Measuring Peer Group Effects: A Study of Teenage Behavior

William N. Evans

University of Maryland

Wallace E. Oates

University of Maryland and Resources for the Future

Robert M. Schwab

University of Maryland

Individuals or households often have some scope for choice of peer groups, whether through the selection of neighborhood of residence, school, or friends. This study addresses the estimation of peer group effects in cases in which measures of peer group influence are potentially endogenous variables. Using a rich data set on individual behavior, the paper explores teenage pregnancy and school dropout behavior. For both cases, the estimation of a straightforward single-equation model yields statistically significant peer group effects; however, these effects disappear under simultaneous equation estimation. The results are robust and suggest the need for careful modeling of the choice of peer groups.

I. Introduction

There is a large empirical literature that seeks to measure the importance of peer group influence in determining the performance of the

We thank the National Science Foundation for its support of this research. We also thank Jonathan Dunn, Andrew Kochera, and Robin Jenkins for their excellent work as our research assistants; the Guttmacher Institute for making the data on family planning clinics we have used in this paper available to us; and Richard Arnott, Christopher Jencks, Robert Plotnick, and Joan Kahn for their very helpful comments and suggestions.

[*Journal of Political Economy*, 1992, vol. 100, no. 5]

individual members of the group. In education, Summers and Wolfe (1977) and Henderson, Mieszkowski, and Sauvageau (1978), for example, have found that, other things equal, students perform at a higher level if their fellow students are high achievers. Their findings provide additional support for the strong claims concerning peer group influence that were part of the famous Coleman report (Coleman et al. 1966) on educational achievement.

Some recent papers in the theory of local finance, including Brueckner and Lee (1989), de Bartolome (1990), and Schwab and Oates (1991), have explored this issue in a broader context. Drawing on the conceptual framework developed in Bradford, Malt, and Oates (1969), these papers take the levels of final outputs of various local public goods to depend not only on local budgetary inputs but also on the characteristics of the people who reside in the community. There is compelling evidence, for instance, that the level of public safety in a particular neighborhood depends not so much on the frequency of police patrols as on the propensity of local residents to engage in illegal activities. The argument, more formally, is that the "production function" for such local outputs includes *both* the standard inputs such as labor and capital (transformed into police patrols, for example) *and* the characteristics of the local population itself.

The basic issue we raise in this paper concerns the testing and measurement of these peer group (or neighborhood) effects. The studies that have looked at this issue typically begin with a "standard" model that seeks to explain "output" (as measured by standardized test scores, for example) in terms of a vector of student and family characteristics, a vector of "public" inputs (measures of pupil/teacher ratios, the training and experience of the teacher, the availability of certain special educational services, etc.), and a vector of variables that summarize the characteristics of the other students in the class. The estimates of the coefficients on this third set of variables and their standard errors are then used to test for the presence of peer group effects and to measure their magnitude. Many studies find that these peer group effects exist and are quite important.[1]

Our concern is that the "peer group" (or, more generally, the set of neighbors or local residents) is often itself a matter of individual choice. The extreme case of the Tiebout (1956) model of local finance makes the point. In a Tiebout world, individual households search costlessly among a wide variety of local communities and select as a jurisdiction of residence a locality that provides a menu of local outputs that best fits their tastes. Although the original Tiebout model did not address the issue of peer group effects, the later work cited

[1] See Hanushek (1986) for an excellent review of the education literature.

above has incorporated these effects into the analytical framework for local public goods. What is important for our purposes is the implication of these later models that individual households, in making their locational choice, are also choosing a peer group for many of the relevant local services, including local public schools. In such a setting, the peer group becomes an endogenous variable, determined in part by household choice. Once this is recognized, it is clear that the estimation of the "standard model" by ordinary least squares or other techniques that do not allow for the endogeneity of the peer group are inappropriate.

The direction of the bias introduced by ignoring simultaneity depends on the relationship between the unobserved factors that determine the peer group and the unobservable factors that determine performance. For the problems this literature has examined, we suspect that simple models are likely to overstate peer group effects. To see this in the context of the education example, consider a child whose parents devote great effort to the welfare of their children. We might reasonably expect to find that this child does well in school for two reasons. First, his parents will see that he attends a school in which the peer group is "better" than expected given the family's observed characteristics. Second, he will do well in school as a result of factors that cannot be observed but that are under his parents' control, such as the time the parents spend with the child. A single-equation model, however, would mistakenly attribute this entire increment to the child's performance to his superior peer group.[2]

The purpose of this paper is to explore these issues at both the conceptual and empirical levels. To this end, we choose to examine peer group effects on teenage pregnancy and the decision to drop out of school. Teenage pregnancy is particularly interesting in this context in that peer group effects are often thought to be very important. As one study notes, "probably the agent most 'blamed' for increases in teen sexual activity over the last decade has been the peer group" (Hofferth 1987, p. 27). Mayer (1991), for example, found that the socioeconomic status of the students in a girl's high school is a key determinant of the probability that she will bear a child while still a teen. But other studies of teenage pregnancy are less clear on this matter, suggesting that peer influences may have been "overrated as a source of increased sexual activity among teenagers" (Hofferth 1987, p. 27). As Jencks and Mayer (1990) point out, the effect of peer

[2] Few papers have looked at the possible endogeneity of the peer group. Case and Katz (1990) find that neighborhood effects are important for socioeconomic outcomes. They argue that their results rule out the possibility that the neighborhood effects they find are just an artifact of the way families sort themselves into communities.

groups on the likelihood of graduating from high school is more problematic, but existing evidence provides some support for the view that the socioeconomic status of schoolmates is inversely related to the probability that a student will drop out of high school.

The paper is structured as follows. Section II describes our data set, and Section III sets forth the estimates of a single-equation model that assumes that the peer group is exogenous. In Section IV we present an expanded model that treats the peer group as a choice variable, we describe the estimation techniques we have used, and we summarize our estimates of this model for teenage pregnancy. Section V extends the econometric analysis to examine the case of school dropouts. Section VI examines some alternative specifications of our model, and Section VII presents some concluding reflections on our findings.

We find that single-equation and simultaneous equation models yield very different pictures of the importance of peer group effects. In the standard model presented in Section III, we find a moderate and statistically significant peer group effect on teenage pregnancy. In the expanded model presented in Section IV, this effect disappears completely! Our results on school dropouts in Section V follow a similar pattern. The standard model estimates, in short, indicate a significant peer group effect that vanishes in the expanded model.

Our results must be interpreted carefully. We are certainly not arguing that peer group effects are unimportant, but rather that these effects are to some extent under the control of the individuals who choose to become members of a particular group. This endogeneity must be embodied in the econometric models we use if we are to obtain reliable estimates of these effects. If we fail to do so, we may mistakenly attribute part of the effect of unobservable family characteristics to the effect of the peer group. The potential for this type of bias is present in virtually all research on the effects of the peer group; in our study, the magnitude of the bias is very large indeed.

