University of Wisconsin-Madison

Institute for Research on Poverty

Discussion Papers

Charles F. Manski, Gary D. Sandefur
Sara McLanahan, Daniel Powers

ALTERNATIVE ESTIMATES OF THE EFFECT OF FAMILY STRUCTURE DURING ADOLESCENCE ON HIGH SCHOOL GRADUATION

DP # 929-90

Institute for Research on Poverty
Discussion Paper No. 929-90

Alternative Estimates of the Effect of Family Structure during Adolescence on High School Graduation

Charles F. Manski
Gary D. Sandefur
Sara McLanahan
Daniel Powers

Institute for Research on Poverty
University of Wisconsin-Madison

November 1990

This paper was originally prepared for presentation at the 1990 Annual Meetings of the Population Association of America. Work on this project was supported by grants from the Assistant Secretary for Planning and Evaluation of the U.S. Department of Health and Social Services and from the National Institute of Child Health and Human Development. The authors are grateful to Arthur Goldberger and James Walker for their useful comments. Ted Shen ably assisted in the empirical analysis. Any opinions expressed in the paper are those of the authors alone.

Abstract

Many studies have reported significant empirical associations between family structure during childhood and children's outcomes later in life. It may be that living in a nonintact family has adverse consequences for children. On the other hand, it may be that some unobserved process jointly determines family structure and children's outcomes. How then should one interpret the empirical evidence on the relationship between family structure and children's outcomes? The answer depends on the question asked and on the prior information available to the researcher.

We seek to interpret the association between family structure and high school graduation found among respondents in the National Longitudinal Survey of Youth. We seek to answer the traditional question of the literature on treatment effects: How would the probability of high school graduation vary with family structure if family structure were not selected by parents but were, instead, an exogenously assigned "treatment," as in a clinical trial or other controlled experiment?

The inferential problem is that the data alone do not suffice to identify the treatment effect. Hence any attempt to estimate a treatment effect depends critically on the prior information available to the researcher. We develop alternative estimates of the effect of family structure on high school graduation, obtained under differing assumptions about the actual process generating family structure and high school outcomes.

We first assume strong prior information and present estimates of a set of parametric latent-variable models explaining family structure and children's outcomes. We then assume no prior information at all and report estimates of nonparametric bounds on the graduation probabilities. Finally, we give nonparametric estimates obtained under the assumption that family structure is exogenous with respect to high school graduation.

Our empirical analysis strengthens the evidence that living in an intact family increases the probability that a child will graduate from high school. We also report that the probability of high school graduation increases markedly with both parents' education, regardless of family structure. At the same time, we stress that no empirical analysis of the effect of family structure on children's outcomes can be conclusive. In the absence of prior information, one can only bound the family structure effect. Any attempt to determine the effect more tightly must bring to bear prior information about the process generating family structure and children's outcomes. As long as social scientists are heterogeneous in their beliefs about this process, their estimates of family-structure effects may vary.

1. Introduction

Many studies in recent years have reported significant empirical associations between family structure during childhood and children's outcomes later in life. Several analyses have shown that, conditional on parental income, education, and other observed family characteristics, persons who live in single-parent families at some time during their childhood are more likely to become single parents themselves (Hetherington, Cox, and Cox, 1978; Hogan and Kitagawa, 1985). Other studies have shown that adolescents who live in single-parent families, step-parent/parent families, or with neither parent at age 14 are less likely to graduate from high school than are those who live with both biological parents at age 14 (Krein and Beller, 1986; McLanahan and Bumpass, 1988). These findings have been replicated with several data sets and appear to be consistent across the major racial and ethnic groups in the United States (Sandefur, McLanahan, and Wojtkiewicz, 1989).

The recent increase in the prevalence of marital disruption and single parenthood makes it increasingly important to understand the documented associations between family structure and children's outcomes. It may be that, as the empirical evidence suggests, living in a nonintact family has adverse consequences for children. On the other hand, it may be that some unobserved process jointly determines family structure and children's outcomes.

For example, parents who are less committed to their families may be more likely to divorce and may also provide less guidance and emotional support to their children. Behavioral and/or medical problems such as alcoholism, depression, drug addiction, anxiety, or low self-esteem may make a person more likely to divorce and less effective as a parent. Another possibility is that parents take the interests of their children into account in making decisions about divorce. In particular, parents may compare the likely impact on their children of maintaining a marriage characterized by constant fighting and hostility with the impact of raising the children in a single-parent household.

How then should one interpret the empirical evidence on the relationship between family structure and children's outcomes? The answer depends on the question asked and on the prior information available to the researcher.

In this paper, we seek to interpret the association between family structure and high school graduation found among respondents in the National Longitudinal Survey of Youth (NLSY); Section 2 describes the NLSY data. We seek to answer the following question: How would the probability of high school graduation vary with family structure if family structure were not selected by parents but were, instead, an exogenously assigned "treatment," as in a clinical trial or other controlled experiment? Section 3 formalizes this question in the manner of Rosenbaum and Rubin (1983), Heckman and Robb (1985), and Manski (1990a) and discusses its relevance. This done, we explain the identification problem that makes any attempt to answer the question so dependent on the available prior information.

Given the formulated question, we develop alternative estimates of the effect of family structure on high school graduation. The estimates are obtained under differing assumptions about the actual process generating family structure and high school outcomes. Section 4, which assumes that the researcher has strong prior information, presents estimates of a set of parametric latent-variable models of the type found in Heckman (1978) and Maddala (1983). Section 5, which assumes that the researcher has no prior information at all, reports estimates of the nonparametric bounds introduced in Manski (1989) and uses the estimated bounds to check the parametric models of Section 4. Section 6 reports nonparametric estimates obtained under the assumption that family structure is exogenous with respect to high school graduation. The concluding Section 7 summarizes the evidence.

We hope that this paper serves both substantive and methodological objectives. Our substantive concern, of course, is to better understand the association between family structure and children's outcomes. Our methodological goal is to provide a case study useful in the design of efforts to interpret empirical associations between treatments and outcomes.

2. Data and Measures

The data for our analysis are taken from the cross-sectional, supplemental black, and supplemental Hispanic panels of the NLSY. The NLSY was initiated in 1979 with a national sample of men and women aged 14-21. We confine our sample to individuals aged 14-17 in 1979 for whom we have information on the respondent's family structure at age 14, high school graduation, and several covariates. The covariates include race and ethnicity, gender, region of birth, region of residence in 1979, and parental education.

Family structure is operationally defined to be a binary variable, taken from the 1979 survey, indicating whether the respondent resided in an intact or nonintact family at age 14. A nonintact family is one that does not include both biological or adoptive parents; that is, a family with one parent, with a parent and stepparent, or with no parents. **High school graduation** is a binary variable, taken from the 1985 survey, indicating whether a respondent received a high school diploma or GED certificate by age 20.

Race and ethnic identity are based on self-reports. **Sex** is measured by a dummy variable indicating whether the respondent is **female**. **Mother's education** and **father's education** are measured by years of schooling, sometimes aggregated into broader categories. **Residence** is measured by a set of dummy variables for the Northeast, North Central, South, and West regions.

Southern born is a dummy variable indicating that an individual was born in the South.

Table 5 (see below, pp. 35-38) shows the composition of the sample by race, sex, parents' schooling, and family status. See the column labeled "Cell Sample Size."

