ALTERNATIVE ESTIMATES OF THE EFFECT OF SCHOOLING ON EARNINGS

Donna K. Ginther*

Abstract—This paper examines how assumptions imposed on the data influence estimates of schooling's effect on earnings. The paper models schooling decisions as treatment effects and imposes assumptions about schooling selection to estimate bounds on the treatment effect. The study begins by using the worst-case bounds derived by Manski (1989, 1990, 1994, 1995) and adds assumptions from the Roy model of schooling self-selection to narrow the bounds on the schooling treatment effect. The bounds are narrowed further by using family structure, college proximity, and school-quality characteristics as exclusion restrictions. The selection problem requires the researcher to make explicit assumptions to estimate the effect of schooling on earnings. This paper demonstrates that different selection assumptions yield very different results.

I. Introduction

On average, individuals with more schooling earn higher wages than individuals with less. This conventional wisdom has prompted the Clinton administration to propose policy changes that support increased educational opportunities for all Americans. In particular, the Balanced Budget Act of 1997 provides a $1,500 tax credit for the first two years of postsecondary education and a 20% annual tax deduction for tuition and fees of up to $10,000. Behind these tax incentives is the assumption that the policy prescription of schooling increases wages for anyone receiving the treatment. However, estimating the treatment effect of schooling on wages is not a straightforward matter.

The human capital model defines schooling as an investment that enhances the productive abilities of workers, who, in turn, earn higher wages. In practice, data limitations such as omitted information on ability and the self-selection of schooling levels make it difficult to identify the returns to schooling. If ability is correlated with schooling choice and earnings, estimates that fail to take account of ability and self-selection will be biased (Griliches (1977)). In order to identify the returns to schooling, a researcher must make explicit assumptions about the selection process.

Many recent papers use different exclusion-restriction assumptions to control for ability bias in estimates of the returns to schooling.[1] The exclusion restrictions are assumed to be correlated with schooling choice and uncorrelated with earnings. Instead of trying to estimate the objective return to schooling with new exclusion restrictions, this paper approaches the question from a different perspective. What is the effect of schooling on earnings under alternative sample-selection assumptions? The paper introduces a unified framework based on work by Manski (1989, 1990, 1994, 1995) for estimating the schooling treatment effect. The paper uses data from the National Longitudinal Survey of Youth (NLSY) and different schooling selection assumptions to implement this approach. The research begins by making very general assumptions about the data and gradually adding plausible assumptions from economic theory to bound the effect of schooling on earnings. Unlike other research, bounds on the treatment effect do not identify point estimates of the return to schooling. Point estimates require restrictive assumptions that could potentially bias estimates of the effect of schooling on earnings. As more information from economic theory and exclusion restrictions is added to the model, it is possible to tighten the bounds on the schooling treatment effect.

This paper demonstrates that different selection assumptions yield very different results. First, treatment effect bounds that assume nonrandom selection are uninformative. As information from the Roy model of self-selection is added to the model, treatment effect bounds narrow significantly and, in some cases, identify a positive treatment effect. Second, OLS estimates of the schooling treatment effect are rejected in favor of nonparametric estimates. Third, bounds that incorporate exclusion restrictions can be used to examine whether the exclusion-restriction assumptions are consistent with the data. The exclusion-restriction bounds are narrower than bounds without exclusion restrictions.

The paper is divided into the following sections: Section II discusses the selection problem under different assumptions from economic theory, section III describes the data set used in this research, section IV details the empirical results, and section V concludes.

II. The Selection Problem and the Schooling Treatment Effect

A. *Schooling as a Treatment Effect*

The treatment effect framework provides an intuitive explanation of the schooling selection problem. Consider the following switching process in which the expected wage for an individual with characteristics x differs depending on the treatment: When $s = 1$, the individual obtains schooling beyond high school, and the wage is y_1; when $s = 0$, the individual stops schooling at high school, and the wage is y_0. In addition, assume that potential outcomes for each individual are unrelated to the treatment status of other members of the population; this is Rubin's (1978) stable unit-treatment value assumption (SUTVA). The population average treatment effect, given in equation (1), is the difference

Received for publication April 12, 1996. Revision accepted for publication February 8, 1999.

* Washington University

Charles Manski, Patrick Lampani, Robert Haveman, John Kennan, Rodger Erickson, John Pepper, three anonymous referees, and participants at seminars at Washington University and Rice University provided useful comments. Any errors are my own.

[1] Card (1995a) provides an excellent survey of these articles.

© 2000 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

between these conditional expectations.

$$T(x) = E(y_1|x) - E(y_0|x) \tag{1}$$

The treatment effect in equation (1) is interpretable as the average difference in outcomes between the treated and untreated across all units in the population. Using the law of total probability, one can rewrite $E(y_1|x)$ and $E(y_0|x)$ in equation (2):

$$\begin{aligned} E(y_1|x) &= E(y_1|x, s = 1)P(s = 1|x) \\ &\quad + E(y_1|x, s = 0)P(s = 0|x) \\ E(y_0|x) &= E(y_0|x, s = 1)P(s = 1|x) \\ &\quad + E(y_0|x, s = 0)P(s = 0|x) \end{aligned} \tag{2}$$

The data-sampling process identifies $E(y_1|x, s = 1)$, $P(s = 1|x)$, $E(y_0|x, s = 0)$, and $P(s = 0|x)$. $E(y_0|x, s = 1)$ and $E(y_1|x, s = 0)$ are not identified by the sampling process. Hence, the population average treatment effect of schooling on earnings is not identified. In order to estimate this treatment effect, the researcher must make assumptions about the selection process.

Even though the population average treatment effect in equation (1) is difficult to identify, it remains the parameter of interest in most studies. When a treatment is shown to be effective, policymakers will attempt to influence population outcomes through policy changes. The population average treatment effect measures the effect of these changes.

Given the difficulty in identifying the population average treatment effect, other treatment effects exist in the literature. The effect of treatment on the treated measures the outcome for the subpopulation that selects the treatment. However, this treatment does not measure the effect of policy changes on the entire population and is not the parameter of interest. Angrist and Imbens (1995) and Imbens and Angrist (1994) specify another alternative—local average treatment effects. Identification of local average treatment effects relies on the assumption that the probability of receiving treatment is affected by a monotonic change in an exclusion restriction. Local average treatment effects have the advantage of being identified by the data for a subpopulation. However, the local average treatment effect cannot be generalized for the population and is not the parameter of interest to policymakers. Thus, this paper estimates bounds on the population average treatment effects, allowing the researcher to bound the parameter of interest to policymakers.

B. *Schooling Treatment Effects Using Alternative Assumptions*

Exogenous Selection Mincer (1974) developed an empirical version of the human capital model that assumes schooling treatment is independent of earnings outcomes—an assumption known as *exogenous selection* or *strongly ignorable treatment.* Models imposing exogenous selection assume all individuals have equal ability, and that schooling wage differentials are equalizing on the direct and opportunity costs of schooling. Workers who obtain more schooling are compensated by higher wages that exactly offset the costs of schooling and foregone earnings. For all individuals, the rate of return to additional schooling is equal to the discount rate. Given perfect capital markets in this framework, individuals are indifferent between schooling levels, and researchers can assume exogenous selection. The switching regression format illustrates the exogenous selection assumption:

$$E(y_1|x) = E(y_1|x, s = 1) = E(y_1|x, s = 0)$$

$$E(y_0|x) = E(y_0|x, s = 1) = E(y_0|x, s = 0).$$

Both $E(y_1|x)$ and $E(y_0|x)$ are identified by the data, because $E(y_1|x, s = 1)$ is assumed to equal $E(y_1|x, s = 0)$ because of random assignment to schooling treatments. The treatment effect is also identified under exogenous selection as

$$T(x) = E(y_1|x, s = 1) - E(y_0|x, s = 0). \tag{3}$$

Often exogenous selection is combined with a log-linear functional form assumption to estimate a constant return on the human capital investment. Researchers recently have found evidence against this specification. Both Park (1994) and Heckman et al. (1996) show different rates of return to schooling for different educational degrees. These ‘‘sheepskin effects’’ indicate the return to an additional year of schooling is not constant.

Worst-Case Bounds Even if the researcher is unwilling to impose exogenous selection, it is possible to bound the treatment effect using the data alone. Manski (1989, 1990, 1994, 1995) derives the worst-case bounds on the mean of a bounded function, $E(f(y)|x)$. Consider $E(f(y_1)|x)$, where $f(\cdot)$ is some function mapping y_1 into a known interval $[K_{1U}, K_{1L}]$ with $-\infty \leq K_{1L} < K_{1U} \leq \infty$. The mean of a bounded function, $E(f(y_1|x) \in [K_{1L}, K_{1U}]$, is contained within the upper and lower bounds, K_{1U} and K_{1L}. Using the law of total probability in equation (2) and the bounds on $f(y_1)$, we can bound the mean of a bounded function, $E(f(y_1)|x)$. $E(f(y_1)|\ x, s = 1)$, $P(s = 1|x)$, and $P(s = 0|x)$ are identified by the data, and this information is used to bound $E(f(y_1)|x)$ in equation (4).

$$\begin{aligned} &E(f(y_1)|x, s = 1)P(s = 1|x) + K_{1L}P(s = 0|x) \\ &\qquad \leq E(f(y_1)|x) \leq \\ &E(f(y_1)|x, s = 1)P(s = 1|x) + K_{1U}P(s = 0|x) \end{aligned} \tag{4}$$

$E(f(y_1)|x)$ takes on the value of the lower bound if $f(y_1)$ equals K_{1L} for all individuals receiving the no additional schooling treatment, $s = 0$; $E(f(y_1)|x)$ takes on the value of