In the course of this study, we have come in contact with a substantial body of empirical literature cutting across the various social sciences that attempts to measure and assess the importance of neighborhood or peer group effects. This literature addresses a wide range of behavior including school performance, crime, labor market performance, and sexual activity. Some of this research finds that such neighborhood effects are important determinants of individual behavior. Crane (1991), for example, develops an "epidemic" theory of social behavior for which he finds supporting evidence from patterns of school and sexual behavior in ghetto neighborhoods. Such evidence is consistent with the emerging view of an urban "underclass" that is trapped in poor and geographically isolated neighborhoods (Wilson 1987).[3]

However, the findings on the importance of peer effects in this literature are mixed. Jencks and Mayer (1990) provide an excellent survey of this work in which they both describe the relevant studies and provide an insightful evaluation of this body of research. We shall refer to some of the specific findings of individual studies at the relevant points in our paper. But certain threads in this work are worth noting here. First, it is hard to reach *any* broad generalizations from the studies of neighborhood and peer group effects. As Jencks and Mayer put it, "Our first and strongest conclusion is that there is no general pattern of neighborhood or school effects that recurs across all outcomes" (p. 174). It is interesting, however, that for the wide range of behavior that their survey encompasses, they find that "despite its limitations, the available evidence suggests to us that neighborhood and classmates probably have a stronger (or at least more consistent) effect on sexual behavior than on cognitive skills, school enrollment decisions, or even criminal activity" (p. 167).

Second, although the importance of neighborhood and peer group effects varies widely across these studies, family background traits nearly always have powerful effects on individual behavior: they typically dwarf the peer group effects in importance. And this, as we shall see, is what we find as well. Jencks and Mayer note in this regard that "social scientists need to be very cautious about estimates of neighborhood or school effects that control only one or two family background characteristics. As a rule, the more aspects of family background we control, the smaller neighborhood and school effects look" (p. 176). Our study reinforces this conclusion.

II. Data

Most of the data for our study were drawn from the National Longitudinal Study of Youth (NLSY). The full NLSY sample includes 12,686 people (6,403 males and 6,283 females) who were between 14 and 21 years old when the survey began in 1979. The NLSY is wide ranging, providing information on such matters as labor market experience, family background, fertility history, and drug and alcohol abuse. It is particularly well suited for our purposes because it included in 1979 a survey of the last secondary school each respondent attended. This survey asked school officials a number of questions

[3] Our study does not directly address the underclass issue inasmuch as ghetto neighborhoods are not the focus of our model or sample. Such a study requires a framework and specifications that allow for sharp, discontinuous effects among different types of neighborhoods (e.g., Crane 1991).

about the composition of the student body. We use the responses to these questions to construct our peer group variables.

We relied on several criteria to select the sample for our study from the larger NLSY sample. First, we eliminated all cases in which the school data were not available.[4] Second, we considered only women who were aged 19 or younger when the NLSY began. Most of the women who were 20 or 21 years of age left high school 3 or 4 years before the school survey was taken in 1979. Third, we included only teenagers who lived in standard metropolitan statistical areas (SMSAs). Fourth, we excluded all NLSY respondents who became pregnant before they enrolled in their 1979 school.

Our final sample included 1,453 cases.[5] Table 1 presents definitions and summary statistics for most of the important variables we have used in this study. PREGNANT is a dichotomous variable that has a value of one if the respondent became pregnant during her teen years. Underreporting is often a problem in pregnancy studies; we suspect that it may be a particular problem in the NLSY since in a large majority of the cases at least one parent was present during the interview. Mott (1983) notes that underreporting in the NLSY appears to be greater among blacks than whites. In our sample, 22.5 percent of the white respondents and 38.2 percent of the black respondents said that they had become pregnant at least once before they were 20 (PREGNANT = 1). Hofferth (1987, p. 420) notes that studies that do not rely on survey methods find that 39.7 percent of whites and 63.1 percent of blacks have teenage pregnancies.[6]

Underreporting is thus a real concern as a source of measurement error in our dependent variable. At the same time, teenage pregnancy is an appealing subject for this study since, as we noted earlier, it is one for which peer group effects are thought to be especially important. Our focus on pregnancies, rather than births, has a further advantage. The decision to carry a pregnancy to term is complicated by a host of variables other than the peer group, many of which are also potentially endogenous. For example, Grossman and Joyce (1990) find that the resolution of a pregnancy is in part a function of

[4] School data were more likely to be available for teenagers from wealthier families, and we have therefore opened up the possibility of some sample selection bias by including only those cases in which the school responded to the survey.

[5] Our sample includes respondents from the cross-section NLSY sample that is intended to be representative of all youths 14–21 as well as the supplemental NLSY sample designed to oversample Hispanic, black, and disadvantaged youths. We did not use the NLSY sample weights in this paper.

[6] These numbers were computed with 1981 data by Jacqueline Forrest for the National Academy of Science's report on teenage pregnancy, *Risking the Future.* She bases her estimates on counts of first births reported in the *Vital Statistics of the United States* and abortion counts tabulated by the Alan Guttmacher Institute.

TABLE 1

VARIABLE DEFINITIONS AND SAMPLE CHARACTERISTICS

Variable Name	Definition	Mean and Standard Deviation
PREGNANT	0–1 dummy variable that equals one if student became pregnant before age 20	.28 (.45)
BLACK	0–1 dummy variable that equals one if student is black	.22 (.41)
SEX EDUCATION	0–1 dummy variable that equals one if student had sex education course	.70 (.46)
RELIGIOUS	0–1 dummy variable that equals one if student attends religious services at least four times a month	.62 (.47)
FEMALE HH	0–1 dummy variable that equals one if student's family is a female-headed household	.17 (.37)
STEPFATHER	0–1 dummy variable that equals one if student's family has a stepfather present	.05 (.23)
OTHER FAMILY	0–1 dummy variable that equals one if student's family is some other structure	.06 (.24)
SIBLINGS	Number of siblings in student's family	3.58 (2.45)
QUARTILE 2	0–1 dummy variable that equals one if student's family income is within second quartile of sample incomes	.25 (.44)
QUARTILE 3	0–1 dummy variable that equals one if student's family income is within third quartile of sample incomes	.26 (.43)
QUARTILE 4	0–1 dummy variable that equals one if student's family income is within fourth quartile of sample incomes	.24 (.43)
DISADVANTAGED	Percentage of students in school that are disadvantaged	22.25 (23.90)
MOTHER DROPOUT	0–1 dummy variable that equals one if the student's mother dropped out of high school	.40 (.49)
MOTHER HIGH SCHOOL	0–1 dummy variable that equals one if the student's mother has 12 years of education	.40 (.49)
MOTHER SOME COLLEGE	0–1 dummy variable that equals one if the student's mother has more than 12 but fewer than 16 years of education	.11 (.32)
AFDC2	State AFDC payment for a family of two, in 1979	235.34 (88.03)
LICENSE	0–1 dummy variable that equals one if state requires a license to sell contraceptives	.14 (.34)
CATHOLIC	0–1 dummy variable that equals one if student's religious affiliation is Catholic	.39 (.49)

TABLE 1 (*Continued*)

Variable Name	Definition	Mean and Standard Deviation
PROTESTANT	0–1 dummy variable that equals one if student's religious affiliation is Protestant	.42 (.49)
OTHER RELIGION	0–1 dummy variable that equals one if student's religious affiliation is some other religion	.11 (.31)
TEACHER/PUPIL	Ratio of teachers to pupils in student's school	.05 (.02)
CLINICS	Ratio of clinics in county to the number of teens at risk in county	.75 (.59)
N	Number of observations	1,453

such variables as medicaid funding for abortions, the availability of abortion clinics, and the presence of a spouse. Births are one more step removed from the variable that is of primary interest.[7]

Rather than abandon our dependent variable, we have responded to this issue by bringing additional evidence to bear on the problem. More specifically, we have employed three different kinds of "tests" to ensure that the results from the overall pregnancy equations are not misleading: (1) We repeated much of the analysis we report here using actual births, rather than pregnancy, as the dependent variable. This avoids some of the underreporting issues: the fraction of teens in the NLSY who report that they gave birth to a child before reaching age 20 is roughly consistent with national estimates. Our findings based on this second dependent variable are similar, as we shall see, to our results using the pregnancy variable. (2) We divided our sample by race and estimated separate equations for blacks and whites, since underreporting is apparently less pervasive for the latter group. (3) We looked at a wholly different form of behavior: the decision to stay in school or drop out, for which underreporting is, we believe, unimportant. As we shall see, the results for our dropout equations are much like those for the pregnancy equations. The results from these three additional exercises give us some confidence that measurement error in the dependent variable is not seriously distorting our findings.