3. The Inferential Problem

The title to this paper refers to the "effect" of family structure on high school graduation. In this section, we formally define an "effect" and, in so doing, pose the question we would like to answer. This done, we then explain the identification problem that makes inference on family-structure effects critically dependent on the available prior information.

Definition of the family-structure effect

We assume that the NLSY respondents are drawn from a population of youth, each of whom is characterized by values for the variables (y_I, y_N, z, x). Here x is the vector of observed covariates describing a person's family (e.g., race and parental education). The binary variable z indicates family structure; $z = 0$ if the person resides in an intact family at age 14 and $z = 1$ otherwise.

Each person is characterized by two hypothetical high school graduation outcomes, y_I and y_N. Variable y_I indicates the outcome if the person were to reside in an intact family; $y_I = 0$ if the person would not graduate and $y_I = 1$ otherwise. Similarly, y_N indicates the outcome if the person were to reside in a nonintact family. Of the two outcomes y_I and y_N, one is realized and the other is latent; y_I is realized if $z = 0$ and y_N is realized if $z = 1$. (A more refined analysis would disaggregate the family-structure categories "intact" and "nonintact" so as to distinguish among several family types. If so, each person would be characterized by several graduation outcomes, of which one is realized and the remainder are latent.)

Now consider $P(y_I=1|x)$ and $P(y_N=1|x)$, the probabilities that a person with covariates x would graduate if he or she were to reside in an intact or nonintact family respectively. We define the effect of family structure on graduation to be the difference $P(y_I=1|x) - P(y_N=1|x)$. This quantity measures how the probability of high school graduation would vary with family structure if family structure were not self-selected but were, instead, exogenously assigned.

Discussion

Interest in knowing $P(y_I=1|x)$ and $P(y_N=1|x)$ can be motivated from social, personal, and scientific perspectives. First, consider the social problem of assessing the effects on children of two extreme policies, of which one requires all families to be intact and the other prohibits intact families. For children in families with attributes x, the difference in the high school graduation rates associated with these policies is $P(y_I=1|x) - P(y_N=1|x)$. Of course, such extreme policies are unrealistic; hence the seriousness of this motivation is questionable.

Second, consider the personal problem of a couple who conceive a child and must then decide whether to live together or apart. The couple will presumably want to know how their decision will affect the child's outcomes. Suppose that the couple knows their value of the attributes x but does not possess other information. Then $P(y_I=1|x) - P(y_N=1|x)$ appropriately measures the effect of the couple's family-structure decision on their child's probability of high school graduation. Note that $P(y_I=1|x) - P(y_N=1|x)$ does not appropriately measure the family-structure effect if the couple possesses information beyond x.

Third, consider the scientific problem of establishing an acceptable convention for the reporting of treatment effects. A treatment effect, whether it be the effect of family structure on high school graduation or the effect of smoking on life-span, is the change in outcomes that occurs when different processes are used to assign persons to alternative treatments. As there are innumerable assignment processes, there are innumerable possible definitions of treatment effects. (See Maddala, 1983, Section 9.2, and Heckman and Robb, 1985 for discussions of this point.) What then should be the convention for scientific reporting?

Exogenous assignment has merit as a convention for two reasons. First, exogenous assignment characterizes randomized experiments and mandated policies, processes which are easily grasped. Second, in physical and biological science applications, exogenous assignment is realistic; there, treatment effects have long been defined as the variation in outcomes associated with alternative exogenous assignment processes. Thus, for reasons of simplicity and consistency with practice elsewhere, exogenous assignment has become the convention for social science reporting of treatment effects.

It should be understood, however, that a convention is just that. Other definitions of treatment effects, particularly ones assuming that treatments are self-selected, may well be more relevant in many social science applications.

The identification problem

The central problem we face in our attempt to learn the effect of family structure on children's outcomes is the failure of the available data to identify $P(y_I=1|x)$ and $P(y_N=1|x)$. By the law of total probability,

$$\text{(1a)}\quad P(y_I=1|x) = P(y_I=1|x,z=0)P(z=0|x) + P(y_I=1|x,z=1)P(z=1|x)$$

$$\text{(1b)}\quad P(y_N=1|x) = P(y_N=1|x,z=0)P(z=0|x) + P(y_N=1|x,z=1)P(z=1|x).$$

The sampling process generating the NLSY data identifies $P(y_I=1|x,z=0)$ and $P(y_N=1|x,z=1)$. It also identifies the probabilities of intact and nonintact family status, $P(z=0|x)$ and $P(z=1|x)$. But the sampling process does not identify

$P(y_I=1|x,z=1)$ and $P(y_N=1|x,z=0)$. The reason, of course, is that y_I is unobservable when $z = 1$ and y_N is unobservable when $z = 0$. Hence, in the absence of prior information, the NLSY data cannot identify $P(y_I=1|x)$ and $P(y_N=1|x)$.

The possibilities for inference on family-structure effects depend critically on the available prior information. See Manski (1989, 1990a, 1990b). It is, of course, impossible for us to examine the inferential problem in all potentially interesting informational situations. We therefore have selected a range of cases for study. Sections 4 through 6 specify the information brought to bear and present the findings.

4. Estimates of Parametric Latent-Variable Models

The effect of family structure on high school graduation is identified if the researcher has available sufficiently strong prior information about the probability distribution of (y_I,y_N,z) conditional on x. Prior information is often expressed through the medium of parametric latent variable models; see Maddala (1983). In this section we present estimates of a set of three such models.

The models estimated are all cases of the following three-equation system:

$$(2a)\quad z = 1 \text{ if } Bx + u > 0$$
$$= 0 \text{ otherwise}$$
$$(2b)\quad y_I = 1 \text{ if } Cx + e_I > 0$$
$$= 0 \text{ otherwise}$$
$$(2c)\quad y_N = 1 \text{ if } Cx + A + e_N > 0$$
$$= 0 \text{ otherwise.}$$

Here B and C are parameter vectors of length commensurate with the observed covariate vector x, which includes an intercept term. The scalar parameter A allows the intercept in the equation explaining y_N to differ from that in the equation explaining y_I. A more general model, not considered here, would replace the parameter vector C appearing in (2b) and (2c) with two parameter vectors, C_I and C_N respectively.

The contribution of unobserved covariates to the determination of family structure and outcomes is represented by the disturbances (u, e_I, e_N). Following convention, we maintain the assumption that these disturbances are statistically independent of x and distributed trivariate normal with mean zero and variances equal to one. It follows that

$$(3a)\quad P(y_I=1|x) = F(Cx)$$
$$(3b)\quad P(y_N=1|x) = F(Cx+A),$$

where F(.) denotes the standard normal distribution function.

The estimated models differ in the assumptions they make about the covariance matrix of the disturbances. The model in Section 4.1 assumes that u is statistically independent of (e_I, e_N). That in Section 4.2 assumes that $e_I = e_N$. The model in Section 4.3 imposes no restrictions on the covariance matrix of (u, e_I, e_N). The analysis below focuses on the parameter estimates for these models. The implied estimates of graduation probabilities and treatment effects will be given in Tables 4 and 6, discussed in Sections 5 and 7.