TABLE 1.—BOUNDS ON THE SCHOOLING TREATMENT EFFECT UNDER ALTERNATIVE SELECTION ASSUMPTIONS

	Bounds on Treatment Effect	Assumptions
Worst-Case Bounds		$E(y_1 \mid x), E(y_0 \mid x) \in [K_L, K_U]$
Lower bound	$E(y_1 \mid x, s = 1)P(s = 1 \mid x) + K_L P(s = 0 \mid x) -$ $E(y_0 \mid x, s = 0)P(s = 0 \mid x) - K_U P(s = 1 \mid x)$	
Upper bound	$E(y_1 \mid x, s = 1)P(s = 1 \mid x) + K_U P(s = 0 \mid x) -$ $E(y_0 \mid x, s = 0)P(s = 0 \mid x) - K_L P(s = 1 \mid x)$	
Ordered Outcome Bounds		$y_1 \geq y_0$
Lower bound	$E(y_1 \mid x, s = 1)P(s = 1 \mid x) + E(y_0 \mid x, s = 0)P(s = 0 \mid x) -$ $E(y_0 \mid x, s = 0)P(s = 0 \mid x) - E(y_1 \mid x, s = 1)P(s = 1 \mid x) = 0$	$\Rightarrow E(y_1 \mid x, s = 1) > E(y_0 \mid x, s = 1)$
Upper bound	$E(y_1 \mid x, s = 1)P(s = 1 \mid x) + K_U P(s = 0 \mid x) -$ $E(y_0 \mid x, s = 0)P(s = 0 \mid x) - K_L P(s = 1 \mid x)$	$\Rightarrow E(y_1 \mid x, s = 0) > E(y_0 \mid x, s = 0)$
Roy Model Bounds		$s = 1$, when $y_1 e^{-ra} \geq y_0$ $s = 0$, when $y_1 < y_0 e^{ra}$
Lower bound	$E(y_1 \mid x, s = 1)P(s = 1 \mid x) + K_L P(s = 0 \mid x) -$ $E(y_0 \mid x, s = 0)P(s = 0 \mid x) - E(y_1 e^{-ra} \mid x, s = 1)P(s = 1 \mid x)$	$\Rightarrow E(y_1 e^{-ra} \mid x, s = 1) \geq E(y_0 \mid x, s = 1)$
Upper bound	$E(y_1 \mid x, s = 1)P(s = 1 \mid x) + E(y_0 e^{ra} \mid x, s = 0)P(s = 0 \mid x) -$ $E(y_0 \mid x, s = 0)P(s = 0 \mid x) - K_L P(s = 1 \mid x)$	$\Rightarrow E(y_0 e^{ra} \mid x, s = 0) \geq E(y_1 \mid x, s = 0)$

the upper bound if $f(y_1)$ equals K_{1U} for all individuals receiving the no additional schooling treatment. By using the censored sampling process alone, it is possible to bound the mean of any bounded function of y_1. $E(f(y_0)|x)$ can be bounded in a manner similar to equation (4).

Equation (5) derives the bounds for the treatment effect under a worst-case scenario that assumes no prior information except for nonrandom selection.

$$\begin{aligned} &E(f(y_1)|x, s = 1)P(s = 1|x) + K_{1L}P(s = 0|x) \\ &- E(f(y_0)|x, s = 0)P(s = 0|x) + K_{0U}P(s = 1|x) \\ &\qquad \leq E(f(y_1)|x) - E(f(y_0)|x) \\ &\leq E(f(y_1)|x, s = 1)P(s = 1|x) + K_{1U}P(s = 0|x) \\ &- E(f(y_0)|x, s = 0)P(s = 0|x) + K_{0L}P(s = 1|x) \end{aligned} \tag{5}$$

The lower bound of the treatment effect is the difference between the lower bound of $E(f(y_1)|x)$ and the upper bound of $E(f(y_0)|x)$. The upper bound is the difference between the upper bound of $E(f(y_1)|x)$ and the lower bound of $E(f(y_0)|x)$. Using this method, the treatment effect is not identified. However, these bounds provide useful insights about the effect of schooling on wages in the worst possible case—the researcher does not assume exogenous selection and has no prior information about the selection process.[2] The worst-case bounds will serve as a benchmark for comparing estimates of exogenous selection and bounds on the schooling treatment effect under alternative assumptions.

Estimating the worst-case bounds on the schooling treatment effect is difficult when the outcome is a continuous variable such as earnings. Let $f(y_1) = y_1$ and let $f(y_0) = y_0$, where y_i is equal to the log wage. Log wages are bounded below by $K_{1L} = K_{0L} = -\infty$ and are bounded above by $K_{1U} = K_{0U} = \infty$. The worst-case bounds on the mean treatment effect given in equation (5) are infinite. However, if $f(\cdot)$ is a function mapping y_i into a trimmed wage distribution, the researcher can obtain bounds on the trimmed mean and on the trimmed mean treatment effect. In order to identify the upper and lower bounds, α percent of both tails are trimmed. The trimmed mean treatment effects are estimated with the remainder of the sample. The empirical quantiles at the location of the trim provide intuitive bounds on the trimmed wage distribution. Let $K_{1U} = K_{0U} = K_U = Y_{1-\alpha}$ be the $(1 - \alpha)$ quantile of the sample wage distribution. The lower bound at $K_{1L} = K_{0L} = K_L = Y_\alpha$ is the α quantile of the sample wage distribution. Henceforth, y_i, $(i = 0, 1)$ in the text refers to $f(y_i)$, where $f(y_i) = y_i$ for $Y_\alpha \leq y_i \leq Y_{1\text{-}\alpha}$ and $f(y_i) = Y_{1\text{-}\alpha}$ for $y_i > Y_{1\text{-}\alpha}$ and $f(y_i) = Y_\alpha$ for $y_i < Y_\alpha$.

The first row of table 1 presents the worst-case bounds for the trimmed mean treatment effect. The width of the worst-case bounds is, by definition, $K_U - K_L$. The researcher is given wide discretion in choosing where to trim the data. In part, the amount of trimming is a function of the data. One can narrow the worst-case bounds by trimming a larger portion of the data off each tail of the distribution. Section III discusses the implications of choosing α in greater detail.

Ordered Outcomes Bounds Economic theory suggests assumptions that allow one to narrow the bounds on the trimmed mean treatment effect. Assume individuals have equal ability and vary by discount rate and family wealth. Then those individuals with a low discount rate and no borrowing constraints obtain additional schooling. In addition, assume more schooling always improves earnings. The researcher can apply Manski's (1990, 1994) assumption that treatments are ordered by outcome: $y_1 \geq y_0$. Receiving the schooling treatment always yields at least as much income as no treatment. This implies $E(y_1|x, s = 1) > E(y_0|x, s = 1)$ and $E(y_1|x, s = 0) > E(y_0|x, s = 0)$ and can be used to tighten the bounds on $E(y_1|x)$ and $E(y_0|x)$. The second row of table 1 shows the ordered outcome bounds. The lower bound on the treatment effect is zero, and the upper bound on the treatment effect is unchanged from the worst-case upper bound. The ordered outcome bounds become narrower when smaller values of K_u are used.

[2] In a similar approach, Rosenbaum and Rubin (1983) analyze the sensitivity of assuming exogenous selection conditional on different assumptions about an unknown covariate.

Roy Model Bounds The paper proceeds by dropping the ordered outcome assumption and assuming that individuals choose schooling based on their economic ability and valuation of future earnings. These assumptions are the basis of the Roy model used by Willis and Rosen (1979). The Roy model assumes that individuals choose the level of schooling to maximize the present discounted value of earnings conditional on ability and rate of time preference. Like Willis and Rosen, this study parameterizes the earnings stream to motivate the Roy model decision rule and incorporate foregone earnings. Individuals can choose from two earnings streams: y_{1t} if they attend college, and y_{0t} if they do not. Let a be the time spent in additional schooling and t be an index for time. Then the individual earnings streams conditional on schooling choice may be characterized as

$$\begin{aligned} y_{1t} &= 0, && 0 < t \le a \\ y_{1t} &= y_1 e^{g(t-a)}, && a \le t < \infty \\ y_{0t} &= y_0 e^{gt}, && 0 \le t < \infty, \end{aligned}$$

where y_1 and y_0 may be interpreted as either initial or annual earnings. The discount rate (r) is constant and the same for all individuals, and g is the constant rate of earnings growth. In addition, assume that $r > g$, and individuals face an infinite time horizon. The direct costs of schooling are ignored, and nonpecuniary aspects of the schooling decision are assumed away. Let V_1 and V_0 be the present discounted value of each earnings stream:

$$V_1 = \int_a^\infty y_{1t} e^{-rt}\, dt = \frac{y_1}{r-g} e^{-ra}$$

$$V_0 = \int_0^\infty y_{0t} e^{-rt}\, dt = \frac{y_0}{r-g}.$$

Individuals attend college, $s = 1$, when $V_1 \ge V_0$. This occurs when the present value of earnings from additional schooling less foregone earnings is greater than or equal to the present value of earnings from no additional schooling. The economic content of the Roy model is contained in this selection decision.[3]

The Roy model alone does not identify the schooling treatment effect, but it does allow one to narrow the bounds on the mean treatment effect. The Roy model states that individuals choose more schooling when $V_1 \ge V_0$, which implies that earnings must be greater for the higher schooling treatment plus foregone earnings: $y_1 \ge y_0 e^{ra}$. Manski (1994) derives the implications of the Roy model assumptions. The individual knows his own abilities and selects the higher schooling treatment, $s = 1$, when $y_1 e^{-ra} \ge y_0$ and selects no additional schooling, $s = 0$, when $y_1 < y_0 e^{ra}$. This implies $E(y_1 e^{-ra}|x, s = 1) \ge E(y_0|x, s = 1)$ and $E(y_0 e^{ra}|x, s = 0) \ge E(y_1|x, s = 0)$, which allows the researcher to tighten the upper bounds on $E(y_1|x)$ and $E(y_0|x)$. The bounds on the treatment effect under the Roy model assumptions are given in the last rows of table 1.

