUMA level social multiplier of 1.675. In regression (2), we find a State level social multiplier of 2.172. As we would expect, the social multiplier rises with the level of aggregation. However, just as in the case of crime, the exponential model is hard to use with aggregations of this size. These social multipliers may be biased upwards because of sorting on unobservables; however, sorting on observables at least is stronger at the PUMA than at the state level. Still, these different coefficients should stand as a warning against using coefficients from one level of aggregation to inform us of effects at a different level of aggregation.

Furthermore, these results continue to suggest that there are human capital spillovers, as suggested by a wide body of other research (e.g. Rauch, 1993, Lucas, 1988, Acemoglu and Angrist, 1999). However, the results from these regressions imply a much larger human capital spillover than the previous research. To us, this discrepancy serves to emphasize the fact that unobserved heterogeneity may be driving our results.

VI. Conclusion

Often empirical work treats the level of aggregation as irrelevant. Routinely, state or national policy interventions are used to infer underlying individual-level parameters in contexts as diverse as labor supply or the returns to schooling. If positive spillovers or strategic complementarities exist, then these forms of inference are improper. State-level regressions yield appropriate answers to questions about state-level policies, but not necessarily anything else. The existence of a social multiplier means that in many contexts, aggregate level coefficients will tend to radically overstate the true individual level response.

This paper has presented a brief analysis of the social multiplier. We presented a series of simple models, all of which tell us how to infer social interactions variables from the level of the estimated social multiplier. In principle, these models can be computed efficiently by using maximum likelihood, but in many contexts, an unbiased measure of the social multiplier can be estimated by comparing ordinary least squares coefficients found at different levels of aggregation.

In the empirical sections of the paper, we found evidence for a social multiplier at three different levels of aggregation. Using Dartmouth roommates data, where roommates are randomized, we found that the impact of at least one predetermined variable had a bigger impact on joining a fraternity or sorority at higher levels of aggregation. In this case, our results were compatible with our model of exponentially declining social influence. Using crime data, we found evidence for a very large social multiplier in the level of crime. We do not necessarily take the estimates as being precise, but they are large

enough to support the idea that a social multiplier exists. Finally, using data on wages and human capital variables, we found further evidence for large social multipliers in the case of wages and human capital. The pattern supports the idea that researchers need to be careful about how social interactions can potentially make inference very difficult, especially when state level variation is used as the source of identification.

References

- Acemoglu, D. and J. Angrist (1999) "How Large are the Social Returns to Education? Evidence from Compulsory Schooling Laws," NBER Working Paper # 7444.
- Becker, G. and K. Murphy (2000) *Social Economics*. Cambridge: Harvard University Press.
- Brock, W. and S. Durlauf (2001) "Discrete Choice with Social Interactions," *Review of Economic Studies* 68(2): 235-260.
- Glaeser, E. and J. Scheinkman (2001) "Non-Market Interactions," *Advances in Economic Theory*, forthcoming.
- Glaeser, E., Sacerdote, B. and J. Scheinkman (1996) "Crime and Social Interactions," *Quarterly Journal of Economics* CXI (2): 507-548.
- Goldin, C. and L. Katz (2002) "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions," *Journal of Political Economy*, forthcoming.
- Katz, L., Kling, J. and J. Leibman (2001) "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment," *Quarterly Journal of Economics* CXVI (2): 607-654.
- Levitt, S. (1999) "The Limited Role of Changing Age Structure in Explaining Aggregate Crime Rates," *Criminology* 37(3): 581-597.
- Lucas, R. (1988) "On the Mechanics of Economic Development," *Journal of Monetary Economics* 22(1): 3-42.
- Ludwig, J., Hirschfeld, P. and G. Duncan (2001) "Urban Poverty and Juvenile Crime: Evidence from a Randomized Housing-Mobility Experiment," *Quarterly Journal of Economics*, CXVI (2): 665-679.
- Manski, C. (1993) "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies* 60(3): 531-542.
- Rauch, J. (1993) "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities," *Journal of Urban Economics* 34(3): 380-400.
- Sacerdote, B. (2001) "Peer Effects With Random Assignment: Results for Dartmouth Roommates" *Quarterly Journal of Economics* CXVI (2): 681-704.