4.1. Probit Model with Exogenous Family Structure

Researchers applying models of the form (2) to study the relationship between family structure and children's outcomes have typically assumed that u is statistically independent of (e_I, e_N). Viewed substantively, this assumption means that the unobserved factors that affect family structure and high school graduation are unrelated. From a probabilistic perspective, the assumption implies that

$$(4a) \quad P(y_I=1|x) = P(y_I=1|x,z)$$

$$(4b) \quad P(y_N=1|x) = P(y_N=1|x,z).$$

Thus, a person's <u>latent</u> high school graduation outcomes (y_I, y_N) are statistically independent of his family structure, conditional on the covariates x. (Equation (4) does not, of course, imply that the realized graduation outcome is independent of family structure; the realized outcome is y_I if $z = 0$ and is y_N if $z = 1$.) When (4) holds, family structure is said to be "exogenous" (Maddala, 1983) or, synonymously, to be "strongly ignorable" (Rosenbaum and Rubin, 1983).

Because $P(y_I=1|x,z=0)$ and $P(y_N=1|x,z=1)$ are identified by the NLSY sampling process, condition (4) suffices to identify $P(y_I=1|x)$ and $P(y_N=1|x)$. Nonparametric estimates of $P(y_I=1|x,z=0)$ and $P(y_N=1|x,z=1)$ will be reported in Section 5. The conventional practice, however, has been to combine (4) with the parametric assumption (3). Taken together, (3) and (4) imply that

(5a) $P(y_I=1|x,z=0) = F(Cx)$

(5b) $P(y_N=1|x,z=1) = F(Cx+A)$.

The parameters C and A may be estimated by maximizing the binary probit likelihood given in (5).

Table 1 reports our estimates of C and A based on the NLSY data. If we use a t-statistic of approximately 2 as the criterion of statistical significance, the results suggest that racial, ethnic, region-of-birth, and region-of-residence differences in high school graduation are not present after controlling for parental education and residence in a nonintact family. Women are more likely to graduate from high school than are men.

TABLE 1: PROBIT MODEL WITH EXOGENOUS FAMILY STRUCTURE

Variable	Parameter Estimate	t-statistic
Constant	.466	5.813
Race and ethnicity (excluded category = white)		
Black	.089	1.307
Hispanic	-.122	-1.569
Female	.225	4.648
Parental education (excluded category = less than high school)		
Mother is high school graduate	.427	5.404
Mother has some college education	.500	3.579
Mother is college graduate	.761	4.473
Father is high school graduate	.383	5.439
Father has some college education	.618	5.240
Father is college graduate	.847	6.147
Mother's education greater than father's	-.044	-0.435
Residence (excluded category = Northeast)		
North Central	.030	0.399
West	-.116	-1.474
South	.083	0.860
Southern born	-.009	-0.112
Nonintact family at 14	-.356	-6.753

Note: The model specification also included categories for American Indian, other races, and no race reported.

The probability of high school graduation increases substantially with both mother's and father's schooling. Observe that the coefficients on the two sets of schooling variables have similar magnitudes; thus, it appears that both parents' education contribute equally to a child's success in school. We also included an educational heterogamy variable, **M's ed gt f's ed**, indicating that mother's education is higher than father's education. This interaction of the parental education variables is known from past work to be related to family structure; our results indicate that it is not related to the probability of high school graduation, conditional on family structure. See Section 6 for further analysis of the relationship between parent's education and children's outcomes.

The main result, apparent in the estimate of A (see the variable **Nonintact at 14**), is that residing in a nonintact family at age 14 has a strong negative effect on the probability of high school graduation. This result is consistent with the findings of past work (Astone and McLanahan, 1989; Sandefur, McLanahan, and Wojtkiewicz, 1989). In fact most of what we know about the relationship between family structure and children's outcomes is based on estimates of models like that in Table 1. Numerous studies have assumed models of the form (5) and have experimented with alternative sets of covariates x, the objective being to determine whether the estimate of A remains consistently negative as the specification of x is varied.

4.2. Bivariate Probit Model with Structural Shift

We suggested in the Introduction that family structure and children's outcomes may be jointly determined by processes that we cannot observe directly. In the context of the latent-variable model (3), this means that the disturbances (u,e_I,e_N) may be statistically dependent. If so, the estimates reported in Table 1 may be invalid.

The "bivariate probit model with structural shift" of Heckman (1978) emerges if we impose no restriction on the covariance between u and e_I but assume that $e_N = e_I$. Substantively, this assumption means that the unobserved factors that affect high school graduation are the same in intact and nonintact families. Probabilistically, this assumption permits us to rewrite the three-equation system (2) as a two-equation system

(6a) $z = 1$ if $Bx + u > 0$

$= 0$ otherwise

(6b) $y = 1$ if $Cx + Az + e > 0$

$= 0$ otherwise,

where $y = y_I(1-z) + y_N z$ is the observed high school graduation outcome and where $e = e_I = e_N$.

Table 2 reports maximum likelihood estimates of a model of the form (6). The specified model imposes some exclusion restrictions; that is, some components of B and C are assumed to equal

TABLE 2: BIVARIATE PROBIT MODEL WITH STRUCTURAL SHIFT

Variable	Family Structure Equation	High School Graduation Equation
Constant	-.722	.534
	-13.01)	(6.10)
Race (excluded category = white)		
Black	.825	.196
	(14.99)	(1.67)
Hispanic	.161	-.151
	(2.35)	(-1.89)
Female		.226
		(4.64)
Parental education (excluded category = less than high school)		
Mother is high school graduate		.401
		(6.97)
Mother has some college education		.455
		(4.18)
Mother is college graduate		.711
		(4.65)
Father is high school graduate		.389
		(6.63)
Father has some college education		.455
		(6.03)
Father is college graduate		.862
		(7.41)
Mother's education greater than father's	.126	
	2.27)	
Residence (excluded category = Northeast)		
North Central	-.266	
	(-4.18)	
West	.026	
	(.38)	
South	-.222	
	(-2.76)	
Southern born	-.151	
	(-2.12)	
Nonintact family at 14		-.677
		(-1.69)
Rho	.184	
	(.77)	

Note: The equations also included categories for American Indian, other races, and no race reported. The numbers in parentheses are the t-statistics for the parameter estimates.

zero. We assume that family structure does not vary with the sex of the family's children and varies with parental education only through the heterogamy variable. We assume that a child's high school graduation outcome does not vary with region of birth, region of residence, or educational heterogamy.

The estimates of B indicate that the probability of living in a nonintact family at age 14 varies with race and ethnicity, educational heterogamy, region of residence, and region of birth. The estimates of C are very similar to those reported in Table 1. The estimate of A is more negative than the estimate in Table 1, but its statistical significance is marginal. The estimate of rho, the correlation between u and e, is small in magnitude and is not statistically significant.

In addition to the model whose estimates are reported in Table 2, we estimated a number of alternative models invoking different exclusion restrictions. The estimate of A remained negative in all cases but its magnitude and statistical significance varied with the model specification.