In order to estimate the Roy model bounds given in table 1, the researcher needs to obtain values for a, the number of years spent in additional schooling, and r, the discount rate. Since the Roy model assumptions affect only the upper bounds on $E(y_1|x)$ and $E(y_0|x)$, it is natural to choose upper bounds on the years spent in additional schooling and the discount rate. The data provide a reasonable upper bound for the amount of additional schooling. In the NLSY, individuals are observed with up to twenty years of schooling. Thus, the upper bound on years spent obtaining additional schooling is the difference between twelve and twenty years of schooling, $a = 8$.

The discount rate (r) is the opportunity cost of foregone earnings. For an individual to select additional years of schooling, the personal rate of return on the schooling investment must be greater than or equal to the discount rate. For the purposes of this paper, the unobserved discount rate is assumed to be constant. To obtain an upper bound on the Roy model, this research assumes two discount rates: 5% and 10%. These discount rates allow one to compare estimates of the Roy model bounds when all individuals face the same discount rate and are relatively "patient," $r = 5\%$, and "impatient," $r = 10\%$.

In table 1, one can easily examine the relationship between the discount rate (r) and the Roy model bounds. Let additional schooling be eight years, and hold the probability of selecting additional schooling constant. As the discount rate increases, the opportunity cost of additional schooling increases, increasing both the upper and lower Roy model bounds. The higher the opportunity cost in terms of foregone earnings, the higher the return to additional schooling needed to select more schooling. This causes an increase in the upper and lower bounds on the treatment effect under the Roy model assumptions.

Two additional facts emerge from the Roy model bounds in table 1. First, the Roy model bounds, by definition, do not identify the sign of the treatment effect. Second, the Roy model bounds are sensitive to the selection of K_L: As K_L gets larger (as more data is trimmed off the tails of the earnings distribution), the width of the Roy model bounds narrows.

C. *Schooling Treatment Effects Using Exclusion Restrictions*

In addition to controlling for selection, the researcher can use exclusion restrictions to narrow the bounds on the treatment effect. An exclusion restriction is some variable z that affects the probability of receiving treatment but is assumed not to affect the distribution of y_1 and y_0. Researchers often use exclusion restrictions combined with func-

[3] In order to examine the effect of self-selection on the earnings distribution, it is necessary to impose precise assumptions about the distribution of ability and wages. These distributional assumptions are not required to estimate bounds on the schooling treatment effect. Heckman and Honore (1990) assume log concavity in order to make predictions about the earnings distribution.

tional form assumptions to identify the schooling treatment effect.[4] However, functional form assumptions are not always necessary to identify treatment effects given exclusion restrictions. Imbens and Angrist (1994) use exclusion restrictions combined with nonparametric assumptions to identify local average treatment effects. In addition, Vella and Verbeek (1997) examine the relationship between exclusion restrictions and control function models of endogenous treatment effects, showing similar results for both approaches.

This paper uses exclusion restrictions to tighten the bounds on the treatment effect. Partition the x matrix into two vectors: $x = (w, z)$. Earnings (y) vary with age (w), but are assumed constant across all values of (z), the exclusion restriction. Manski (1990, 1994) shows that exclusion restrictions allow the researcher to tighten the worst-case bounds by replacing the bounds with the intersection of the worst-case bounds across all values of z. Equation (6) shows the worst-case bounds on $E(y_1|w)$ under an exclusion restriction:

$$\begin{aligned} &\sup_z \, [E(y_1|(w, z), s = 1)P(s = 1|(w, z)) \\ &\quad + K_L P(s = 0|(w, z))] \leq E(y_1|w) \\ &\quad \leq \inf_z \, [E(y_1|(w, z), s = 1)P(s = 1|(w, z)) \\ &\quad + K_U P(s = 0|(w, z))]. \end{aligned} \tag{6}$$

The lower bound is the greatest lower bound across z, and the upper bound is the least upper bound across z. Besides potentially narrowing the bounds on the treatment effect, the exclusion-restriction bounds also examine whether the exclusion-restriction assumptions are consistent with the data.[5] If the lower bound is greater than the upper bound, no constant value of $E(y_1|w, z)$ is possible, and the exclusion restriction is inconsistent with the data. In this case, z violates the exclusion-restriction assumption; it is correlated with both wages and selection into schooling treatment. The bounds given in equation (6) will be used to evaluate exclusion restrictions described in the following section.

III. The Data

This paper uses data on white, employed males from the 1994 National Longitudinal Survey of Youth Geographic Micro-Data (NLSY). The NLSY contains variables including wages and salary for 1993, highest level of schooling completed, age, hours worked, and weeks worked. In order to focus on estimating the effect of additional schooling on earnings, only workers with positive wages who are not self-employed and not in the military are included in the sample. White males are used because they are not likely to experience wage discrimination. The study includes workers with at least 35 weeks and at least 1,400 hours of work in the previous year. Workers with less than twelve years of schooling are excluded from the sample. The full NLSY sample contains 1,411 individuals with complete records for schooling, age, and weekly wages.

The dependent variable is log weekly wages, and the primary independent variable is age. This study deviates from other studies of the returns to schooling by using age instead of potential experience as an independent variable. There are valid reasons for doing so. First, if schooling is endogenous, potential experience defined as age less years of schooling, less six, is a function of the endogenous variable. Second, the probability of attending school, $P(s = 1|x)$, is likely to vary more across cohorts than with potential work experience.

This study uses family structure, proximity of accredited four-year colleges and universities, and school-quality measures as exclusion restrictions from the NLSY data. Exclusion restrictions are assumed to affect schooling choice and not wage outcomes by changing the potential costs or benefits of the schooling choice. This study uses presence of both parents in the home until age eighteen as an exclusion restriction. This variable is assumed to lower the cost of additional schooling because two-parent families usually have higher incomes than single parent families.

The college proximity and public college proximity exclusion restrictions were first used by Card (1995b) to estimate the returns to schooling using data from the NLS. The NLS contains a question regarding the presence of a four-year accredited college or university in the individual's local labor market (SMSA) in 1966, the first year of the NLS survey. The NLSY does not contain this variable but does contain detailed geocode information on the state and county of residence in 1979. For the first time, the 1996–1997 Integrated Post-Secondary Education Data System (IPEDS) contains the state and county geocodes of four-year colleges and universities. By merging the geocode information from the NLSY and IPEDS, it is possible to create exclusion restrictions similar to those used by Card (1995b): presence of an accredited four-year college or university in the county of residence in 1979 and presence of an accredited public four-year college or university in the county of residence in 1979. These exclusion restrictions could be measured with error because residence is measured in 1979 and colleges are observed in 1996. Implicitly, the study assumes that the population of four-year accredited colleges and universities has not changed significantly between 1979 and 1996. One can investigate this assumption by examining the change in the population of accredited four-year colleges in the IPEDS data. According to the 1995 Digest of Education Statistics, between 1979–1980 and 1992–1993, eighty four-year institutions of higher education have closed; none were public institutions. Thus, the measure of presence of a four-year accredited college in county of residence 1979 is likely to understate the true measure.

[4] See Card (1995a) for a survey of this research.

[5] In intuitively similar approaches, Imbens and Rubin (1994) and Balke and Pearl (1994) use exclusion restrictions to generate restrictions on the data-generating process and evaluate the exclusion-restriction assumptions.

Recently researchers have been interested in estimating the effect of school quality—state and school district investments in teachers and students—on the returns to schooling (Card and Krueger (1992), Betts (1995), and Heckman et al. (1996)). In a cross-state school quality model, Card and Krueger find positive and significant effects of school quality on earnings. Heckman et al. find little evidence in support of the effect of school quality on earnings, instead arguing that school quality is likely to affect earnings by increasing the level of schooling attainment. Betts (1995) examines the effect of school quality on earnings using the NLSY. While his study does not find evidence of a positive effect of school quality on earnings, his results do not rule out the potential influence of school quality on level of schooling attainment. Based on this evidence, this study assumes school quality is correlated with choice of additional schooling and not with earnings. Three measures of school quality from the NLSY are used as exclusion restrictions: the teacher-to-pupil ratio, percent of teachers with postgraduate degrees, and beginning teacher salaries. The NLSY does not have school-quality measures for all individuals in the sample. The study creates a subsample of 920 individuals with complete school-quality records.

In order to estimate bounds on the trimmed mean treatment effect, the research must specify α, the percentage of the distribution to be trimmed off each tail of the distribution. The data provide some guide for choosing α. The NLSY is topcoded by the mean wage of individuals earning above the topcode of $100,000.[6] Over 2% of the NLSY sample used in this research is topcoded. Table A.1 includes estimates of the worst-case bounds using α equal to 2.5%, 5%, and 10%. The worst-case bounds have a width of $K_U - K_L$. Accordingly, the bounds narrow as the size of α increases. Even though α quadruples in size, the worst-case bounds in table A.1 remain quite wide, having a width of 1.500. The sign of the treatment effect is never identified. Because the goal of this research is to compare very general assumptions about sample selection with more-restrictive assumptions used in the literature, the paper will estimate bounds on the 2.5% trimmed mean.

Estimates of the returns to schooling are treatment effects that are the difference between mean wages at $s + 1$ and s years of schooling. In most cases, estimates of returns to schooling assume the returns to each year are constant. Instead of estimating the returns to each year of schooling, this study estimates binary treatment effects. The first treatment effect compares workers with twelve years of schooling to those with more than twelve, and the second compares those workers with between twelve and fifteen years of schooling to those with more than fifteen. Table 2 contains descriptive statistics and the number of individuals

[6] By assigning the mean of the topcoded portion of the distribution to all topcoded individuals, mean wages are identified when using the entire sample. However, when trimming the distribution, the entire top-coded portion must be trimmed to identify the trimmed mean wage. When the value of the topcode is assigned to all individuals whose earnings are greater than or equal to the topcode, the mean wage is not identified.