The conceptual framework we presented in the Introduction sug-

[7] Christopher Jencks has taken issue with us on this matter. In some written comments to us, he points out interestingly that "many sociologists actually make the opposite argument, suggesting that pregnancies occur because of ignorance and hormones, but that the decision to carry a pregnancy to term rather than abort is highly social."

gests that we should consider three sets of determinants of teenage behavior: individual and family characteristics, publicly provided inputs, and peer group effects. The literature on teenage pregnancy has examined peer group effects at several different levels, including the influence of community characteristics, peers at school, and close friends.[8] Our measure of the peer group effect, LOG(DISADVANTAGED), is the natural log of the percentage of students in the respondent's school who were classified as economically disadvantaged.[9] There is substantial evidence that teens from low-income families are more likely to be sexually active and to become pregnant. This paper asks whether this behavior and the attitudes that lead to this behavior have important effects on other students at their school.[10]

We have included a number of family characteristics in the study.[11] The variables FEMALE HH, STEPFATHER, and OTHER FAMILY are a set of dummy variables that describe family structure; the suppressed category includes families in which both biological parents are present. Similarly, MOTHER DROPOUT, MOTHER HIGH

[8] Hogan, Astone, and Kitagawa (1985) concluded that black girls 15–19 years old living in a poverty area of Chicago were much more likely to be sexually active than black teens living outside poverty areas. Furstenberg et al. (1987) found that blacks who attended a segregated school were 13 times more likely to have had sexual intercourse than whites from segregated schools but that the differences between blacks and whites from integrated schools were much smaller. Billy and Udry (1985) found that friends' sexual attitudes and behavior were important elements in decisions of white teenage girls to have sexual intercourse.

[9] Schools were asked to estimate the percentage of students who were disadvantaged under Elementary and Secondary Education Act (ESEA) guidelines. The federal government disperses ESEA funds to states to assist areas with high concentrations of low-income families. States may distribute funds to districts according to a range of different criteria, and at the school level it is very difficult to determine the financial characteristics of students' families. According to the U.S. Department of Education, the best and most widely used measure of economic disadvantage at the school level is the number of students participating in the school lunch program; these data would always be available to school principals.

[10] We would have liked to look at more direct measures of peer group effects such as the pregnancy rate within the school. High School and Beyond is the only data set of which we are aware that would allow us to do so. We cannot use High School and Beyond because it is not possible to identify a respondent's city or metropolitan area in that data set, and, as we explain below, we require that information in order to estimate our simultaneous equation model.

[11] We suspect that the average number of siblings in our sample is so large (though close to the average in High School and Beyond) for several reasons. First, the average number of siblings per child is always greater than the average number of children in families with children (minus one). Suppose that all families had either one child or nine children and that there were equal numbers of both types of families. On average, children would have 7.2 siblings but the average family would have only five children. Second, as we noted above, our sample includes respondents from the NLSY cross-section sample as well as the supplemental NLSY sample of Hispanic, black, and disadvantaged youth. Catholics are overrepresented in our sample for the same reasons; in addition, we restricted our sample to NLSY respondents who lived in SMSAs, and Catholics are more likely to live in urban areas.

SCHOOL, and MOTHER SOME COLLEGE represent mother's education; the suppressed category is college graduate. We also include a measure of religious attendance and the number of siblings as explanatory variables in light of the results of past research.[12]

The reported measure of family income in the NLSY is troubling. Income was not reported for 26 percent of the cases in our sample, either because the respondent did not answer this question or because at the time of the interview the respondent was living outside of her parents' home and her answer therefore reflected her own income. We used data for those respondents in our sample who did report income to estimate an equation that described income in terms of other household characteristics, and then we used the estimated values from this equation to impute the missing values for the income variable.[13] This introduces some measurement error into our estimates. We tried to address this issue by defining a set of dummy variables (QUARTILE 2, QUARTILE 3, and QUARTILE 4) that together characterize a respondent's place in the income distribution. This approach reduces measurement error because we are unlikely to place a family in the wrong quartile even if we cannot estimate its income precisely. The procedure comes at some cost, though, since it suppresses information by treating alike all families within a quartile. Testing suggests that our basic results are not sensitive to our procedure for constructing the income variable.

We have included a number of measures of publicly provided inputs. Education variables include whether or not a respondent has taken a sex education course and the teacher/pupil ratio in the respondent's school. We also consider the availability of family planning clinics, contraceptive laws, and the level of AFDC payments.

III. Single-Equation Estimation

In this section we treat peer group effects as an exogenous variable in a single-equation probit model. This section is thus a direct parallel to the work on peer group effects in the education literature. Our model can be described as follows. Let y_1^* represent a teenager's unobservable propensity to become pregnant. We assume that y_1^* is a linear function of observed variables such as family characteristics and pub-

[12] See Lundberg and Plotnick (1990) for a review of this evidence.

[13] This approach will lead to misleading results if, when observables are held constant, those who did not respond to the income question systematically have higher or lower incomes than those who did. We tried to address this question by estimating a model suggested by Griliches, Hall, and Hausman (1978) in which we treated nonreporting as an endogenous variable. The results were disappointing because the model did a very poor job of predicting the decision not to respond. We therefore fell back to the straightforward solution described above.

licly provided inputs $\mathbf{x}_1$, a scalar measure of peer group effects y_2, and an error term ϵ_1 that is $N(0, 1)$:

$$y_1^* = \mathbf{x}_1\beta_1 + y_2\beta_2 + \epsilon_1. \tag{1}$$

Let y_1 equal one if the student becomes pregnant during her teen years. It is defined by

$$\begin{aligned} y_1 &= 1 \quad \text{if } y_1^* \geq 0 \\ &= 0 \quad \text{if } y_1^* < 0. \end{aligned} \tag{2}$$

Then the probability that a teenager will become pregnant is

$$\begin{aligned} \text{prob}(y_1 = 1) &= \text{prob}(\epsilon_1 \geq -\mathbf{x}_1\beta_1 - y_2\beta_2) \\ &= 1 - \Phi(-\mathbf{x}_1\beta_1 - y_2\beta_2) \\ &= \Phi(\mathbf{x}_1\beta_1 + y_2\beta_2), \end{aligned} \tag{3}$$

where Φ is the cumulative distribution function of the standard normal distribution.