4.3. Trivariate Probit Model with Structural Shift

Table 3 presents maximum likelihood estimates of the full three-equation system (2), with no restrictions imposed on the covariance matrix of (u,e_I,e_N). The model specification here maintains the same exclusion restrictions as in the model of Table 2. The only difference between the two cases is that we

TABLE 3: TRIVARIATE PROBIT MODEL WITH STRUCTURAL SHIFT

Variable	Family Structure Equation	High School Graduation Equation
Constant	-.725 (-14.80)	.497 (11.61)
Race (excluded category = white)		
Black	.827 (17.32)	.126 (3.03)
Hispanic	.169 (2.70)	-.175 (-3.70)
Female		.228 (7.28)
Parental education (excluded category = less than high school)		
Mother is high school graduate		.399 (10.81)
Mother has some college education		.447 (6.48)
Mother is college graduate		.707 (7.20)
Father is high school graduate		.393 (10.61)
Father has some college education		.627 (9.55)
Father is college graduate		.871 (11.45)
Mother's education greater than father's		.123 (2.61)
Residence (excluded category = Northeast)		
North Central	-.267 (-4.80)	
West	.021 (.36)	
South	-.216 (-3.11)	
Southern born	-.153 (-2.44)	
Nonintact family at 14		-.383 (2.27)

(Table continues)

TABLE 3 (continued)

Rho [e(nint), e(hsgrad-nint)]	-.011 (-.26)
Rho [e(nint), e(hsgrad-int)]	.085 (.92)

Note: The equations also included categories for American Indian, other races, and no race reported. The numbers in parentheses are the t-statistics for the parameter estimates.

now have estimates of two rho parameters, one giving the correlation between u and e_I and the other that between u and e_N. Note that the correlation between e_I and e_N is not identified, as each NLSY respondent realizes only one of y_I and y_N.

Comparison of Tables 2 and 3 shows that the estimates of B and C are very similar. The estimate of A in Table 3 is negative and statistically significant; its magnitude is close to that of the estimate in Table 1. The estimates of the two rho parameters are both small in magnitude and statistically insignificant. Thus, the estimates of Tables 2 and 3 provide no evidence against the hypothesis, invoked in Table 1, that family structure is exogenous.

5. Nonparametric Bounds on the Graduation Probabilities

The foregoing models impose strong assumptions on the process determining family structure and high school graduation. These assumptions may well be inappropriate. If so, the parameter estimates in Tables 1 through 3 may all be misleading.

Suppose that one has available no prior information at all about the probability distribution of (y_I, y_N, z) conditional on x. Then, as we found in Section 3, $P(y_I=1|x)$ and $P(y_N=1|x)$ are not identified. Nevertheless, these probabilities may still be bounded (Manski, 1989). Section 5.1 gives the simple derivation of the bounds. Section 5.2 describes estimation of the bounds.

Section 5.3 presents the estimates and uses them to check the parametric-model specifications of Section 4.

5.1. The Bounds

Let us reconsider equation (1). Recall that the sampling process generating the NLSY data identifies $P(y_I=1|x,z=0)$, $P(y_N=1|x,z=1)$, $P(z=0|x)$, and $P(z=1|x)$ but not $P(y_I=1|x,z=1)$ and $P(y_N=1|x,z=0)$. The unidentified conditional probabilities must, however, lie in the interval [0,1]. It follows that

$$\text{(7a)}\quad P(y_I=1|x,z=0)P(z=0|x) \le P(y_I=1|x) \le P(y_I=1|x,z=0)P(z=0|x) + P(z=1|x).$$

$$\text{(7b)}\quad P(y_N=1|x,z=1)P(z=1|x) \le P(y_N=1|x) \le P(y_N=1|x,z=1)P(z=1|x) + P(z=0|x).$$

These bounds are remarkable in that they impose no assumptions on the process generating y_I, y_N, and z; they are functions of quantities identified by the sampling process alone. The bounds can be applied given any specification of the covariates x.

Observe that the bound on $P(y_I=1|x)$ has width $P(z=1|x)$ while that on $P(y_N=1|x)$ has width $P(z=0|x)$. This finding is intuitive. The probability of living in an intact family is the probability that we observe a realization of y_I rather than y_N. Hence, as $P(z=0|x)$ increases, the sampling process reveals more about $P(y_I=1|x)$ and less about $P(y_N=1|x)$.

The bounds (7) imply a bound on the treatment effect $P(y_I|x) - P(y_N|x)$, namely

$$\begin{aligned} &P(y_I=1|x,z=0)P(z=0|x) - P(y_N=1|x,z=1)P(z=1|x) - P(z=0|x) \\ (8) \qquad &\leq P(y_I|x) - P(y_N|x) \leq \\ &P(y_I=1|x,z=0)P(z=0|x) + P(z=1|x) - P(y_N=1|x,z=1)P(z=1|x). \end{aligned}$$

The lower bound in (8) is obtained by subtracting the upper bound in (7b) from the lower bound in (7a); the upper bound is obtained similarly.

Inspection of (8) reveals that this bound has width one for all values of $P(z=1|x)$. If the sample data were not available, the effect of family structure on high school graduation could only be said to lie in the interval [-1,1], which has width two. Thus, using sample data alone, we can cut in half the range of uncertainty regarding the treatment effect. Tighter bounds can be obtained only if prior information is available. The implications of various forms of prior information are studied in Manski (1990a, 1990b).

5.2. Estimation of the Bounds

Estimation of the bounds entails estimation of $P(y_I=1|x,z=0)$, $P(y_N=1|x,z=1)$, $P(z=0|x)$, and $P(z=1|x)$. Each of these conditional probabilities is estimable using any of the numerous nonparametric regression estimators developed over the past twenty-five years. See, for example, Hardle (1990).

The estimates presented here specify the covariates x as in Tables 2 and 3. That is, we estimate bounds on probabilities of high school graduation conditional on race, sex, and parental education. Race and sex are discrete variables; hence nonparametric estimation implies separation of the NLSY respondents into (race,sex) cells. The parental education data are father's and mother's years of schooling. We treat these two covariates as continuous variables.

Within each (race,sex) cell, we compute kernel estimates of the bounds. The kernel is the standard spherical bivariate normal density function and the bandwidth is fixed at one. Thus, suppose we wish to estimate the bounds for a child i whose mother and father have M_i and F_i years of schooling respectively. The weight given to an observation j with parents' schooling (M_j,F_j) is proportional to $g(M_i-M_j)g(F_i-F_j)$, where $g(.)$ is the univariate standard normal density function.

Tables 4A through 4D report estimates for the following (race,sex) cells: white males, white females, black males, and black females. We have not estimated bounds for other

TABLE 4: BOUNDS ON AND ESTIMATES OF THE GRADUATION PROBABILITIES
$P(y_I=1|x)$ and $P(y_N=1|x)$

A. WHITE MALES

Parents' Schooling F	Parents' Schooling M	Family Status	(L-)	Nonparametric Bounds L	Nonparametric Bounds U	(U+)	Nonparametric Model	Parametric Models 1	Parametric Models 2	Parametric Models 3
<12	<12	I	.427	.463	.753	.820	.651	.713	.767	.733
		N	.106	.144	.855	.910	.497	.582	.521	.595
	12	I	.623	.678	.897	.929	.869	.829	.871	.847
		N	.103	.149	.929	.960	.680	.723	.675	.739
	>12	I	.579	.648	.942	.966	.919	.866	.896	.875
		N	.075	.137	.842	.954	.466	.775	.722	.779
12	<12	I	.642	.699	.872	.919	.845	.834	.868	.845
		N	.078	.097	.924	.962	.561	.730	.671	.737
	12	I	.689	.729	.913	.938	.892	.915	.936	.921
		N	.109	.141	.958	.975	.766	.845	.800	.849
	>12	I	.669	.720	.923	.953	.905	.926	.950	.937
		N	.112	.158	.954	.982	.778	.863	.834	.875
>12	<12	I	.651	.696	.934	.963	.913	.891	.904	.908
		N	.102	.206	.968	.989	.866	.810	.739	.828
	12	I	.773	.823	.949	.967	.942	.956	.961	.961
		N	.088	.114	.988	.993	.905	.913	.864	.917
	>12	I	.802	.848	.978	.990	.975	.974	.978	.976
		N	.090	.120	.990	.998	.923	.945	.914	.946