Table 2.—Trimmed Mean Characteristics for NLSY Samples, White Male Workers

Variable	Full Sample	School Quality Sample
Years schooling	14.55	14.57
	(2.66)	(2.59)
Age	32.90	32.96
	(2.29)	(2.21)
Log weekly wages	6.43	6.39
	(0.50)	(0.46)
Four-year college in county	0.87	0.86
	(0.34)	(0.35)
Public four-year college in county	0.23	0.23
	(0.42)	(0.42)
Presence of both parents in home until age 18	0.74	0.72
	(0.44)	(0.45)
Teacher-pupil		0.05
		(0.01)
Teacher education		50.92
		(22.94)
Teacher beginning salaries		10,921
		(1,149)
Observations	1,341	874
K_L = 2.5% quantile	5.234	5.426
K_U = 97.5% quantile	7.478	7.562

Number of Observations in Each Schooling Treatment

Treatment	N	Mean Wage	N	Mean Wage
Schooling = 12	532	6.289	332	6.289
Schooling >12	809	6.519	542	6.450
Schooling ≤15	800	6.286	506	6.288
Schooling ≥16	541	6.637	368	6.528

α = 2.5% of each tail trimmed from the distribution.

receiving each schooling treatment for the NLSY Full and School Quality samples after trimming 2.5% from each tail. Sample sizes are reduced to 1,341 and 874 after trimming. K_U and K_L for each sample are reported.

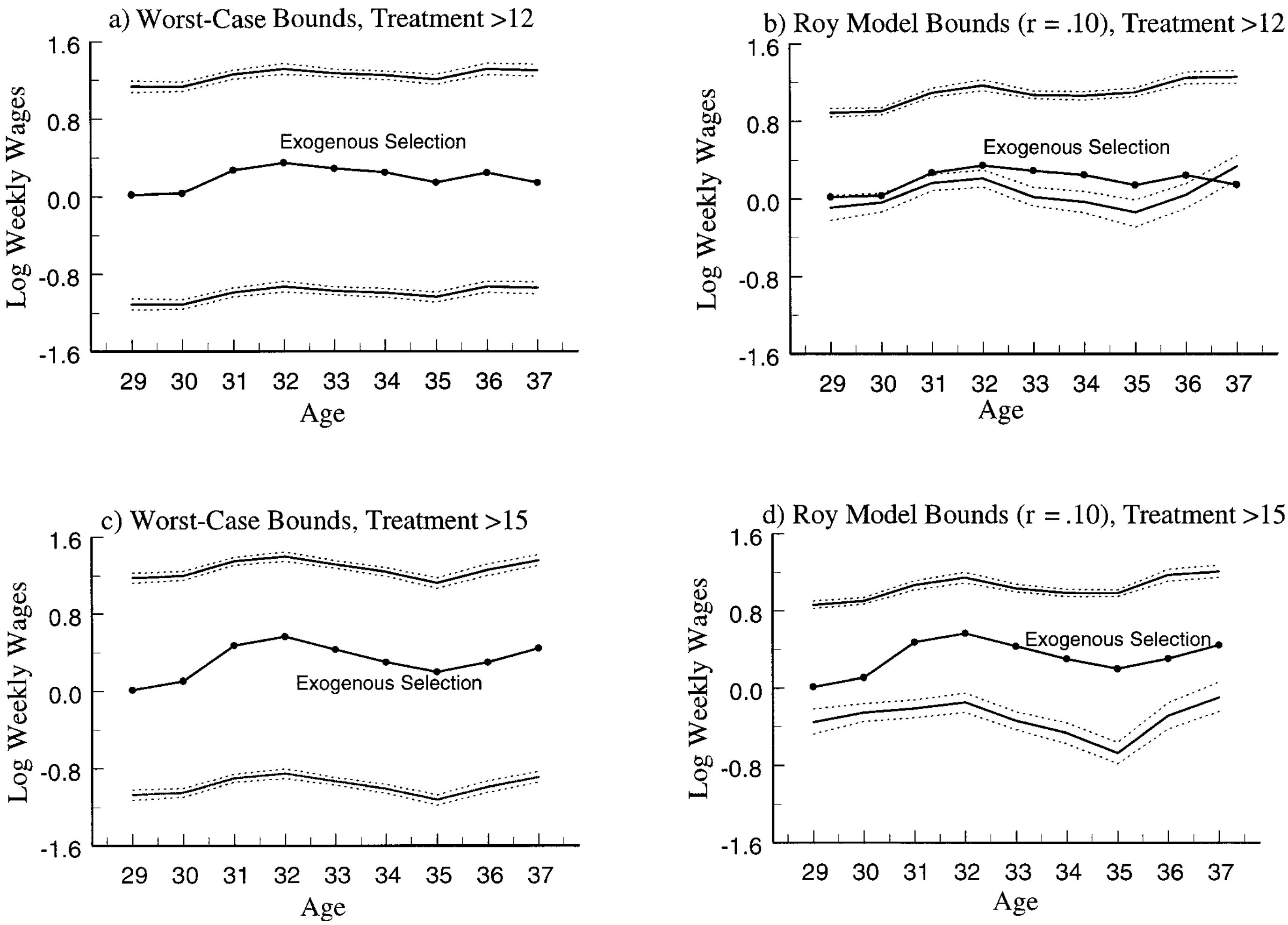
IV. Empirical Results Using Alternative Selection Assumptions

A. *Estimation*

The treatment effect bounds are defined as functions of the conditional expectation of wages and the probability of selecting schooling treatment. It is straightforward to estimate these quantities using nonparametric regression. This study uses the kernel method to estimate the expected value of wages conditioning on the covariate x = age and the schooling treatment, s, and the conditional probability of selecting the schooling treatment conditioning on age. Both the conditional expectation of wages and the conditional probability of selecting the schooling treatment are estimated using the Nadaraya-Watson estimator in equation (7).

$$\hat{\mu}(\mathbf{x}) \frac{n^{-1} \sum_{i=1}^{n} y_i K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right)}{n^{-1} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right)} \tag{7}$$

FIGURE 1.—NLSY SCHOOLING TREATMENT EFFECT BOUNDS[1]



[1]Estimated using NLSY full sample.

Choice of the bandwidth (h) is crucial when implementing kernel estimators. Too small or too large a bandwidth prevents reliable interpretation of the empirical results. This study chooses the bandwidth subjectively using Silverman's (1986) rule-of-thumb bandwidth for kernel estimators: $h = 1.06\sigma_x n^{-1/5}$. The rule-of-thumb bandwidth (h) is a function of the sample size (n) and the standard deviation of x. The Gaussian kernel, $K(\cdot)$, is used as a weighting function in the nonparametric estimates.

Although nonparametric estimation methods allow the researcher to estimate conditional expectations and probabilities without functional form and distributional assumptions, these methods are limited in the number of possible conditioning variables. Thus, the estimates reported in this paper cannot be compared directly to estimates of the returns to schooling found in the literature. OLS estimates of the returns to schooling often control for region of the country, marital status, race, sex, and family background characteristics, whereas the estimates reported here are limited to include only white males. Likewise, dividing the data by marital status and region of the country would greatly complicate the reporting of the results. Thus, the covariates are limited to schooling treatment, age, and the exclusion restrictions in order to evaluate the effects of assumptions on estimates.

B. *Empirical Results Using Treatment Effect Bounds*

The study estimates two treatment effects using the assumptions described above. The first treatment effect is labeled as treatment >12 years schooling, and the second treatment effect is labeled as treatment >15 years schooling. Both treatments are estimated across the ages of the NLSY sample, 29 to 37 years. Estimates of $E(y_1|x, s = 1)$, $P(s = 1|x)$, and $E(y_0|x, s = 0)$ appear in Table A.2.

Figure 1 presents the worst-case, ordered outcome, and Roy model bounds for both trimmed mean treatment effects. The bounds are shown with 95% bootstrap confidence intervals. In these graphs, the estimated bounds are compared with the nonparametric trimmed mean treatment effect assuming exogenous selection. The worst-case bounds in figure 1(a) and 1(c) are, by definition, $K_U - K_L = 7.478 -$

TABLE 3.—NLSY TREATMENT: =12 YEARS AND >12 YEARS OF SCHOOLING UNDER ALTERNATIVE SELECTION ASSUMPTIONS
DEPENDENT VARIABLE: LOG WAGE 1993