Estimates of alternative versions of this single-equation model are set forth in table 2. Model 1 in column 1 is our basic model. The key variable in these models is LOG(DISADVANTAGED), our measure of the peer group effect. This variable is significant and has the expected sign; everything else equal, placing a teenage girl in a school in which a higher proportion of the students are disadvantaged increases the probability that she will become pregnant.

We have included in column 2 of table 2 the derivative of the probability of pregnancy with respect to each variable and the associated standard error of the estimate.[14] Consider a teenage girl who moves from a school with the mean percentage disadvantaged to a school in which the percentage of students who are disadvantaged is 25 percentage points higher; this implies that LOG(DISADVANTAGED) rises by .75. Our results indicate that the probability that she will become pregnant would rise by 1.7 percentage points. The magnitude of the response is roughly equivalent to the addition of 1.5 more siblings to the household.

While these estimates suggest that the peer group effect is a statistically significant and nontrivial determinant of teenage pregnancy, its importance is small in comparison with variables describing family structure. Teenagers from households in which at least one biological

[14] All these derivatives are calculated for a white teen who attends church regularly (RELIGION = 1), has taken a sex education course, and lives with both parents and whose mother was graduated from high school but did not attend college, whose family is in the third income quartile, and whose high school has the mean percentage of disadvantaged students.

parent is absent have a 7–11-percentage-point greater chance of becoming pregnant than their counterparts who have both natural parents present. To take another instance, a teen whose mother was graduated from high school but did not attend college faces a 10-percentage-point higher chance of becoming pregnant in comparison with an identical teen whose mother was a college graduate.

Although the race variable is significant only at the 10 percent level, we suspect that this variable may understate racial differences given the discussion of underreporting presented above. With all other factors such as education and family structure held constant, income seems to have no impact on the probability of pregnancy. All the results above are consistent with the literature (see Hofferth 1987; Lundberg and Plotnick 1990).

Model 1 also points to the important role that sex education courses can play in reducing teen pregnancies. The estimates in column 2 of table 2 imply that providing a teenager with a course in sex education reduces the probability that she will become pregnant by 5.8 percentage points. This is not a trivial effect; it suggests, for example, that a teen from a female-headed household who has taken a sex education course would have roughly the same probability of becoming pregnant as a teen who came from a household in which both biological parents were present but who had not received sex education.

In the remaining columns of table 2, some extensions and alternatives to our basic model are presented. Model 2 includes a set of dummy variables for religious affiliation (the suppressed category is teens with no religious affiliation). None of these dummies is significant, and a log likelihood test shows that jointly they add no explanatory power to the basic model. Lundberg and Plotnick (1990) find that teens are more likely to become pregnant and give birth if welfare benefits are high or if it is difficult to obtain contraceptives. Model 3 includes the natural log of AFDC payments for a family of two and a dummy variable, LICENSE, that has a value of one if the respondent lives in a state in which a merchant must have a license to sell contraceptives. Model 4 includes two additional measures of publicly provided inputs: CLINICS (the number of family planning clinics per teenager in the respondent's county) and TEACHER/PUPIL (the teacher/pupil ratio in the respondent's school). In all these alternative versions of our basic model, none of the additional variables is statistically significant, and the estimates of the coefficients on the variables in the basic model are very similar to the estimates for model 1. In particular, the estimated coefficient of LOG(DISADVANTAGED) is always positive, significant, and roughly equal to .085.

In all, the findings from our "standard model" suggest that peer

TABLE 2

Single-Equation Estimates of Pregnancy Equation

Independent Variable	Model 1 (1)	$\frac{d\Pr()}{d\mathbf{x}_j}$ (2)	Model 2 (3)	Model 3 (4)	Model 4 (5)
Intercept	−1.438	. . .	−1.500	−1.801	−1.112
	(.233)		(.262)	(.517)	(.288)
BLACK	.158	.043	.066	.178	.158
	(.092)	(.024)	(.101)	(.095)	(.098)
MOTHER DROPOUT	.826	.222	.840	.818	.790
	(.177)	(.051)	(.178)	(.177)	(.181)
MOTHER HIGH SCHOOL	.373	.100	.381	.356	.357
	(.171)	(.048)	(.172)	(.172)	(.174)
MOTHER SOME COLLEGE	.182	.049	.183	.155	.175
	(.199)	(.053)	(.200)	(.200)	(.203)
RELIGIOUS	−.139	−.037	−.129	−.125	−.120
	(.077)	(.019)	(.081)	(.078)	(.080)
SEX EDUCATION	−.217	−.058	−.219	−.232	−.222
	(.080)	(.021)	(.080)	(.081)	(.083)
FEMALE HH	.273	.073	.280	.265	.284
	(.108)	(.029)	(.108)	(.108)	(.114)
STEPFATHER	.391	.105	.381	.382	.385
	(.153)	(.041)	(.153)	(.153)	(.158)
OTHER FAMILY	.422	.113	.426	.415	.404
	(.151)	(.041)	(.152)	(.151)	(.156)

SIBLINGS	.040	.010	.043	.036	.032
	(.015)	(.004)	(.015)	(.015)	(.016)
QUARTILE 2	.193	.052	.191	.187	.208
	(.105)	(.028)	(.105)	(.105)	(.110)
QUARTILE 3	.141	.038	.138	.128	.146
	(.119)	(.033)	(.119)	(.119)	(.125)
QUARTILE 4	−.056	−.015	−.057	−.082	−.012
	(.139)	(.037)	(.139)	(.140)	(.146)
LOG(DISADVANTAGED)	.082	.022	.086	.085	.087
	(.032)	(.009)	(.032)	(.032)	(.034)
CATHOLIC	...	...	−.057	...	...
			(.161)		
PROTESTANT	...	...	.154	...	...
			(.157)		
OTHER RELIGION	...	...	−.011	...	...
			(.189)		
LOG(AFDC2)	...	...	...	.067	...
				(.089)	
LICENSE	...	...	...	.189	...
				(.111)	
TEACHER/PUPIL	...	...	...	...	−5.233
					(2.941)
CLINICS	...	...	...	...	−33.411
					(63.392)
Log likelihood	−778.9	...	−775.9	−776.8	−715.5
N	1,453	1,453	1,453	1,453	1,322

NOTE.—Asymptotic standard errors are in parentheses.

group effects matter for teenage pregnancy; regardless of how we specify the single-equation model, the results consistently show that placing a teen in a school in which fewer of her peers come from disadvantaged families reduces the probability that she will become pregnant. However, this type of model assumes implicitly that the peer group is an exogenous variable. To some extent, the composition of the peer group is under a family's control; we therefore need to take the endogeneity of the peer group into account. We present a model in the next section that does so.