Key to symbols

F = father M = mother
I = intact family N = nonintact family
(L-) = .05-quantile of bootstrapped distribution of lower-bound estimate
L = actual lower-bound estimate
U = actual upper-bound estimate
(U+) = .95-quantile of bootstrapped distribution of upper-bound estimate
models 1 through 3 = estimated probabilities using parameter estimates in Tables 1 through 3

(Table continues)

TABLE 4 (continued)

B. WHITE FEMALES

Parents' Schooling F	Parents' Schooling M	Family Status	(L-)	Nonparametric Bounds L	Nonparametric Bounds U	(U+)	Nonparametric Model	Parametric Models 1	Parametric Models 2	Parametric Models 3
<12	<12	I	.616	.657	.861	.904	.826	.782	.830	.803
		N	.099	.128	.923	.954	.627	.665	.610	.680
	12	I	.649	.697	.924	.957	.903	.880	.913	.894
		N	.109	.161	.933	.969	.709	.794	.752	.807
	>12	I	.589	.728	.966	.990	.955	.914	.940	.925
		N	.116	.220	.982	.993	.924	.845	.811	.856
12	<12	I	.665	.700	.925	.955	.903	.880	.911	.893
		N	.113	.157	.932	.968	.698	.794	.748	.805
	12	I	.729	.769	.951	.969	.939	.946	.960	.950
		N	.112	.143	.962	.979	.786	.894	.857	.896
	>12	I	.665	.709	.974	.983	.965	.954	.969	.961
		N	.148	.238	.973	.988	.898	.908	.884	.917
>12	<12	I	.621	.679	.927	.977	.904	.929	.939	.942
		N	.099	.165	.916	.926	.665	.867	.812	.883
	12	I	.749	.810	.972	.984	.968	.974	.977	.977
		N	.116	.150	.987	.993	.926	.945	.909	.947
	>12	I	.787	.836	.981	.994	.978	.984	.987	.986
		N	.103	.142	.997	.998	.979	.965	.942	.966

(Table continues)

TABLE 4 (continued)

C. BLACK MALES

Parents' Schooling F	M	Family Status	(L-)	Nonparametric Bounds L	U	(U+)	Nonparametric Model	Parametric Models 1	2	3
<12	<12	I	.396	.443	.832	.875	.724	.688	.703	.690
		N	.226	.268	.880	.916	.689	.553	.443	.545
	12	I	.457	.533	.885	.902	.776	.814	.825	.815
		N	.216	.245	.932	.953	.696	.705	.602	.696
	>12	I	.492	.573	.860	.929	.804	.857	.857	.847
		N	.166	.230	.943	.979	.801	.762	.653	.740
12	<12	I	.351	.382	.931	.949	.847	.811	.822	.813
		N	.344	.434	.885	.922	.791	.700	.597	.694
	12	I	.413	.474	.926	.951	.865	.905	.907	.901
		N	.305	.353	.901	.936	.781	.830	.741	.818
	>12	I	.348	.423	.931	.971	.860	.920	.925	.919
		N	.314	.424	.916	.960	.835	.853	.778	.846
>12	<12	I	.320	.342	.922	.966	.814	.877	.862	.882
		N	.319	.435	.855	.940	.750	.791	.663	.790
	12	I	.403	.516	.973	.980	.950	.942	.929	.942
		N	.283	.385	.928	.957	.842	.888	.787	.883
	>12	I	.529	.641	.990	.996	.985	.967	.961	.966
		N	.234	.337	.988	.991	.966	.931	.866	.926

(Table continues)

TABLE 4 (continued)

D. BLACK FEMALES

Parents' Schooling F	Parents' Schooling M	Family Status	Nonparametric Bounds (L-)	Nonparametric Bounds L	Nonparametric Bounds U	Nonparametric Bounds (U+)	Nonparametric Model	Parametric Models 1	Parametric Models 2	Parametric Models 3
<12	<12	I	.425	.468	.894	.930	.814	.763	.776	.766
		N	.281	.321	.896	.925	.754	.641	.533	.634
	12	I	.413	.469	.937	.968	.882	.871	.877	.869
		N	.309	.363	.895	.930	.776	.780	.686	.771
	>12	I	.383	.480	.975	.990	.950	.891	.898	.891
		N	.284	.416	.921	.974	.840	.811	.725	.802
12	<12	I	.350	.399	.958	.983	.905	.868	.875	.868
		N	.375	.432	.873	.928	.773	.777	.682	.769
	12	I	.375	.440	.977	.991	.950	.939	.939	.935
		N	.389	.470	.933	.960	.875	.883	.809	.872
	>12	I	.373	.493	.995	.999	.990	.947	.952	.949
		N	.350	.451	.949	.982	.898	.897	.840	.894
>12	<12	I	.473	.618	.970	.990	.956	.918	.896	.917
		N	.259	.295	.942	.967	.838	.851	.722	.841
	12	I	.440	.545	.958	.994	.928	.968	.960	.967
		N	.310	.389	.976	.987	.942	.933	.860	.927
	>12	I	.496	.620	.994	.999	.990	.977	.972	.977
		N	.258	.357	.983	.994	.955	.950	.896	.948

racial/ethnic groups as the NLSY sample sizes were small. Within each of the four cells examined, we have estimated the bound at each value of parental years-of-schooling (e.g., father = 11 years, mother = 14 years). To simplify the presentation, Table 4 reports estimates for broader years-of-schooling categories: less than 12 years, 12 years, greater than 12 years. The aggregated estimates are computed by averaging the raw estimates over all NLSY respondents whose parental years-of-schooling falls within the broader category.

The literature on nonparametric regression analysis has not yet settled on a convention for reporting the precision of an estimate. Several approaches are examined in Hardle (1990), Chapters 4 and 5. In this paper, we report bootstrapped confidence intervals for the bounds. Our procedure has the following steps:

(a) estimate the conditional probabilities $P[(y=i,z=j)|x]$, $i = 0,1$, $j = 0,1$ nonparametrically (recall that y is the observed high school graduation outcome);

(b) apply the estimated $P[(y,z)|x]$ to draw a simulated realization of (y,z) for each member of the NLSY sample, hence generating a pseudo NLSY sample;

(b) estimate the bounds on the pseudo-sample data;

(d) repeat steps b and b five hundred times, thereby yielding a bootstrapped sampling distribution for the bounds;

(e) report in Table 4 the .05-quantile of the bootstrapped distribution of the lower bound and the .95-quantile of the bootstrapped distribution of the upper bound.

5.3. Findings

In this section, we interpret the findings reported in Table 4. We first consider the sampling precision of the bound estimates. This done, we examine the tightness of the bounds on $P(y_I=1|x)$ and $P(y_N=1|x)$. We then use the bounds to check the parametric models of Section 4.

Sampling precision of the bound estimates

Substantive discussion of our estimates of the bounds is worthwhile only if these estimates are reasonably precise. The bootstrapped confidence intervals indicate that the estimates are precise. In 57 of 72 cases, the .05-quantile of the bootstrapped distribution of the lower-bound estimates lies less than .10 below the actual lower-bound estimate; in every case, the .05-quantile lies less than .15 below the estimate. In 65 of 72 cases, the .95-quantile of the bootstrapped distribution of the upper-bound estimates lies less than .05 above the actual upper-bound estimate; in every case, the .95-quantile lies less than .12 above the estimate.