	Worst Case		Roy Model, r = .10		Roy Model, r = .05			
Age	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Exogenous	Linear
29	−1.111	1.134	−0.086	0.889	−0.284	0.687	0.021*	0.216
	(−1.169, −1.053)	(1.076, 1.191)	(−0.217, 0.040)	(0.844, 0.935)	(−0.389, −0.186)	(0.626, 0.751)	(−0.089, 0.128)	
30	−1.111	1.133	−0.035	0.904	−0.245	0.714	0.037*	
	(−1.159, −1.064)	(1.085, 1.181)	(−0.132, 0.058)	(0.870, 0.941)	(−0.322, −0.173)	(0.671, 0.763)	(−0.046, 0.122)	
31	−0.986	1.259	**0.169**	**1.097**	−0.089	0.956	0.272	
	(−1.034, −0.940)	(1.211, 1.304)	(0.087, 0.254)	(1.052, 1.144)	(−0.156, −0.020)	(0.900, 1.016)	(0.182, 0.357)	
32	−0.929	1.316	**0.215**	**1.172**	−0.055	1.043	0.349*	
	(−0.985, −0.873)	(1.259, 1.372)	(0.126, 0.303)	(1.117, 1.231)	(−0.124, 0.015)	(0.977, 1.111)	(0.249, 0.442)	
33	−0.974	1.270	**0.024**	**1.074**	−0.203	0.901	0.291	
	(−1.014, −0.932)	(1.231, 1.312)	(−0.068, 0.120)	(1.037, 1.115)	(−0.277, −0.128)	(0.847, 0.958)	(0.212, 0.374)	
34	−0.993	1.251	−0.027	1.065	−0.247	0.885	0.251	
	(−1.039, −0.949)	(1.205, 1.295)	(−0.138, 0.079)	(1.024, 1.110)	(−0.337, −0.162)	(0.823, 0.946)	(0.160, 0.340)	
35	−1.038	1.206	−0.136	1.102	−0.351	0.916	0.144	
	(−1.091, −0.989)	(1.154, 1.256)	(−0.288, −0.006)	(1.059, 1.145)	(−0.473, −0.244)	(0.854, 0.979)	(0.046, 0.237)	
36	−0.928	1.317	**0.045**	**1.250**	−0.208	1.103	0.248	
	(−0.989, −0.871)	(1.256, 1.374)	(−0.094, 0.162)	(1.186, 1.310)	(−0.320, −0.115)	(1.019, 1.182)	(0.151, 0.348)	
37	−0.942	1.303	**0.337**	**1.258**	**0.027**	**1.167**	0.146	
	(−1.004, −0.879)	(1.241, 1.365)	(0.197, 0.452)	(1.195, 1.321)	(−0.086, 0.121)	(1.088, 1.247)	(0.040, 0.258)	

Estimates use NLSY full sample. Numbers in parentheses are 95%-confidence intervals estimated using 1,000 bootstrap subsamples.
* Linear estimates lie outside of 95%-confidence intervals around nonparametric estimates assuming exogenous selection.
Underline indicates sign of treatment effect is identified.

5.234 = 2.244 wide. On average, the upper worst-case bound is approximately 85% larger than the exogenous treatment effect. The lower worst-case bound is negative and large in absolute value. The width of the worst-case bounds underscores the severity of the selection problem when the researcher is unwilling to assume anything; without additional assumptions, these bounds contain no information about the sign and little information about the size of the treatment effect. The ordered outcome bounds are indirectly reported in figure 1: By definition, the lower ordered outcome bound is zero, and the upper ordered outcome bound is equal to the upper worst-case bound. The sign of the treatment effect is positive, but it falls within the wide interval of [0, 7.478]. The Roy model bounds reported in figure 1(b) and (d) are calculated at a 10% discount rate. They are approximately 50% of the size of the worst-case bounds. Clearly, imposing restrictions from economic theory results in tighter bounds.

Table 3 reports the worst-case, Roy model bounds, and 95% confidence intervals on the >12 trimmed mean treatment effect using the full NLSY sample. The worst-case upper bounds range from 1.133 to 1.317 for the >12 treatment and have narrow confidence intervals with an average width of 0.10. The Roy model upper bounds estimated at a 10% discount rate range in size from 0.889 to 1.258 for the >12 treatment with an average of 0.10 difference between the confidence intervals. The Roy model upper bounds estimated at a 5% discount rate range in size from 0.687 to 1.167. The sign of the treatment effect is identified by the ordered outcomes assumption. A positive sign is identified using the Roy model assumptions for both discount rates at selected ages.

Table 4 reports the worst-case, Roy model bounds, and 95% confidence intervals on the >15 trimmed mean treatment effect using the full NLSY sample. The worst-case upper bounds range from 1.123 to 1.396 for the >15 treatment, and the confidence intervals have an average width of 0.10. The upper Roy model bounds at the 10% discount rate range from 0.861 to 1.207, and the confidence intervals have an average width of 0.09. The Roy model bounds at the 5% discount rate range from 0.599 to 1.019 and have confidence intervals with an average width of 0.13. Unlike the >12 treatment, the sign of the >15 treatment effect is not identified by the Roy model bounds. Table 3 and 4 indicate that adding the ordered outcomes and Roy model assumptions provides tighter bounds on the population average treatment effect. Even though the worst-case bounds are precisely estimated, they provide little insight when compared to the Roy model bounds.

C. *Empirical Results Assuming Exogenous Selection*

The bounds on the treatment effect under alternative selection assumptions cannot be compared with estimates of the returns to schooling in the literature. Typically, researchers estimating the returns to schooling assume that the mean wage is linear conditioning on schooling and a quadratic in experience, where schooling is often assumed to be assigned exogenously. In order to compare the bounds estimated here to estimates similar to those found in the literature, this study imposes exogenous selection and a linear functional form when estimating binary schooling treatment effects. These estimates allow the researcher to evaluate the effect of the usual assumptions on the estimated outcomes. The last two columns of table 3 and 4 present these results. Nonparametric estimates of the treatment effect that assume exogenous selection range from 0.02 to 0.35 for the >12 treatment and range from 0.01 to 0.57 in the >15 treatment conditioning

TABLE 4.—NLSY TREATMENT: >15 YEARS AND <16 YEARS OF SCHOOLING UNDER ALTERNATIVE SELECTION ASSUMPTIONS
DEPENDENT VARIABLE: LOG WAGE 1993

	Worst Case		Roy Model, r = .10		Roy Model, r = .05			
Age	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Exogenous	Linear
29	−1.069	1.176	−0.352	0.861	−0.491	0.599	0.013*	0.348
	(−1.123, −1.015)	(1.122, 1.230)	(−0.473, −0.215)	(0.826, 0.901)	(−0.588, −0.384)	(0.549, 0.660)	(−0.103, 0.127)	
30	−1.044	1.200	−0.254	0.902	−0.413	0.661	0.107*	
	(−1.091, −0.999)	(1.154, 1.245)	(−0.347, −0.161)	(0.868, 0.936)	(−0.487, −0.341)	(0.617, 0.708)	(0.012, 0.208)	
31	−0.896	1.349	−0.211	1.065	−0.380	0.834	0.474*	
	(−0.936, −0.855)	(1.308, 1.390)	(−0.307, −0.121)	(1.018, 1.113)	(−0.457, −0.309)	(0.773, 0.897)	(0.390, 0.559)	
32	−0.848	1.396	−0.149	1.144	−0.335	0.930	0.566*	
	(−0.899, −0.797)	(1.346, 1.447)	(−0.253, −0.054)	(1.089, 1.201)	(−0.419, −0.259)	(0.855, 1.002)	(0.460, 0.662)	
33	−0.931	1.313	−0.339	1.033	−0.486	0.780	0.432	
	(−0.970, −0.891)	(1.274, 1.353)	(−0.430, −0.247)	(0.996, 1.075)	(−0.559, −0.413)	(0.723, 0.839)	(0.348, 0.513)	
34	−1.007	1.238	−0.463	0.985	−0.593	0.715	0.300	
	(−1.052, −0.963)	(1.192, 1.281)	(−0.578, −0.362)	(0.946, 1.024)	(−0.686, −0.512)	(0.657, 0.774)	(0.202, 0.394)	
35	−1.122	1.123	−0.671	0.981	−0.784	0.694	0.197*	
	(−1.176, −1.068)	(1.068, 1.176)	(−0.780, −0.560)	(0.946, 1.017)	(−0.877, −0.691)	(0.641, 0.750)	(0.101, 0.289)	
36	−0.986	1.259	−0.287	1.170	−0.477	0.961	0.302	
	(−1.042, −0.923)	(1.202, 1.322)	(−0.422, −0.153)	(1.108, 1.231)	(−0.589, −0.370)	(0.877, 1.046)	(0.211, 0.395)	
37	−0.886	1.358	−0.100	1.207	−0.312	1.019	0.444	
	(−0.941, −0.828)	(1.303, 1.417)	(−0.244, 0.063)	(1.147, 1.273)	(−0.431, −0.186)	(0.932, 1.116)	(0.346, 0.545)	

Estimates use NLSY full sample. Numbers in parentheses are 95%-confidence intervals estimated using 1,000 bootstrap subsamples.
* Linear estimates lie outside of 95 percent confidence intervals around nonparametric estimates assuming exogenous selection.

on age. The treatment effect is constant at 0.22 for the >12 treatment and 0.35 for the >15 treatment in the linear model. In table 3 and 4, linear estimates lie outside the nonparametric 95%-confidence intervals, which indicates that the linear functional form assumption does bias the estimates of the schooling treatment effect. Even if exogenous selection is the correct assumption, assuming a linear functional form to estimate the returns to schooling will bias the results.

D. *Empirical Results Using Exclusion Restrictions*

Applying exclusion restrictions to narrow the bounds on the schooling treatment effect is a two-step process. First, the paper examines whether the exclusion-restriction assumptions are consistent with the data. This involves estimating the exclusion-restriction bounds on $E(y_1|w)$ and $E(y_0|w)$ using equation (6) for all treatments. The bounds are consistent with the data when the upper bound exceeds the lower bound; in this case, a constant value of $E(y_1|w)$ and $E(y_0|w)$ is possible given the exclusion-restriction assumption. The statistical significance of this result is tested by using 1,000 bootstrap samples to recalculate the bounds on $E(y_1|w)$ and $E(y_0|w)$. Under the null hypothesis, the upper bound should exceed the lower bound in at least 95% of the bootstrap estimates. If the null hypothesis is rejected, this suggests that the data are inconsistent with the exclusion-restriction assumption.