IV. Simultaneous Equation Estimation

In our system of simultaneous equations, we model a teenager's peer group y_2 as an endogenous variable. We assume that y_2 is a linear function of the variables that determine pregnancy $\mathbf{x}_1$, a set of other exogenous variables $\mathbf{x}_2$, and an error term ϵ_2. Our extended model then consists of equation (1) and

$$y_2 = \mathbf{x}_1\gamma_1 + \mathbf{x}_2\gamma_2 + \epsilon_2. \tag{4}$$

The argument we presented in the Introduction suggests that cov(ϵ_1, ϵ_2) is positive. That is, we would expect that if parents have chosen a school for their daughter in which many of her peers are disadvantaged (ϵ_2 is large), the probability that their daughter will become pregnant given her peer group is also high (ϵ_1 is large). In order to allow for this possibility, we assume that ϵ_1 and ϵ_2 are bivariate normal, each with mean zero and covariance $\mathbf{\Sigma}$, where

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho\sigma_2 \\ \rho\sigma_2 & \sigma_2^2 \end{bmatrix}. \tag{5}$$

The likelihood function for an individual teen in our simultaneous equation model is

$$L = [g(y_1 = 1, y_2)]^{y_1} \cdot [g(y_1 = 0, y_2)]^{1-y_1}, \tag{6}$$

where we have defined the joint density functions

$$g(y_1 = 1, y_2) = \int_{-\mathbf{x}_1\beta_1 - y_2\beta_2}^{\infty} f(\epsilon_1, y_2 - \mathbf{x}_1\gamma_1 - \mathbf{x}_2\gamma_2)d\epsilon_1 \tag{7}$$

and

$$g(y_1 = 0, y_2) = \int_{-\infty}^{-\mathbf{x}_1\beta_1 - y_2\beta_2} f(\epsilon_1, y_2 - \mathbf{x}_1\gamma_1 - \mathbf{x}_2\gamma_2)d\epsilon_1, \tag{8}$$

and $f(\epsilon_1, \epsilon_2)$ is the bivariate normal density function.

To simplify the calculation of the densities in equations (7) and

(8), it is convenient to express $f(\epsilon_1, \epsilon_2)$ in terms of the conditional distribution for ϵ_1 and the marginal distribution for ϵ_2:

$$f(\epsilon_1, \epsilon_2) = f(\epsilon_1 | \epsilon_2) \cdot f(\epsilon_2). \tag{9}$$

The distribution of ϵ_1 conditional on ϵ_2 is normal with mean $\epsilon_2 \rho/\sigma_2$ and variance $1 - \rho^2$. We can then rewrite the joint density functions as

$$g(y_1 = 1, y_2) = \Phi(z) \cdot [\phi(y_2 - \mathbf{x}_1\gamma_1 - \mathbf{x}_2\gamma_2)] \tag{10}$$

and

$$g(y_1 = 0, y_2) = [1 - \Phi(z)] \cdot [\phi(y_2 - \mathbf{x}_1\gamma_1 - \mathbf{x}_2\gamma_2)], \tag{11}$$

where $\Phi(z)$ is the cumulative standard normal distribution, $\phi(z)$ is the standard normal density function, and z is defined as

$$z = \frac{\mathbf{x}_1\beta_1 + y_2\beta_2 + [(y_2 - \mathbf{x}_1\gamma_1 - \mathbf{x}_2\gamma_2)\rho/\sigma_2]}{(1 - \rho^2)^{.5}}. \tag{12}$$

Equations (6), (10), and (11) together imply that the log likelihood function for an individual in this model is

$$\begin{aligned} \log L = {} & y_1 \ln \Phi(z) + (1 - y_1)\ln[1 - \Phi(z)] \\ & - .5 \ln(2\pi\sigma_2^2) - .5\left(\frac{1}{\sigma_2^2}\right)(y_2 - \mathbf{x}_1\gamma_1 - \mathbf{x}_2\gamma_2)^2. \end{aligned} \tag{13}$$

Equation (13) makes the nature of the estimation problem clearer. Suppose for the moment that ϵ_1 and ϵ_2 are uncorrelated and thus $\rho = 0$. In this special case, the likelihood function can be broken into two separate pieces. The first line in (13) would simply be the likelihood function for our single-equation probit model, and the second would be the likelihood function for the normal linear least-squares model associated with y_2. Thus if $\rho = 0$, there is no gain in moving to the simultaneous equation model. If $\rho \neq 0$, however, then the two models would lead to different parameter estimates.

This model requires a set of variables ($\mathbf{x}_2$ in the notation we have been using) that are exogenous determinants of the characteristics of a teenager's peer group but that are not determinants of the probability of pregnancy. We have included in $\mathbf{x}_2$ the metropolitan area unemployment rate, median family income, poverty rate, and the percentage of adults who completed college. These variables are likely to be correlated with the characteristics of a teen's school; everything else equal, we would expect to find that someone who lives in a metropolitan area in which the poverty rate is high is more likely to attend a high school in which many of the other students are disadvantaged.

The metropolitan area variables are therefore designed to measure the range of choices available to an individual household. We are assuming implicitly that parents take their choice of metropolitan area as given when choosing a school for their daughter but treat their location within the metropolitan area (and thus their daughter's school) as a decision variable.[15]

Tests using this set of instrumental variables were reassuring. They are, not surprisingly, important determinants of our peer group variable, the log of the percentage of students disadvantaged.[16] However, they have little explanatory power in the basic pregnancy equation. When placed in the pregnancy equation, the estimated coefficient of each of the instrumental variables was smaller than its estimated standard error, and the instrumental variables taken as a group did not add significantly to the explanatory power of the equation.[17]

We have estimated the simultaneous equation model using a maximum likelihood technique suggested by Berndt et al. (1974). Column 2 of table 3 presents the system estimates of the pregnancy equation. For comparison, we have repeated in column 1 of that table the results from the first single-equation probit model presented in table 2.

Several important results emerge in table 3. The correlation measure ρ is positive and significant, and the error terms in the equation that explain LOG(DISADVANTAGED) and the equation that describes the probability of pregnancy are positively correlated. A χ^2 test shows that the simultaneous equation estimates are significantly different from the single-equation estimates.

When these error terms are correlated, single-equation models may mistakenly attribute part of the impact of unobservable household characteristics to the peer group. A comparison of the estimates of the coefficients on LOG(DISADVANTAGED) in the two models suggests

[15] We would argue that this is a sensible assumption given that intrametropolitan mobility is much more common than intermetropolitan mobility. Consider those families who lived in metropolitan areas in 1980 and moved between 1975 and 1980. Two-thirds of those families moved within the same metropolitan area.

[16] The R^2 in a first-stage regression of LOG(DISADVANTAGED) on the vector $\mathbf{x}_1$ and the instruments is .266. The coefficients (standard errors) on the SMSA-level variables in the first-stage regression are log of the percentage of families in poverty, .829 (.154); log of median family income, −.917 (.395); log of the unemployment rate, .138 (.121); and log of the percentage of adults with a college degree, .197 (.173).

[17] The coefficients (standard errors) on the SMSA-level variables in the pregnancy probit are log of the unemployment rate, −.017 (.147); log of median family income, .353 (.489); log of the percentage of families in poverty, −.184 (.193); and log of the percentage of adults with a college degree, −.062 (.212). The −2 log likelihood test statistic that these four variables are jointly zero is 5.47, which is well below the 95 percent critical value of 9.49 for a χ^2 distribution with four degrees of freedom. As a further check, we experimented with some alternative sets of instrumental variables consisting of characteristics at the county and state levels. Use of these alternative instruments had little impact on our basic results.