Tightness of the estimated bounds

Inspection of Table 4 shows that the estimated bound on $P(y_I=1|x)$ is in most cases tighter than that on $P(y_N=1|x)$. This is so because the probability of residing in an intact family exceeds the probability of residing in a nonintact family for most values of the covariates x.

The estimated probability of residing in an intact family varies considerably with x; hence the width of the bounds varies considerably with x. The bound on $P(y_I=1|x)$ is tightest, and that on $P(y_N=1|x)$ correspondingly loosest, when x = (white, male, father's schooling > 12, mother's schooling = 12). In this case, the estimated probability of residing in an intact family is .874 and the estimated bounds on graduation probabilities are [.823,.949] and [.114,.988] respectively. Thus, in the absence of prior information, we can pin down $P(y_I=1|x)$ quite well but can say little about $P(y_N=1|x)$.

The bound on $P(y_I=1|x)$ is loosest, and that on $P(y_N=1|x)$ tightest, when x = (black, male, father's schooling > 12, mother's schooling < 12). Here the estimated probability of residing in an intact family is .420 and the estimated bounds on graduation probabilities are [.342,.922] and [.435,.855]. Thus, although we cannot pin down either $P(y_I=1|x)$ or $P(y_N=1|x)$ very well without prior information, we can restrict both probabilities to intervals of length less that .6.

Consistency of the parametric models with the bounds

The estimated bounds have two uses in empirical analysis. One use, discussed above, is to provide interval estimates of $P(y_I=1|x)$ and $P(y_N=1|x)$; estimates that are valid in the absence of prior information. The second use, discussed here, is to test hypotheses on the process generating family structure and children's outcomes.

Let it be hypothesized, for the moment, that $P(y_I=1|x) = f_I(x)$ and that $P(y_N=1|x) = f_N(x)$, where $f_I(.)$ and $f_N(.)$ are specified functions of x. Also suppose that the nonparametric bounds on $P(y_I=1|x)$ and $P(y_N=1|x)$ are known. Then the stated hypothesis is testable. We can conclude that the hypothesis is incorrect if, for any value of x, either $f_I(x)$ or $f_N(x)$ lies outside the nonparametric bound on $P(y_I=1|x)$ or $P(y_N=1|x)$ respectively. On the other hand, we cannot reject the hypothesis if $f_I(x)$ and $f_N(x)$ lie within the bounds for all x.

Now consider the realistic situation in which a parametric model explaining the determination of family structure and high school outcomes is hypothesized and the parameters of this model are estimated. Also suppose that the nonparametric bounds on $P(y_I=1|x)$ and $P(y_N=1|x)$ are estimated. Then we may use the estimated bounds to test the hypothesized model. In particular, we may reject the model if its implied estimates for $P(y_I=1|x)$ and $P(y_N=1|x)$ lie "too far" outside the estimated bounds on these probabilities. To choose a formal rejection region would require knowledge of the sampling distribution of the estimates and specification of a significance level for the test. Here we shall proceed less formally.

The last three columns of Table 4 give the estimates of $P(y_I=1|x)$ and $P(y_N=1|x)$ implied by the three parametric models of Section 4. To obtain these estimates, we first compute estimates for each NLSY respondent by applying equation (3) to the parameter estimates in Tables 1, 2, and 3. We then average the estimates over all NLSY respondents whose racial/ethnic group, sex, and parental education values lie in the appropriate cell of Table 4.

We find that the parametric estimates of $P(y_I=1|x)$ and $P(y_N=1|x)$ lie within the estimated bounds 67 of 72 times for model 1, 64 of 72 times for model 2, and 65 of 72 times for model 3. Not surprisingly, the cases in which the estimated bounds are violated tend to be those in which the bounds are relatively tight. All of the violations concern the estimates for $P(y_I=1|x)$; all but one occurs in the white group.

Are the violations of the bounds sufficiently large as to warrant rejecting the parametric models? Probably not. In no case does a parametric estimate of $P(y_I=1|x)$ or $P(y_N=1|x)$ lie outside the bootstrapped confidence interval for the bounds, although a few estimates are close to the edge. As we interpret it, the evidence in Table 4 does not suffice to reject any of the three models.

6. Nonparametric Estimates Assuming Family Structure Is Exogenous

Section 4 and 5 considered two polar informational situations. In this section, we examine the important intermediate case in which family structure is known to be exogenous but no further information is available. That is, we assume that (4) holds but do not restrict the form of the probability of high school graduation, as was done in (3).

Recall that, given (4), $P(y_I=1|x)$ and $P(y_N=1|x)$ equal $P(y_I=1|x,z=0)$ and $P(y_N=1|x,z=1)$ respectively. Thus (4) alone, without any additional information, identifies $P(y_I=1|x)$ and $P(y_N=1|x)$. Inspection of the bounds (7) shows that $P(y_I=1|x,z=0)$ and $P(y_N=1|x,z=1)$ necessarily lie within the bounds on $P(y_I=1|x)$ and $P(y_N=1|x)$ respectively. Hence, in the absence of prior information, the hypothesis that (4) holds cannot be rejected.

Estimates of $P(y_I=1|x,z=0)$ and $P(y_N=1|x,z=1)$ are reported in Table 4 under the heading "Nonparametric Model." These estimates are algebraically related to our bound estimates. The estimate of $P(y_I=1|x,z=0)$ is the lower bound on $P(y_I=1|x)$ divided by the width of the estimated bound on $P(y_N=1|x)$. The estimate of $P(y_N=1|x,z=1)$ is computed analogously. Table 5 characterizes the sampling uncertainty of the nonparametric estimates by presenting quantiles of their bootstrapped sampling distributions.

The remainder of this section discusses three aspects of the findings.

Sampling precision of the estimates

Examination of Table 5 shows that the precision of the estimates varies substantially with family structure and with parental education. The estimates are more precise for intact families and for ones in which both parents have similar years of schooling. This pattern reflects the distribution of x in the NLSY data, which is presented in the last column of the table. As is well known, the precision of kernel estimates increases with the concentration of data around the x-value of interest.

A familiar quantitative summary of precision is the length of the interval between the .05 and .95 quantile of the sampling distribution. For whites in intact families, the interval length varies from .03 to .14, depending on parents' education. For whites in nonintact families, most of the intervals have length between .15 and .25, although a few are shorter or longer. For blacks, the interval lengths range from .03 to .18 in intact families and from to .08 to .26 in nonintact ones.