Table 5 presents the results of the bootstrap test for each estimate of $E(y_1|w)$ and $E(y_0|w)$ where the schooling treatment is equal to >12, =12, <16, and >15 years.[7] The bootstrap test does not reject presence of both parents at home until age eighteen, presence of a four-year college in county of residence, and presence of a four-year, public college in the county of residence. The data are consistent with the assumption that these exclusion restrictions are correlated with schooling choice and uncorrelated with wages. The bootstrap test rejects the consistency of the school-quality exclusion restrictions with the data, indicating that measures of school quality are correlated with schooling choice and wages.[8]

Since the school-quality exclusion restrictions are inconsistent with the data, the paper proceeds with the second step of estimating exclusion-restriction bounds with the NLSY

TABLE 5.—BOOTSTRAP TEST OF THE CONSISTENCY OF EXCLUSION RESTRICTIONS WITH THE DATA[1,4]: THE PERCENTAGE OF BOOTSTRAP ESTIMATES WHERE THE EXCLUSION RESTRICTION ASSUMPTION IS INCONSISTENT WITH THE DATA

Exclusion Restriction	$E(y_1|w)$ Schooling >12	$E(y_0|w)$ Schooling =12	$E(y_1|w)$ Schooling >15	$E(y_0|w)$ Schooling <16
Both parents at home until age 18[2]	0	0	0	0
Proximity to four-year college[2]	0	0	0	0
Proximity to four-year public college[2]	0	0	0	0
Teachers with advanced degrees[3]	90	20	5	95
Teacher starting salaries[3]	100	84	36	100
Teacher/pupil ratio[3]	100	100	98	100

[1] Bootstrap test consists of 1,000 bootstrap subsamples used to recalculate $E(y_1|w)$ and $E(y_0|w)$. Numbers in the table are the percentage of bootstrap estimates where the lower bound exceeds the upper bound. Exclusion restrictions are inconsistent with the data when more than 5% of the lower bounds exceed the upper bounds.
[2] Estimated using the full NLSY sample.
[3] Estimated using the NLSY school-quality sample.
[4] An appendix that graphs the exclusion-restriction bounds is available from the author by request.

[7] An appendix graphing the exclusion-restriction bounds for all treatments is available from the author by request.

[8] The school-quality exclusion-restriction results are at odds with Betts (1995), who finds that school quality has no significant effect on earnings, while the exclusion-restriction bounds indicate that school quality is correlated with earnings. This discrepancy may be due to the differences in specification used here and in Betts, who specifies an OLS wage equation that uses more covariates than used in this study.

TABLE 6.—NLSY SCHOOLING TREATMENT EFFECT: LIVED WITH BOTH PARENTS UNTIL AGE 18 EXCLUSION RESTRICTION
DEPENDENT VARIABLE: LOG WAGE 1993

	Worst Case		Roy Model, $r = .10$		Roy Model, $r = .05$	
Age	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Upper Bound
Treatment >12						
29	−1.013	0.941	**0.156**	**0.710**	−0.055	0.521
	(−1.065, −0.885)	(0.747, 1.073)	(−0.028, 0.325)	(0.549, 0.804)	(−0.203, 0.092)	(0.384, 0.601)
30	−1.028	0.865	**0.196**	**0.655**	−0.050	0.501
	(−1.066, −0.954)	(0.719, 1.022)	(0.070, 0.328)	(0.535, 0.784)	(−0.156, 0.065)	(0.395, 0.611)
31	−0.915	1.149	**0.322**	**0.949**	**0.067**	**0.804**
	(−0.995, −0.773)	(1.026, 1.248)	(0.210, 0.458)	(0.839, 1.045)	(−0.040, 0.194)	(0.692, 0.902)
32	−0.865	1.017	**0.565**	**0.854**	**0.291**	**0.728**
	(−0.914, −0.714)	(0.839, 1.165)	(0.388, 0.738)	(0.707, 0.977)	(0.132, 0.451)	(0.603, 0.836)
33	−0.940	0.990	**0.312**	**0.812**	**0.049**	**0.674**
	(−0.977, −0.842)	(0.807, 1.155)	(0.172, 0.472)	(0.661, 0.957)	(−0.071, 0.194)	(0.541, 0.803)
34	−0.965	1.106	−0.016	0.985	−0.236	0.804
	(−1.007, −0.848)	(0.988, 1.193)	(−0.106, 0.145)	(0.822, 1.045)	(−0.309, −0.098)	(0.643, 0.880)
35	−0.876	0.781	**0.083**	**0.643**	−0.165	0.491
	(−0.981, −0.736)	(0.596, 0.974)	(−0.035, 0.192)	(0.458, 0.831)	(−0.260, −0.078)	(0.309, 0.678)
36	−0.843	1.223	**0.182**	**1.147**	−0.073	1.002
	(−0.955, −0.602)	(1.065, 1.305)	(0.008, 0.399)	(0.989, 1.224)	(−0.234, 0.149)	(0.845, 1.084)
37	−0.912	1.070	**0.531**	**1.005**	**0.201**	**0.934**
	(−0.953, −0.739)	(0.837, 1.311)	(0.359, 0.735)	(0.797, 1.217)	(0.064, 0.390)	(0.748, 1.140)
Treatment >15						
29	−0.972	0.927	−0.103	0.637	−0.272	0.406
	(−1.026, −0.902)	(0.744, 1.085)	(−0.268, 0.066)	(0.476, 0.777)	(−0.415, −0.121)	(0.268, 0.530)
30	−0.914	0.825	−0.005	0.577	−0.217	0.390
	(−0.968, −0.794)	(0.670, 0.976)	(−0.137, 0.147)	(0.455, 0.704)	(−0.318, −0.099)	(0.284, 0.501)
31	−0.868	1.105	−0.001	0.822	−0.199	0.620
	(−0.904, −0.754)	(0.996, 1.212)	(−0.112, 0.131)	(0.715, 0.928)	(−0.294, −0.076)	(0.515, 0.729)
32	−0.542	1.339	**0.115**	**0.928**	−0.027	0.670
	(−0.733, −0.365)	(1.180, 1.401)	(−0.038, 0.262)	(0.802, 0.994)	(−0.170, 0.109)	(0.549, 0.747)
33	−0.788	1.230	−0.130	0.855	−0.271	0.597
	(−0.932, −0.595)	(1.044, 1.327)	(−0.275, 0.022)	(0.718, 0.964)	(−0.401, −0.135)	(0.475, 0.709)
34	−0.890	1.066	−0.424	0.818	−0.562	0.556
	(−1.014, −0.727)	(0.918, 1.196)	(−0.538, −0.305)	(0.673, 0.948)	(−0.651, −0.461)	(0.411, 0.677)
35	−1.030	0.796	−0.464	0.613	−0.606	0.356
	(−1.084, −0.893)	(0.651, 0.944)	(−0.581, −0.342)	(0.468, 0.762)	(−0.703, −0.505)	(0.218, 0.504)
36	−0.915	1.040	−0.240	0.943	−0.437	0.740
	(−0.986, −0.701)	(0.850, 1.216)	(−0.353, −0.002)	(0.757, 1.115)	(−0.528, −0.201)	(0.557, 0.909)
37	−0.311	1.061	**0.018**	**0.989**	−0.145	0.752
	(−0.547, −0.105)	(0.906, 1.206)	(−0.132, 0.179)	(0.828, 1.111)	(−0.300, 0.037)	(0.608, 0.905)

Estimates use NLSY full sample. Numbers in parentheses are 95%-confidence intervals estimated using 1,000 bootstrap subsamples. **Underline** indicates sign of treatment effect is identified.

full sample. In the estimates that follow, the paper combines the exclusion-restriction assumptions with the Roy model in order to narrow the treatment effect bounds.[9] In addition, it is straightforward to combine the scalar exclusion-restriction assumptions; this effect can be obtained by taking the intersection of the bounds across each exclusion restriction imposed (Manski and Pepper (1998)). However, the paper does not combine exclusion restrictions as a multidimensional exclusion restriction. Even though this is possible using nonparametric estimation, it is not feasible given the sparseness of the data in higher dimensions of the NLSY sample.[10]

Table 6 contains bounds estimated for >12 and >15 years schooling treatments using the presence of both parents in the home until age eighteen as an exclusion restriction. The worst-case bounds using the exclusion restriction are narrower than those presented in table 3 and 4. The Roy model bounds estimated at the 10% discount rate identify a positive sign on the schooling treatment effect except at age 34. The Roy model bounds at the 5% discount rate also identify a positive sign on the schooling treatment effect at ages 31 to 33 and 37. For the >15 treatment, the worst-case bounds are 18% narrower on average than those reported in table 4. The Roy model bounds are also narrower. A positive sign is identified using the Roy model bounds assuming a 10% discount rate at ages 32 and 37.

Table 7 reports estimates for >12 and >15 treatments and college in the county of residence exclusion restriction. The worst-case lower bounds using presence of college are greater than the worst-case lower bounds reported in table 3 and 4. Worst-case exclusion restrictions on the >12 treatment using college proximity are approximately 17% narrower than those reported in table 3. These bounds are also narrower on average than the exclusion-restriction bounds

[9] Willis and Rosen (1979) include variables in their Roy model selection equation that are assumed to affect schooling choices and not wages. Using exclusion restrictions combined with the Roy model does not allow the researcher to identify which assumption is responsible for the selection of schooling treatment.

[10] This shortcoming of nonparametric estimation is known as the "curse of dimensionality." It is also the reason why the number of covariates is limited when estimating the schooling treatment effect.