TABLE 3

COMPARISON OF SINGLE-EQUATION AND SYSTEM ESTIMATE OF THE PREGNANCY EQUATION

Independent Variable	Single-Equation Probit (1)	System Estimate (2)
Intercept	−1.438	−.816
	(.233)	(.372)
BLACK	.158	.295
	(.092)	(.108)
MOTHER DROPOUT	.826	.863
	(.177)	(.182)
MOTHER HIGH SCHOOL	.373	.388
	(.171)	(.178)
MOTHER SOME COLLEGE	.182	.151
	(.199)	(.204)
RELIGIOUS	−.139	−.140
	(.077)	(.078)
SEX EDUCATION	−.217	−.248
	(.080)	(.079)
FEMALE HH	.273	.244
	(.108)	(.106)
STEPFATHER	.391	.423
	(.153)	(.156)
OTHER FAMILY	.422	.447
	(.151)	(.143)
SIBLINGS	.040	.046
	(.015)	(.015)
QUARTILE 2	.193	.145
	(.105)	(.103)
QUARTILE 3	.141	.005
	(.119)	(.129)
QUARTILE 4	−.056	−.248
	(.139)	(.161)
LOG(DISADVANTAGED)	.082	−.146
	(.032)	(.104)
ρ	. . .	.279
		(.116)
Log likelihood	−3,015.66*	−3,012.97

* This log likelihood value corresponds to the system likelihood with ρ constrained to equal zero. Asymptotic standard errors are in parentheses.

that in this particular example, the problem is quite serious. The estimated coefficient of LOG(DISADVANTAGED) is positive and significant in the single-equation model. In the simultaneous equation model, however, it has the wrong sign and is insignificant. The peer group effect vanishes in this case when we take into account the endogeneity of the peer group. In our sample, the entire effect of the peer group in the single-equation model can be attributed to the choices families make: teens whose parents choose schools in which their daughters associate with relatively few low-income peers have a lower probability of becoming pregnant than we would have expected given their observed characteristics.

Most of the other results are roughly similar in the two models. None of the other variables changes sign, and all the variables that were significant in the single-equation model are significant in the simultaneous equation model. Moreover, the magnitude of most of the coefficients is roughly the same in both models. The one exception is the race variable: its coefficient increases substantially when we move to a simultaneous equation model.

V. A Second Case of Peer Group Effects: School Dropouts

We have also examined a second issue in which peer group influences are thought to be important: school dropouts. We use the same sample, data, and models to explore the dropout issue that we used above. We define DROPOUT as a dichotomous variable that has a value of one if a respondent did not complete high school before she was 20 years old. It appears that, in contrast to pregnancy, underreporting among dropouts is not a serious problem in the NLSY. In our sample, 16.1 percent of the whites, 29.4 percent of the Hispanics, and 18.0 percent of the blacks did not finish high school by the time they became 20 (DROPOUT = 1); U.S. Department of Education (1987) reports comparable rates of 12.6 percent, 25.8 percent, and 17.3 percent.[18]

Table 4 summarizes our basic results. The structure of the table directly parallels table 3. Column 1 reports estimates of a single-equation probit model. Column 2 summarizes the probit equation when estimated as part of a simultaneous equation model that treats the peer group as a choice variable.

The estimated coefficient of LOG(DISADVANTAGED) in the single-equation model is positive and statistically significant. When we treat the peer group as an exogenous variable, we find that placing a student in a school in which a greater proportion of students are disadvantaged increases the probability that she will not complete high school. Moreover, the effect is fairly large from a policy perspective: our results imply that moving a teen from a school in which 22

[18] We would expect to find a higher dropout rate in our study because we have drawn respondents from both the NLSY cross-section sample and the NLSY supplemental sample designed to oversample low-income teens. Not surprisingly, there is a clear correlation in our sample between becoming pregnant and dropping out of school. The relationship is far from perfect, however. Roughly 38 percent of the teens who became pregnant dropped out; 43 percent of those who dropped out became pregnant.

TABLE 4

COMPARISON OF SINGLE-EQUATION AND SYSTEM ESTIMATE OF THE DROPOUT EQUATION

Independent Variable	Single-Equation Probit (1)	System Estimate (2)
Intercept	−2.584	−1.986
	(.413)	(.533)
BLACK	−.513	−.366
	(.112)	(.138)
MOTHER DROPOUT	1.561	1.575
	(.379)	(.376)
MOTHER HIGH SCHOOL	1.039	1.027
	(.379)	(.381)
MOTHER SOME COLLEGE	.846	.791
	(.403)	(.397)
RELIGIOUS	−.280	−.271
	(.087)	(.088)
FEMALE HH	.264	.240
	(.119)	(.118)
STEPFATHER	.469	.495
	(.169)	(.171)
OTHER FAMILY	.441	.464
	(.168)	(.161)
SIBLINGS	.081	.087
	(.017)	(.017)
QUARTILE 2	−.058	−.100
	(.113)	(.114)
QUARTILE 3	−.320	−.437
	(.134)	(.140)
QUARTILE 4	−.459	−.637
	(.163)	(.181)
LOG(DISADVANTAGED)	.175	−.042
	(.038)	(.114)
ρ	...	.267
		(.128)
Log likelihood	−2,797.71*	−2,795.85

* This log likelihood value corresponds to the system likelihood with ρ constrained to equal zero. Asymptotic standard errors are in parentheses.

percent of the students are disadvantaged (the mean in our sample) to one in which 47 percent are disadvantaged increases the chances she will drop out of high school by 2.2 percentage points.

This result disappears, however, in the simultaneous equation model, where the estimated coefficient has the wrong sign and is insignificant. The correlation coefficient ρ is positive and significant in the simultaneous equation model. The results in our dropout study are thus very similar to those in our pregnancy study. The unobservable factors that determine a student's peer group and those that determine whether or not she will drop out of school are positively correlated. When we ignore this correlation in a single-equation model, we conclude that the peer group matters; when we take this correlation into account, the peer group effect vanishes.[19]

Our estimates of the effects of most of the remaining variables are consistent with the education literature. In both the single-equation and simultaneous equation models, we find that the estimated coefficients for the family structure variables (FEMALE HH, STEPFATHER, and OTHER FAMILY), the religion variable, family size, and mother's education are significant and have the expected sign. QUARTILE 3 and QUARTILE 4 are negative and significant in both models; the magnitudes of the coefficients on the income quartile variables suggest that the probability of dropping out is a decreasing function of income. The race variable is negative and significant in both models, though smaller in absolute value in the simultaneous equation model. We thus find that the fact that black girls in our sample are more likely to drop out is not a result of racial differences (in fact, racial differences alone would lead us to find that blacks are less likely to drop out) but is instead a result of differences in family structure, parents' education, and income.

VI. Are Our Results Robust?

Finally, we have explored the sensitivity of our results to changes in the specification of our model. Some of the results of these tests are shown in table 5. That table presents the peer group coefficient from a single-equation probit model (col. 3), the peer group coefficient from a probit estimated as part of a simultaneous equation model (col. 4), the correlation between the error terms in the simultaneous equation model ρ (col. 5), and the statistic required to test the hypothesis that $\rho = 0$ (col. 6). All the equations reported in table 5 included all the explanatory variables from our basic model in tables 2, 3, and 4.

The results reported in table 5 suggest that our basic findings are robust. The peer group measure is always positive and significant in a single-equation probit model; it is always negative and insignificant when we treat the peer group as an endogenous variable in a simultaneous equation model.