Table I
Social Multipliers in Fraternity Participation

**Dartmouth Roommate Data: Effect of Background Characteristics on Participation
in Fraternities at the Individual Level and Three Levels of Aggregation**

Column (1) shows the OLS regression of individual fraternity participation on own use of beer in high school, own SAT scores, own high school GPA and own family income (self reported). Column (2) regresses the average participation at the dorm room level on dorm room averages of high school beer use, SAT scores, HS GPAs, and family income. Columns (3) and (4) increase the level of aggregation to the dorm floor and dorm building respectively.

	(1)	(2)	(3)	(4)
	Member of fraternity or sorority	Room average level membership	Floor average membership	Dorm average membership
Drank beer in high school	0.1040 (0.0258)	0.0984 (0.0399)	0.1454 (0.0812)	0.2320 (0.1930)
Male	0.0510 (0.0256)	0.0701 (0.0286)	0.0253 (0.0540)	-0.2066 (0.2038)
SAT verbal score	-0.0001 (0.0002)	-0.0000 (0.0003)	-0.0000 (0.0006)	-0.0002 (0.0011)
SAT math score	0.0005 (0.0002)	0.0002 (0.0003)	-0.0006 (0.0006)	-0.0022 (0.0014)
High school GPA	0.0004 (0.0001)	0.0003 (0.0002)	0.0003 (0.0003)	0.0004 (0.0005)
Family Income '000	0.0006 (0.0002)	0.0008 (0.0003)	0.0000 (0.0006)	-0.0004 (0.0013)
Constant	0.0482 (0.1455)	0.1980 (0.2266)	0.7993 (0.4594)	2.2277 (1.1421)
R-squared	0.04	0.05	0.03	0.08
Observations	1579	700	197	57
Average group size	1	2.3	8.0	28

Notes: Data are for Dartmouth Freshmen. Roommates and dormmates are randomly assigned as described in Sacerdote [2001]. SAT scores are from Dartmouth Admissions data. Family income, use of beer, and high school GPA are self reported on the UCLA Higher Education Research Institute's Survey of Incoming Freshmen. Standard errors in parentheses.

Table II**Regression of Crimes Rates on Predicted Crime Rates**

Predicted crime rates for counties (or states or US) are formed by multiplying percentage of persons in each of eight age categories by the crime rate for persons in that age category. Data are from Census Bureau and Uniform Crime Reports. Expected crime rate conditional on age is based on age distribution of arrestees for the U.S.

Columns (1)-(4) are cross sectional and crime data are for 1994 and demographic (age) data are for 1990. Columns (5) and (6) are the time series data for the US as a whole.

	(1)	(2)	(3)	(4)
	County	State	US Crime	US
	Crime	Crime	Rate	Homicide
	Rate	Rate	1960-1999	Rate
				1960-1999
Predicted crime rate (or homicides)	1.732 (0.088)	2.811 (1.070)	8.163 (0.998)	4.467 (0.637)
Constant	-0.039 (0.004)	-0.078 (0.045)	-0.304 (0.043)	-0.000 (0.000)
R-squared	.12	.13	0.64	0.56
Observations	2756	50	40	40
Average Group Size	86,700	5,207,000	226,275,000	226,275,000

Table III

Regression of Wages on Predicted Wages

Predicted wages are formed by regressing individual level wages on gender, race dummies, marital status, education, education squared and education cubed, age, age squared, age cubed and age to the fourth power. We then aggregate to the PUMA (state) level and regress mean wages on the mean of predicted wages.

	(1) PUMA Mean Wages	(2) State Mean Wages
Predicted wages	1.675 (.0270)	2.172 (.246)
Constant	-2227.140 (301.170)	-8250.57 (2675.232)
R-squared	.69	.61
Observations	1726	51
Average group size	82,800	2,802,000