Comparison with probit-model estimates

The nonparametric-model estimates of high school graduation probabilities may be compared with those derived from the probit model of Section 4.1. The latter model invokes not only (4) but also the functional form assumption (3). The two sets of estimates differ by more than .05 in 18 of 72 cases and by more than .10 in 4 of 72 cases. We cannot say whether these discrepancies are severe enough to reject the functional form assumptions of

TABLE 5: BOOTSTRAPPED SAMPLING DISTRIBUTIONS
FOR THE NONPARAMETRIC MODEL ESTIMATES

A. WHITE MALES

Parent's Schooling F	Parent's Schooling M	Family Status	Quantile 5%	10%	25%	50%	75%	90%	95%	Cell Sample Size
<12	<12	I	.609	.620	.649	.676	.706	.728	.744	89
		N	.403	.428	.474	.524	.581	.629	.655	36
	12	I	.825	.836	.850	.867	.884	.898	.907	73
		N	.559	.583	.635	.680	.728	.777	.799	21
	>12	I	.862	.877	.896	.916	.935	.948	.955	11
		N	.315	.347	.411	.503	.604	.689	.750	7
12	<12	I	.799	.809	.827	.850	.872	.890	.899	52
		N	.467	.500	.564	.622	.677	.723	.752	10
	12	I	.864	.870	.880	.893	.905	.915	.924	187
		N	.676	.701	.733	.768	.808	.835	.859	43
	>12	I	.867	.880	.893	.911	.925	.936	.943	21
		N	.631	.668	.738	.798	.848	.879	.901	5
>12	<12	I	.859	.870	.896	.917	.934	.947	.952	11
		N	.691	.723	.783	.843	.894	.929	.944	4
	12	I	.911	.917	.928	.938	.949	.957	.961	124
		N	.804	.823	.858	.889	.913	.934	.944	17
	>12	I	.958	.962	.968	.974	.981	.986	.988	145
		N	.822	.847	.881	.918	.962	.980	.984	19
Sample size										869

(Table continues)

TABLE 5 (continued)

B. WHITE FEMALES

Parent's Schooling F	Parent's Schooling M	Family Status	Quantile 5%	10%	25%	50%	75%	90%	95%	Cell Sample Size
<12	<12	I	.800	.810	.825	.841	.858	.870	.879	120
		N	.537	.566	.607	.649	.699	.739	.760	26
	12	I	.862	.870	.887	.905	.923	.935	.943	61
		N	.589	.627	.680	.734	.782	.821	.838	23
	>12	I	.879	.927	.961	.976	.982	.986	.987	9
		N	.834	.868	.907	.937	.958	.969	.976	1
12	<12	I	.868	.879	.893	.908	.924	.935	.943	51
		N	.605	.625	.671	.716	.767	.809	.836	19
	12	I	.912	.919	.929	.939	.950	.958	.963	163
		N	.691	.720	.751	.789	.827	.865	.881	34
	>12	I	.934	.939	.949	.958	.966	.974	.977	28
		N	.807	.825	.856	.884	.916	.936	.947	16
>12	<12	I	.830	.855	.887	.913	.950	.964	.970	12
		N	.513	.541	.580	.628	.665	.697	.713	4
	12	I	.939	.943	.952	.962	.969	.977	.981	101
		N	.851	.866	.890	.913	.936	.950	.957	17
	>12	I	.953	.959	.969	.976	.984	.991	.993	118
		N	.932	.943	.956	.970	.979	.986	.988	20
Sample size										823

(Table continues)

TABLE 5 (continued)

C. BLACK MALES

Parent's Schooling F	Parent's Schooling M	Family Status	Quantile 5%	10%	25%	50%	75%	90%	95%	Cell Sample Size
<12	<12	I	.674	.689	.711	.735	.758	.779	.791	119
		N	.627	.642	.673	.701	.730	.756	.769	73
	12	I	.724	.743	.763	.790	.818	.842	.851	46
		N	.702	.722	.751	.784	.810	.838	.854	18
	>12	I	.724	.742	.776	.822	.857	.883	.896	7
		N	.677	.719	.769	.828	.877	.908	.928	5
12	<12	I	.773	.793	.813	.839	.864	.881	.893	25
		N	.717	.735	.760	.790	.813	.834	.847	45
	12	I	.803	.814	.839	.859	.880	.897	.909	68
		N	.719	.733	.759	.790	.819	.841	.854	49
	>12	I	.764	.782	.823	.862	.895	.924	.941	13
		N	.721	.751	.789	.833	.874	.902	.916	14
>12	<12	I	.749	.770	.792	.817	.847	.899	.928	4
		N	.687	.707	.747	.789	.828	.867	.881	11
	12	I	.874	.890	.911	.932	.946	.957	.962	12
		N	.700	.729	.778	.819	.860	.884	.903	11
	>12	I	.946	.957	.969	.980	.987	.991	.994	23
		N	.877	.895	.918	.938	.957	.968	.974	12
Sample size										555

(Table continues)

TABLE 5 (continued)

D. BLACK FEMALES

Parent's Schooling F	Parent's Schooling M	Family Status	5%	10%	25%	50%	75%	90%	95%	Cell Sample Size
<12	<12	I	.770	.780	.803	.826	.843	.863	.875	124
		N	.700	.716	.738	.762	.783	.804	.817	93
	12	I	.830	.844	.866	.892	.913	.932	.939	32
		N	.697	.714	.747	.777	.806	.831	.849	30
	>12	I	.858	.879	.912	.948	.966	.977	.982	6
		N	.688	.718	.777	.838	.891	.922	.942	8
12	<12	I	.852	.867	.895	.917	.937	.952	.962	24
		N	.719	.739	.770	.799	.827	.852	.865	31
	12	I	.901	.909	.928	.947	.962	.975	.981	39
		N	.801	.815	.838	.863	.888	.913	.922	46
	>12	I	.948	.960	.973	.982	.988	.993	.997	13
		N	.816	.839	.873	.906	.936	.953	.962	10
>12	<12	I	.905	.914	.932	.954	.969	.980	.984	8
		N	.709	.746	.798	.843	.876	.901	.914	3
	12	I	.864	.879	.909	.939	.965	.984	.989	25
		N	.868	.881	.903	.926	.945	.962	.970	13
	>12	I	.962	.974	.984	.989	.993	.996	.997	20
		N	.902	.918	.936	.955	.969	.979	.984	14
Sample size										539

the probit model. The literature does not offer a formal test of the hypothesis (3) given the maintained hypothesis (4).

Variation of graduation probability with parental education

In our Section 4.1 discussion of the probit model estimates, we observed that mother's and father's education appear to contribute substantially and equally to a child's success in school. The more flexible nonparametric model offers the opportunity to explore in greater depth the relationship between parents' education and children's outcomes. In particular, the nonparametric model allows us to examine separately the effects of parental education in intact and nonintact families. (This was not done earlier because our probit model specification constrained the parameter vector C to be the same in equations (2b) and (2c). The family-structure specific effects of parental education could be investigated within the probit model framework by removing this constraint.)

The mass of evidence in Table 4 clearly corroborates the earlier probit findings. In addition, two new findings warrant attention. First, we find that the effect of parental education on the probability of high school graduation is larger in nonintact families than in intact ones. Second, we find that, in nonintact families, the effect of father's education is at least as large as that of mother's education. This last finding is most intriguing. We had expected that mother's background would be the dominant influence in nonintact families, as children in single-parent homes almost always live with their mothers.

The magnitudes of these effects become apparent through a set of examples. For each of the four race and sex groups, let the mother have twelve years of schooling and let the father's education vary from <12 to >12. We find that, for white males, the probability of high school graduation rises from .869 to .942 in intact families and from .680 to .905 in nonintact ones. For white females, the graduation probability rises from .903 to .968 in intact families and from .709 to .926 in nonintact ones. For black males, the graduation probability rises from .776 to .950 in intact families and from .696 to .842 in nonintact ones. For black females, the probability rises from .882 to .928 in intact families and from .776 to .942 in nonintact ones.

Now, contrariwise, let the father have twelve years of schooling and let the mother's education vary from <12 to >12. For white males, the graduation probability rises from .845 to .905 in intact families and from .561 to .778 in nonintact ones. For white females, the graduation probability rises from .903 to .965 in intact families and from .698 to .898 in nonintact ones. For black males, the graduation probability rises from .847 to .860 in intact families and from .791 to .835 in nonintact ones. For black females, the probability rises from .905 to .990 in intact families and from .772 to .898 in nonintact ones.