The first two rows of table 5 repeat our estimates of the pregnancy and dropout equations. The rest of the table is structured as follows.

[19] On the related issue of college plans and attendance, peer group effects also seem not to matter. Jencks and Mayer (1990, pp. 178–79) note that "this is a case in which sociologists can truly claim to have learned something nobody knew to begin with, namely that a high school's socioeconomic mix has very little net effect on whether graduating seniors plan to attend college, actually attend college, or graduate from college."

TABLE 5

Sensitivity Analysis of Simultaneous Equation Model

Dependent Variable	Sample (1)	Peer Group Measure (2)	Parameter Estimates and Standard Errors: Probit Estimate of Peer Group Effect (3)	System Estimate of Peer Group Effect (4)	ρ (5)	−2 Log Likelihood Test* (6)
PREGNANT	Full sample (N = 1,453)	LOG(DISADVANTAGED)	.082 (.032)	−.146 (.104)	.279 (.116)	5.38
DROPOUT	Full sample (N = 1,453)	LOG(DISADVANTAGED)	.175 (.038)	−.042 (.114)	.267 (.128)	3.72
BIRTH	Full sample (N = 1,453)	LOG(DISADVANTAGED)	.103 (.035)	−.131 (.112)	.285 (.126)	4.90
PREGNANT	Full sample (N = 1,453)	LOG[DISADVANTAGED/ (100 − DISADVANTAGED)]	.053 (.021)	−.105 (.076)	.261 (.116)	4.88
BIRTH	Full sample (N = 1,453)	LOG[DISADVANTAGED/ (100 − DISADVANTAGED)]	.066 (.024)	−.095 (.082)	.267 (.125)	4.44
DROPOUT	Full sample (N = 1,453)	LOG[DISADVANTAGED/ (100 − DISADVANTAGED)]	.109 (.035)	−.053 (.083)	.268 (.127)	3.90
DROPOUT	Full sample (N = 1,386)	LOG(SCHOOL DROPOUT RATE)	.185 (.040)	.110 (.207)	.084 (.221)	.12
PREGNANT	White subsample (N = 1,133)	LOG(DISADVANTAGED)	.079 (.036)	−.089 (.110)	.211 (.125)	2.78
BIRTH	White subsample (N = 1,133)	LOG(DISADVANTAGED)	.086 (.040)	−.091 (.122)	.225 (.141)	2.64
DROPOUT	White subsample (N = 1,133)	LOG(DISADVANTAGED)	.150 (.042)	−.086 (.112)	.299 (.129)	4.65

* Null hypothesis: $\rho = 0$. The test statistic is distributed as χ^2 with one degree of freedom. The critical value of the test at the 95 percent confidence level is 3.84.

Births.—As we noted above, it appears that our pregnancy variable contains a great deal of measurement error. One alternative is to focus on births rather than pregnancy. Measurement error in births in the NLSY is apparently very small. In our sample, 19.1 percent of all women reported that they had given birth while a teen; using birth record data from *Vital Statistics* for a similar time period, we calculate that roughly 20 percent of women gave birth before the age of 20.[20] The dependent variable BIRTH in rows 3, 5, and 9 of table 5 has a value of one if the respondent gave birth before she reached the age of 20, and zero otherwise. This definition is consistent with the way we defined PREGNANT (except in those cases in which a respondent became pregnant for the first time and gave birth while she was 20).

Functional form.—The peer group literature provides little guidance as to appropriate functional form. In rows 4, 5, and 6 we use the log-odds transformation LOG[DISADVANTAGED/(100 − DISADVANTAGED)] as a measure of the peer group. This transformation is an increasing "S-shaped" function of DISADVANTAGED: it is concave when DISADVANTAGED $\leq$ 50 and convex when DISADVANTAGED $\geq$ 50. This functional form is consistent with some of Crane's (1991) findings.[21] In row 7 we use in our dropout model the proportion of students in a school who drop out as an alternative measure of the peer group (estimates of the proportion of students in a school who become pregnant or give birth are not available in the NLSY).

Race.—We tried to estimate our models using separate white and black subsamples. Our work with the black subsample was not very successful: variables were rarely significant and sometimes had the wrong sign even in single-equation probit models.[22] Our results for the white subsample shown in rows 8–10 of table 5 are very similar to those for the full sample.

One additional finding is noteworthy. When we estimated separate equations for the white and black subsamples to explain the individual dropout decisions, we obtained results similar to those reported earlier in the paper. The estimated coefficient on the peer group variable was significant in the single-equation model and diminished in size when estimated as part of the system for both subsamples. It is interesting to note, however, that the systems estimate was larger

[20] *Vital Statistics* reports raw counts of first births to women by age. Following those women aged 15 in 1978 for the rest of their teen years, we calculate that for this cohort 20.4 percent gave birth before the age of 20.

[21] We also considered models that included DISADVANTAGED, DISADVANTAGED squared, and DISADVANTAGED cubed. The parameter estimates for this model generated a response that was very similar in shape to the log-odds transformation reported in rows 4, 5, and 6 of table 5.

[22] Lundberg and Plotnick (1990) also had little success estimating a separate equation for their black subsample.

in the black than in the white equation and was marginally significant, suggesting perhaps that choice was more restricted for blacks than for whites and that the associated problem of endogeneity was less severe in the black subsample.

Alternative estimation techniques.—Instrumental variables is one alternative to our maximum likelihood technique. Using that approach, we would first estimate equation (4) and then use the predicted values of DISADVANTAGED in the estimation of the probit equation. Our instrumental variables estimate of the DISADVANTAGED coefficient in a model in which the dependent variable is PREGNANT is −.156, very close to the estimate shown in row 1 of table 5.

VII. Concluding Remarks

The central point of this paper can be made in terms of a homely example. Consider an adolescent with a propensity to use drugs. We might expect such an individual to seek out friends in school with a similar inclination. Now, suppose that we were conducting a study of drug use. If we had perfect measures of a person's peer group (his or her group of friends), we would most certainly find that the peer group variable had a high degree of explanatory power, and we would conclude that peer group effects are quite important. But the real circumstances are clearly much more complex, for the group is itself the object (at least to some extent) of the individual's choosing. This must be recognized explicitly in the modeling and estimation of the relationship (Kandel 1978). As Jencks and Mayer (1990, p. 117) note, "Even if individuals restrict themselves to choosing among familiar alternatives, however, a neighborhood's social composition may not have much effect on individual behavior. Most people prefer friends like themselves. So long as neighborhoods and schools are moderately heterogeneous, most young people can indulge this preference."

The discussion points up a second issue: the difficulty of defining "peer group" in the most relevant and in an operational way. It is by no means clear, for example, whether the group with the most influence on an individual's behavior is the community in which the person resides, those in the school the person attends, or a select group of close friends. We are acutely aware that the measure we have chosen for our study, which was dictated largely by the availability of data, has its flaws. We have used a measure of the socioeconomic makeup of the high school the girl attends. There may be measures of other groups, close friends or neighbors, that would capture "the" peer group effect more fully.

Moreover, it is possible that different econometric approaches to this problem might yield different results. Christopher Jencks, for example, has raised the following possibility with us. Suppose that a key factor in the teenage pregnancy and dropout problems is the disparity between a teenager's school and the socioeconomic characteristics of his or her metropolitan area. In that world, a model (such as ours) that uses metropolitan area characteristics as instruments might miss important peer group effects that would emerge in other models.