TABLE 7.—NLSY SCHOOLING TREATMENT EFFECT: FOUR-YEAR COLLEGE IN COUNTY OF RESIDENCE EXCLUSION RESTRICTION
DEPENDENT VARIABLE: LOG WAGE 1993

	Worst Case		Roy Model, r = .10		Roy Model, r = .05	
Age	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Upper Bound
Treatment >12						
29	−0.966	1.096	**0.018**	**0.864**	−0.184	0.667
	(−1.066, −0.756)	(0.893, 1.148)	(−0.139, 0.277)	(0.677, 0.898)	(−0.323, 0.068)	(0.485, 0.712)
30	−0.988	1.125	**0.024**	**0.910**	−0.192	0.726
	(−1.075, −0.834)	(0.983, 1.165)	(−0.070, 0.236)	(0.768, 0.938)	(−0.266, 0.003)	(0.591, 0.759)
31	−0.810	1.247	**0.260**	**1.006**	**0.031**	**0.834**
	(−0.982, −0.592)	(1.094, 1.272)	(0.139, 0.400)	(0.874, 1.084)	(−0.089, 0.164)	(0.700, 0.922)
32	−0.649	1.202	**0.252**	**1.052**	**0.047**	**0.858**
	(−0.856, −0.440)	(1.049, 1.318)	(0.165, 0.374)	(0.917, 1.134)	(−0.060, 0.167)	(0.714, 0.985)
33	−0.922	1.251	**0.052**	**1.059**	−0.168	0.879
	(−0.985, −0.724)	(1.072, 1.287)	(−0.019, 0.244)	(0.895, 1.090)	(−0.234, −0.007)	(0.722, 0.934)
34	−0.895	1.161	−0.015	0.983	−0.233	0.801
	(−1.006, −0.692)	(0.966, 1.259)	(−0.109, 0.152)	(0.792, 1.059)	(−0.309, −0.068)	(0.609, 0.884)
35	−0.571	1.066	−0.049	0.960	−0.171	0.682
	(−0.807, −0.358)	(0.864, 1.217)	(−0.165, 0.102)	(0.789, 1.046)	(−0.321, −0.028)	(0.559, 0.818)
36	−0.374	1.001	**0.102**	**0.957**	−0.046	0.705
	(−0.654, −0.107)	(0.740, 1.252)	(−0.011, 0.274)	(0.694, 1.131)	(−0.211, 0.158)	(0.510, 0.941)
37	−0.511	0.906	**0.381**	**0.871**	**0.062**	**0.790**
	(−0.768, −0.215)	(0.571, 1.212)	(0.260, 0.510)	(0.538, 1.120)	(−0.015, 0.240)	(0.458, 0.992)
Treatment >15						
29	−0.914	1.029	−0.284	0.734	−0.438	0.488
	(−1.035, −0.693)	(0.806, 1.173)	(−0.393, −0.083)	(0.529, 0.865)	(−0.524, −0.249)	(0.289, 0.618)
30	−0.879	1.182	−0.244	0.901	−0.407	0.665
	(−0.960, −0.688)	(1.032, 1.225)	(−0.313, −0.038)	(0.754, 0.931)	(−0.458, −0.218)	(0.523, 0.699)
31	−0.717	1.327	−0.105	0.965	−0.252	0.712
	(−0.880, −0.515)	(1.226, 1.362)	(−0.214, 0.028)	(0.856, 1.043)	(−0.357, −0.134)	(0.595, 0.806)
32	−0.627	1.286	−0.100	1.005	−0.231	0.735
	(−0.811, −0.421)	(1.131, 1.394)	(−0.205, 0.026)	(0.886, 1.083)	(−0.352, −0.111)	(0.599, 0.854)
33	−0.796	1.165	−0.290	0.893	−0.447	0.650
	(−0.931, −0.617)	(0.994, 1.297)	(−0.375, −0.185)	(0.727, 1.015)	(−0.511, −0.336)	(0.487, 0.772)
34	−0.794	0.986	−0.421	0.743	−0.560	0.482
	(−0.970, −0.628)	(0.808, 1.167)	(−0.528, −0.310)	(0.556, 0.908)	(−0.644, −0.446)	(0.295, 0.609)
35	−0.677	0.985	−0.542	0.826	−0.570	0.454
	(−0.883, −0.508)	(0.819, 1.137)	(−0.659, −0.360)	(0.692, 0.851)	(−0.723, −0.402)	(0.402, 0.510)
36	−0.412	0.801	−0.233	0.725	−0.349	0.440
	(−0.654, −0.186)	(0.592, 1.024)	(−0.364, −0.081)	(0.520, 0.841)	(−0.508, −0.119)	(0.322, 0.493)
37	−0.435	0.786	−0.008	0.650	−0.239	0.482
	(−0.682, −0.193)	(0.563, 1.028)	(−0.158, 0.134)	(0.434, 0.895)	(−0.359, −0.094)	(0.261, 0.680)

Estimates use NLSY full sample. Numbers in parentheses are 95%-confidence intervals estimated using 1,000 bootstrap subsamples. **Underline** indicates sign of treatment effect is identified.

for both parents at home until age eighteen. The sign of the Roy model bounds using presence of college and a 10% discount rate is positive for all ages except 34 and 35. However, at a discount rate of 5%, a positive sign is identified only when age is equal to 31, 32, and 37. For the >15 treatment and college in the county of residence exclusion restriction, the Roy model bounds do not identify a positive sign on the schooling treatment effect. Similar to the >12 treatment, the presence of college in county of residence provide tighter bounds on average on the >15 treatment than presence of both parents in home until age eighteen.

Table 8 reports estimates for the >12 and >15 treatment and public college in the county of residence exclusion restriction. Compared with the results in table 7, public college in the county of residence does not narrow the exclusion-restriction bounds as much as presence of any four-year college. The Roy model identifies a positive sign at the 10% discount rate for most ages above 29 at the >12 treatment. The Roy model does not identify a positive sign at the >15 treatment. Taking the intersection of the bounds estimated using all of the scalar exclusion restrictions shows that the narrowest bounds are provided by the presence of a four-year college in the county of residence.

VI. Conclusions

The unified framework presented in this paper allows the estimation of the schooling treatment effect using different assumptions about selection into treatments. Unlike previous research that has relied on multiple assumptions using myriad data sets, this approach systematically varies the assumptions and compares the results. By imposing various assumptions using economic theory, it is possible to significantly narrow the bounds on the schooling treatment effect. These findings are summarized below.

- Restrictions from economic theory significantly narrow the bounds on the treatment effect.

TABLE 8.—NLSY SCHOOLING TREATMENT EFFECT: FOUR-YEAR PUBLIC COLLEGE IN COUNTY OF RESIDENCE EXCLUSION RESTRICTION, DEPENDENT VARIABLE: LOG WAGE 1993

	Worst Case		Roy Model, r = .10		Roy Model, r = .05	
Age	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Upper Bound
Treatment >12						
29	−1.013	1.050	−0.006	0.812	−0.211	0.617
	(−1.081, −0.831)	(0.843, 1.151)	(−0.144, 0.215)	(0.620, 0.901)	(−0.328, 0.002)	(0.426, 0.712)
30	−1.005	1.117	**0.098**	**0.893**	−0.117	0.707
	(−1.072, −0.835)	(0.972, 1.165)	(−0.051, 0.280)	(0.754, 0.929)	(−0.252, 0.053)	(0.575, 0.749)
31	−0.949	1.165	**0.205**	**0.998**	−0.055	0.858
	(−0.991, −0.795)	(1.023, 1.259)	(0.125, 0.372)	(0.870, 1.088)	(−0.123, 0.112)	(0.731, 0.953)
32	−0.871	1.303	**0.263**	**1.150**	−0.002	1.014
	(−0.950, −0.706)	(1.170, 1.340)	(0.156, 0.440)	(1.034, 1.194)	(−0.099, 0.171)	(0.907, 1.071)
33	−0.880	1.226	**0.129**	**1.051**	−0.117	0.897
	(−0.973, −0.737)	(1.094, 1.287)	(0.011, 0.302)	(0.932, 1.097)	(−0.212, 0.033)	(0.789, 0.935)
34	−0.965	1.222	**0.020**	**1.024**	−0.206	0.850
	(−1.003, −0.783)	(1.073, 1.262)	(−0.089, 0.208)	(0.906, 1.077)	(−0.294, −0.047)	(0.739, 0.904)
35	−0.972	1.163	−0.006	1.019	−0.237	0.849
	(−1.025, −0.813)	(1.010, 1.231)	(−0.154, 0.206)	(0.874, 1.117)	(−0.357, −0.068)	(0.718, 0.946)
36	−0.886	1.254	**0.142**	**1.172**	−0.126	1.040
	(−0.955, −0.685)	(1.116, 1.335)	(−0.012, 0.373)	(1.036, 1.259)	(−0.244, 0.066)	(0.903, 1.130)
37	−0.850	1.295	**0.385**	**1.247**	**0.085**	**1.147**
	(−0.957, −0.630)	(1.105, 1.340)	(0.253, 0.606)	(1.058, 1.293)	(−0.038, 0.287)	(0.978, 1.207)
Treatment >15						
29	−0.926	1.141	−0.202	0.838	−0.347	0.582
	(−1.011, −0.761)	(0.961, 1.195)	(−0.372, 0.006)	(0.665, 0.871)	(−0.501, −0.144)	(0.416, 0.627)
30	−0.899	1.180	−0.134	0.889	−0.294	0.649
	(−0.961, −0.761)	(1.038, 1.230)	(−0.275, 0.044)	(0.769, 0.921)	(−0.430, −0.121)	(0.531, 0.687)
31	−0.843	1.289	−0.131	0.997	−0.301	0.768
	(−0.914, −0.656)	(1.161, 1.360)	(−0.236, 0.038)	(0.874, 1.069)	(−0.397, −0.131)	(0.646, 0.850)
32	−0.854	1.335	−0.147	1.074	−0.331	0.858
	(−0.874, −0.656)	(1.207, 1.407)	(−0.227, 0.042)	(0.944, 1.146)	(−0.396, −0.139)	(0.733, 0.940)
33	−0.837	1.284	−0.231	1.006	−0.394	0.770
	(−0.917, −0.676)	(1.177, 1.331)	(−0.348, −0.049)	(0.897, 1.049)	(−0.488, −0.230)	(0.665, 0.810)
34	−0.873	1.124	−0.290	0.866	−0.450	0.625
	(−0.986, −0.700)	(1.000, 1.224)	(−0.462, −0.088)	(0.751, 0.965)	(−0.586, −0.284)	(0.510, 0.710)
35	−1.011	1.047	−0.442	0.861	−0.588	0.607
	(−1.090, −0.835)	(0.922, 1.148)	(−0.631, −0.205)	(0.725, 0.979)	(−0.744, −0.393)	(0.479, 0.709)
36	−0.963	1.224	−0.250	1.111	−0.443	0.904
	(−1.011, −0.753)	(1.073, 1.279)	(−0.393, 0.027)	(0.961, 1.179)	(−0.555, −0.209)	(0.760, 0.979)
37	−0.789	1.330	−0.016	1.191	−0.225	0.999
	(−0.894, −0.528)	(1.179, 1.376)	(−0.183, 0.273)	(1.034, 1.239)	(−0.374, 0.019)	(0.839, 1.078)

Estimates use NLSY full sample. Numbers in parentheses are 95%-confidence intervals estimated using 1,000 bootstrap sub samples. **Underline** indicates sign of treatment effect is identified.