Thus, the probability that a child graduates from high school increases markedly with both parents' education, regardless of family structure. Our empirical findings do not, of course, reveal the mechanisms at work. It may be that educated parents value schooling more highly and transmit this value to their children. It may be that educated parents are wealthier and therefore are more able to support their children's continuation in school. The mechanisms at work for mothers need not be the same as those operating for fathers. To interpret our findings at a more basic level would require much richer data than are presently available.

7. Conclusion: Evidence on the Effect of Family Structure

Table 6 translates the findings of Table 4 into estimates of the effect of family structure on high school graduation. The estimates from the parametric and nonparametric models are essentially all non-negative. The estimates vary moderately in magnitude across models and across values of the covariates but rarely are below .03 or above .20. There is some tendency for the estimates to fall in magnitude as parents' schooling increases.

The estimated bounds have width one and so cannot determine the sign of the treatment effect. The bounds are informative nonetheless. For example, the estimated bound for black females whose parents both have 12 years of schooling is [-.493,.507]. This does <u>not</u> imply that a positive treatment effect is as likely as a negative one. It does imply that, in the absence of prior

TABLE 6: BOUNDS ON AND ESTIMATES OF THE TREATMENT EFFECT
$P(y_I=1|x) - P(y_N=1|x)$

A. WHITE MALES

Parents' Schooling F	M	Nonparametric Bounds L	U	Nonparametric Model	Parametric Models 1	2	3
<12	<12	-.392	.609	.154	.131	.246	.138
	12	-.251	.748	.189	.106	.196	.108
	>12	-.194	.805	.453	.091	.174	.096
12	<12	-.225	.775	.284	.104	.197	.108
	12	-.229	.772	.126	.070	.136	.072
	>12	-.234	.765	.127	.063	.116	.062
>12	<12	-.272	.728	.047	.081	.165	.080
	12	-.165	.835	.037	.043	.097	.044
	>12	-.142	.858	.052	.029	.064	.030

B. WHITE FEMALES

Parents' Schooling F	M	Nonparametric Bounds L	U	Nonparametric Model	Parametric Models 1	2	3
<12	<12	-.266	.733	.199	.117	.120	.123
	12	-.236	.763	.194	.086	.161	.087
	>12	-.254	.746	.031	.069	.129	.069
12	<12	-.232	.768	.205	.086	.163	.088
	12	-.193	.808	.153	.052	.103	.054
	>12	-.264	.736	.067	.046	.085	.044
>12	<12	-.237	.762	.239	.062	.127	.059
	12	-.177	.822	.042	.029	.068	.030
	>12	-.161	.839	-.001	.019	.045	.020

(Table continues)

TABLE 6 (continued)

C. BLACK MALES

Parents' Schooling F	M	Nonparametric Bounds L	U	Nonparametric Model	Parametric Models 1	2	3
<12	<12	-.437	.564	.035	.135	.260	.145
	12	-.399	.640	.080	.109	.223	.109
	>12	-.370	.630	.003	.095	.204	.107
12	<12	-.503	.497	.056	.111	.225	.119
	12	-.427	.573	.084	.075	.166	.083
	>12	-.493	.507	.025	.067	.147	.073
>12	<12	-.513	.487	.064	.086	.199	.092
	12	-.412	.588	.108	.054	.142	.059
	>12	-.347	.653	.019	.036	.095	.040

D. BLACK FEMALES

Parents' Schooling F	M	Nonparametric Bounds L	U	Nonparametric Model	Parametric Models 1	2	3
<12	<12	-.428	.573	.060	.122	.243	.112
	12	-.426	.574	.106	.091	.191	.098
	>12	-.441	.559	.110	.080	.173	.089
12	<12	-.474	.526	.132	.091	.193	.099
	12	-.493	.507	.075	.056	.130	.063
	>12	-.456	.544	.092	.050	.112	.055
>12	<12	-.324	.675	.118	.067	.174	.076
	12	-.431	.569	-.014	.035	.100	.040
	>12	-.363	.637	.035	.022	.076	.029

information, the effect of family structure on the probability of high school graduation can be pinned down to the range [-.493,.507].

In summary, our empirical analysis strengthens the evidence that living in an intact family increases the probability that a child will graduate from high school. Earlier studies of the family-structure effect were performed under the strong assumptions of Section 4.1; namely that family structure is exogenous and that the high school graduation probability has a particular functional form. Our estimates of latent-variable models that do not impose the exogeneity assumption, reported in Sections 4.2 and 4.3, suggest that the exogeneity assumption is not far off the mark. Our nonparametric-bounds tests, reported in Section 5, do not reject the latent-variable-model specifications. Our nonparametric-model estimates, reported in Section 6, are of particular interest. These estimates indicate the existence of similar family-structure effects in all four of the race and sex groups studied.

The foregoing remarks must be tempered by the realization that no empirical analysis of the effect of family structure on children's outcomes can be conclusive. In the absence of prior information, one can only estimate the bounds of Section 5. Any attempt to determine the family-structure effect more tightly must bring to bear prior information about the process generating family structure and children's outcomes. As long as social scientists are heterogeneous in their beliefs about this process, their estimates of family-structure effects may vary.

References

Astone, N. M. and S. McLanahan. 1989. "Family Structure and High School Completion: The Role of Parental Practices." Madison, Wisconsin: Institute for Research on Poverty Discussion Paper #905-89.

Hardle, W. 1990. Applied Nonparametric Regression Analysis. New York: Cambridge University Press.

Heckman, J. J. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." Econometrica 46: 931-959.

Heckman, J. and R. Robb. 1985. "Alternative Methods for Evaluating the Impact of Interventions." In J. Heckman and B. Singer (editors), Longitudinal Analysis of Labor Market Data. New York: Cambridge University Press.

Hetherington, E. M., M. Cox, and R. Cox. 1978. "The Aftermath of Divorce." In J. H. Stevens and M. Mathews (editors), Mother-Child Relations. Washington, D.C.: National Association for the Education of Young Children.

Hogan, D. P. and E. M. Kitagawa. 1985. "The Impact of Social Status, Family Structure, and Neighborhood on the Fertility of Black Adolescents." American Journal of Sociology 90: 825-855.

Krein, S. F. and A. H. Beller. 1986. "Family Structure and Educational Attainment of Children: Differences by Duration, Age, and Gender." Paper presented at the Annual Meeting of the Population Association of America, San Francisco, Calif.

Maddala, G. S. 1983. Limited-Dependent and Qualitative Variables in Econometrics. New York: Cambridge University Press.

Manski, C. F. 1989. "Anatomy of the Selection Problem." Journal of Human Resources 24: 343-360.

Manski, C. F. 1990a. "Nonparametric Bounds on Treatment Effects." American Economic Review 80(2): 319-323.

Manski, C. F. 1990b. "The Selection Problem." In J. Laffont and C. Sims (editors), Advances in Econometrics. New York: Cambridge University Press, forthcoming.

McLanahan, S. and L. Bumpass. 1988. "Intergenerational Consequences of Family Disruption." American Journal of Sociology 93: 130-152.

Rosenbaum, P. and D. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." Biometrika 70: 41-55.

Sandefur, G. D., S. McLanahan, and R. A. Wojtkiewicz. 1989. "Race, Ethnicity, Family Structure, and High School Graduation." Madison, Wisconsin: Institute for Research on Poverty, Discussion Paper #893-89.