For these reasons, we surely do not regard the results of our study as "definitive." We most emphatically do not conclude from our findings that peer group effects are inconsequential. The significance of our results, we think, lies elsewhere. We have used as our operational measure of peer group influence what we see as a reasonable or plausible, but certainly not perfect, variable. To the extent that certain peer group effects elude our measure, it is likely to understate the extent of these effects. But what we find striking and noteworthy is the fact that for both the teenage pregnancy and dropout equations, the simple, single-equation results produce statistically significant peer group effects, but when we allow for choice and the implied endogeneity of the peer group, these effects disappear. Our systems estimates of peer group influence are thus very different from those in which the peer group is taken to be exogenous. This is, moreover, a robust result: we have experimented with a variety of specifications, variables, and estimation techniques, and the basic findings are not sensitive to these experiments. This suggests strongly to us that endogeneity is a real issue here: we need to exercise care to account for the scope that individuals have in the choice of the groups of which they are members.

References

Berndt, Ernst R.; Hall, Bronwyn H.; Hall, Robert E.; and Hausman, Jerry A. "Estimation and Inference in Nonlinear Structural Models." *Ann. Econ. and Soc. Measurement* 3 (October 1974): 653–65.

Billy, John O. G., and Udry, J. Richard. "The Influence of Male and Female Best Friends on Adolescent Sexual Behavior." *Adolescence* 20 (Spring 1985): 21–32.

Bradford, David F.; Malt, R. A.; and Oates, Wallace E. "The Rising Cost of Local Public Services: Some Evidence and Reflections." *Nat. Tax J.* 22 (June 1969): 185–202.

Brueckner, Jan K., and Lee, Kangoh. "Club Theory with a Peer-Group Effect." *Regional Sci. and Urban Econ.* 19 (August 1989): 399–420.

Case, Anne C., and Katz, Lawrence. "The Company You Keep: The Effects of Family and Neighborhoods on Disadvantaged Youth." Manuscript. Cambridge, Mass.: NBER, 1990.

Coleman, James S., et al. *Equality of Educational Opportunity.* Washington: Government Printing Office, 1966.

Crane, Jonathan. "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out and Teenage Child Bearing." *American J. Sociology* 96 (March 1991): 1226–59.

de Bartolome, Charles A. M. "Equilibrium and Inefficiency in a Community Model with Peer Group Effects." *J.P.E.* 98 (February 1990): 110–33.

Furstenberg, F. F., Jr.; Morgan, S. P.; Moore, K. A.; and Peterson, K. A. "Exploring Race Differences in the Timing of Intercourse." *American Sociological Rev.* 52 (August 1987): 511–18.

Griliches, Zvi; Hall, Bronwyn H.; and Hausman, Jerry A. "Missing Data and Self-Selection in Large Panels." *Ann. l'INSEE,* nos. 30–31 (1978), pp. 137–76.

Grossman, Michael, and Joyce, Theodore J. "Unobservables, Pregnancy Resolutions, and Birth Weight Production Functions in New York City." *J.P.E.* 98, no. 5, pt. 1 (October 1990): 983–1007.

Hanushek, Eric A. "The Economics of Schooling: Production and Efficiency in Public Schools." *J. Econ. Literature* 24 (September 1986): 1141–77.

Henderson, Vernon; Mieszkowski, Peter; and Sauvageau, Yvon. "Peer Group Effects and Educational Production Functions." *J. Public Econ.* 10 (August 1978): 97–106.

Hofferth, Sandra L. "Factors Affecting Initiation of Sexual Intercourse." In *Risking the Future: Adolescent Sexuality, Pregnancy, and Childbearing,* edited by Sandra L. Hofferth and Cheryl D. Hayes. Washington: Nat. Acad. Press, 1987.

Hogan, D. P.; Astone, N. M.; and Kitagawa, E. M. "Social and Environmental Factors Influencing Contraceptive Use among Black Adolescents." *Family Planning Perspectives* 17 (January/February 1985): 165–69.

Jencks, Christopher, and Mayer, Susan E. "The Social Consequences of Growing Up in a Poor Neighborhood." In *Inner-City Poverty in the United States,* edited by Laurence E. Lynn, Jr. and Michael G. H. McGeary. Washington: Nat. Acad. Press, 1990.

Kandel, Denise B. "Homophily, Selection, and Socialization in Adolescent Friendships." *American J. Sociology* 84 (September 1978): 427–36.

Lundberg, Shelly, and Plotnick, Robert D. "Adolescent Premarital Childbearing: Do Opportunity Costs Matter?" Manuscript. Seattle: Univ. Washington, 1990.

Mayer, Susan E. "How Much Does a High School's Racial and Socioeconomic Mix Affect Graduation and Teenage Fertility Rates?" In *The Urban Underclass,* edited by Christopher Jencks and Paul E. Peterson. Washington: Brookings Inst., 1991.

Mott, Frank L. "Fertility Related Data in the 1982 National Longitudinal Surveys of Work Experience of Youth: An Evaluation of Data Quality and Some Preliminary Analytical Results." Columbus: Ohio State Univ., Center Human Resource Res., 1983.

Schwab, Robert M., and Oates, Wallace E. "Community Composition and the Provision of Local Public Goods: A Normative Analysis." *J. Public Econ.* 44 (March 1991): 217–37.

Summers, Anita A., and Wolfe, Barbara L. "Do Schools Make a Difference?" *A.E.R.* 67 (September 1977): 639–52.

Tiebout, Charles M. "A Pure Theory of Local Expenditures." *J.P.E.* 64 (October 1956): 416–24.

U.S. Department of Education. *The Condition of Education: A Statistical Report.* Washington: Government Printing Office, 1987.

Wilson, William Julius. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy.* Chicago: Univ. Chicago Press, 1987.

LINKED CITATIONS

You have printed the following article:

Measuring Peer Group Effects: A Study of Teenage Behavior
William N. Evans; Wallace E. Oates; Robert M. Schwab
The Journal of Political Economy, Vol. 100, No. 5. (Oct., 1992), pp. 966-991.
Stable URL:
http://links.jstor.org/sici?sici=0022-3808%28199210%29100%3A5%3C966%3AMPGEAS%3E2.0.CO%3B2-A

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

[Footnotes]

[8] **Social and Environmental Factors Influencing Contraceptive Use Among Black Adolescents**
Dennis P. Hogan; Nan Marie Astone; Evelyn M. Kitagawa
Family Planning Perspectives, Vol. 17, No. 4. (Jul. - Aug., 1985), pp. 165-169.
Stable URL:
http://links.jstor.org/sici?sici=0014-7354%28198507%2F08%2917%3A4%3C165%3ASAEFIC%3E2.0.CO%3B2-7

References

Social and Environmental Factors Influencing Contraceptive Use Among Black Adolescents
Dennis P. Hogan; Nan Marie Astone; Evelyn M. Kitagawa
Family Planning Perspectives, Vol. 17, No. 4. (Jul. - Aug., 1985), pp. 165-169.
Stable URL:
http://links.jstor.org/sici?sici=0014-7354%28198507%2F08%2917%3A4%3C165%3ASAEFIC%3E2.0.CO%3B2-7

NOTE: *The reference numbering from the original has been maintained in this citation list.*