- Both the upper and lower Roy model treatment effect bounds become larger as the discount rate increases. The higher the opportunity cost, the higher the return to schooling needed in order to select additional schooling.
- Using the Roy model, in which individuals are assumed to select schooling levels based on comparative advantage, the bounds occasionally identify a positive sign on the schooling treatment effect assuming a 10% discount rate.
- Nonparametric estimates of the treatment effect assuming exogenous selection are not constant across age. Significant differences exist between the nonparametric and linear estimates of the treatment effect. Estimates of the returns to schooling and the college wage premium that use a linear functional form are potentially biased by this assumption.
- The study uses family structure, college proximity, and school-quality measures as exclusion restrictions. Exclusion-restriction bounds are used to examine whether the exclusion-restriction assumptions are consistent with the data.
- Holding other wage covariates besides age constant, this study rejects the validity of school-quality measures as exclusion restrictions, indicating school quality might be correlated with earnings.
- Exclusion restrictions tighten the bounds relative to the worst-case and Roy model bounds. In some cases, the exclusion-restriction bounds identify a positive treatment effect.

When estimating the effect of schooling on earnings, the selection problem requires the researcher to make explicit assumptions about assignment to schooling levels. This paper has presented a number of striking results using alternative assumptions about selection. Clearly, the worst-case bounds provide little insight into the schooling treatment effect. By definition, the ordered-outcome bounds identify a positive sign with a large upper bound. The Roy model bounds identify a positive sign for the >12 treatment;

these bounds become tighter when combined with exclusion restrictions. The treatment effect bounds demonstrate the limits of understanding of the effect of schooling on earnings. In the end, researchers must make some kind of simplifying assumptions to estimate the effect of schooling on earnings, knowing that policymakers' decisions will be based on these estimates. They must choose either to ignore self-selection by assuming exogenous assignment to schooling levels or to incorporate economic theory on the self-selection of schooling levels. As shown in this research, these assumptions provide very different results.

REFERENCES

Angrist, Joshua D., and Guido W. Imbens, ''Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity,'' *Journal of the American Statistical Association* 90 (1995), 431–442.

Balanced Budget Act of 1997, Public Law No. 105–34.

Balke, Alexander, and Judea Pearl, ''Nonparametric Bounds on Causal Effects from Partial Compliance Data,'' Technical Report R-199, University of California, Los Angeles, Computer Science Department (1994).

Betts, Julian R., ''Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth,'' this REVIEW 77(2) (1995), 231–250.

Card, David, ''Earnings, Schooling, and Ability Revisited,'' in *Research in Labor Economics,* vol. 14 (Greenwich, CT: JAI Press, 1995a) 23–48.

——— ''Using Geographic Variation in College Proximity to Estimate the Return to Schooling,'' in L.N. Christofides, E.K. Grant, and R. Swidinsky (Eds.), *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp* (Toronto: University of Toronto Press, 1995b), 201–222.

Card, David, and Alan Krueger, ''Does School Quality Matter: Returns to Education and the Characteristics of Public Schools in the United States,'' *Journal of Political Economy* 100(1) (1992), 1–40.

Griliches, Zvi, ''Estimating the Returns to Schooling: Some Econometric Problems,'' *Econometrica* 45 (1977), 1–22.

Heckman, James J., and Bo E. Honore, ''The Empirical Content of the Roy Model,'' *Econometrica* 58(5) (1990), 1121–1149.

Heckman, James J., Anne Layne-Farrar, and Petra Todd, ''Human Capital Pricing Equations with an Application to Estimating the Effect of School Quality on Earnings,'' this REVIEW 58 (1996), 562–610.

Imbens, Guido W., and Joshua D. Angrist, ''Identification and Estimation of Local Average Treatment Effects,'' *Econometrica,* 62 (1994), 467–475.

Imbens, Guido W., and Donald B. Rubin, ''On the Fragility of Instrumental Variables Estimators,'' Harvard University Discussion Paper No. 1675 (1994).

Manski, Charles F., ''Anatomy of the Selection Problem,'' *Journal of Human Resources* 24(3) (1989), 343–360.

——— ''Nonparametric Bounds on Treatment Effects,'' *American Economic Review, Papers and Proceedings* 80 (1990), 319–323.

——— ''The Selection Problem,'' in Christopher Sims, *Advances in Econometrics* (New York: Cambridge University Press, 1994), 143–170.

——— *Identification Problems in the Social Sciences* (Cambridge, MA: Harvard University Press, 1995).

Manski, Charles F., and John V. Pepper, ''Monotone Instrumental Variables with an Application to the Returns to Schooling,'' NBER Technical Working Paper, no. 224 (February 1998).

Mincer, Jacob, *Schooling, Experience, and Earnings* (New York: National Bureau of Economic Research, 1974).

Park, Jin Heum, ''Estimation of Sheepskin Effects and Returns to Schooling Using the Old and the New CPS Measures of Educational Attainment,'' working paper, Princeton University (1994).

Rosenbaum, Paul R., and Donald B. Rubin, ''Assessing the Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome,'' *Journal of the Royal Statistical Society, Series B,* 45(2) (1983), 212–218.

Rubin, Donald B., ''Bayesian Inference for Causal Effects,'' *The Annals of Statistics* 6(1) (1978), 34–58.

Silverman, B. W., *Density Estimation for Statistics and Data Analysis* (New York: Chapman and Hall, 1986).

Vella, Francis, and Marno Verbeek, ''Estimating and Interpreting Models with Endogenous Treatment Effects,'' Rutgers University working paper (1997).

Willis, Robert J., and Sherwin Rosen, ''Education and Self-Selection,'' *Journal of Political Economy* 87(5) (1979) S7–S37.

APPENDIX

TABLE A.1.—WORST-CASE BOUNDS USING DIFFERENT TRIMMING ASSUMPTIONS DEPENDENT VARIABLE: LOG WAGE 1993

	$\alpha = 2.5\%$		$\alpha = 5\%$		$\alpha = 10\%$	
Age	Lower Bound $K_L = 5.234$	Upper Bound $K_U = 7.478$	Lower Bound $K_L = 5.521$	Upper Bound $K_U = 7.478$	Lower Bound $K_L = 5.691$	Upper Bound $K_U = 7.191$
Treatment >12						
29	−1.111	1.134	−0.963	0.994	−0.745	0.755
30	−1.111	1.133	−0.983	0.974	−0.762	0.737
31	−0.986	1.259	−0.928	1.029	−0.733	0.767
32	−0.929	1.316	−0.941	1.016	−0.687	0.813
33	−0.974	1.270	−0.900	1.057	−0.672	0.827
34	−0.993	1.251	−0.883	1.074	−0.683	0.817
35	−1.038	1.206	−0.917	1.040	−0.735	0.765
36	−0.928	1.317	−0.829	1.128	−0.684	0.816
37	−0.942	1.303	−0.867	1.090	−0.664	0.836
Treatment >15						
29	−1.069	1.176	−0.908	1.049	−0.710	0.790
30	−1.044	1.200	−0.893	1.064	−0.694	0.805
31	−0.896	1.349	−0.780	1.177	−0.623	0.877
32	−0.848	1.396	−0.772	1.185	−0.647	0.853
33	−0.931	1.313	−0.794	1.163	−0.641	0.858
34	−1.007	1.238	−0.832	1.125	−0.665	0.835
35	−1.122	1.123	−0.930	1.026	−0.762	0.738
36	−0.986	1.259	−0.855	1.102	−0.689	0.811
37	−0.886	1.358	−0.764	1.193	−0.571	0.928

Estimates use NLSY full sample. Worst-case bounds are by definition, $K_U - K_L$ wide.

TABLE A.2.—ESTIMATES OF CONDITIONAL PROBABILITIES AND EXPECTATIONS

Age	NLSY, Schooling >12, =12			NLSY, Schooling >15, <16		
	$P(s = 1 \mid x)$	$E(y_1 \mid x, s = 1)$	$E(y_0 \mid x, s = 1)$	$P(s = 1 \mid x)$	$E(y_1 \mid x, s = 1)$	$E(y_0 \mid x, s = 1)$
29	0.496	6.212	6.192	0.346	6.210	6.197
30	0.526	6.232	6.195	0.397	6.291	6.184
31	0.647	6.493	6.221	0.423	6.659	6.186
32	0.676	6.585	6.236	0.465	6.773	6.207
33	0.567	6.517	6.226	0.368	6.667	6.235
34	0.549	6.518	6.267	0.325	6.605	6.305
35	0.536	6.596	6.453	0.281	6.678	6.481
36	0.634	6.743	6.495	0.476	6.811	6.509
37	0.773	6.626	6.480	0.532	6.799	6.354

Estimates use NLSY full sample.