NBER TECHNICAL PAPER SERIES

AVERAGE CAUSAL RESPONSE WITH VARIABLE TREATMENT INTENSITY

Joshua D. Angrist

Guido W. Imbens

Technical Paper No. 127

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1992

We thank Don Rubin for raising the question that this paper answers, and seminar participants at the Harvard-MIT Econometrics Workshop and the University of Wisconsin for helpful comments. Financial support from the National Science Foundation and the Nederlandse organisatie voor Wetenschappelijk Onderzoek is also gratefully acknowledged. Part of the paper was written while the second author was visiting the Hebrew University Department of Economics. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Paper #127
September 1992

AVERAGE CAUSAL RESPONSE WITH VARIABLE TREATMENT INTENSITY

ABSTRACT

In evaluation research, an average causal effect is usually defined as the expected difference between the outcomes of the treated, and what these outcomes would have been in the absence of treatment. This definition of causal effects makes sense for binary treatments only. In this paper, we extend the definition of average causal effects to the case of variable treatments such as drug dosage, hours of exam preparation, cigarette smoking, and years of schooling. We show that given mild regularity assumptions, instrumental variables independence assumptions identify a weighted average of per-unit causal effects along the length of an appropriately defined causal response function. Conventional instrumental variables and Two-Stage Least Squares procedures can be interpreted as estimating the average causal response to a variable treatment.

Joshua D. Angrist
Department of
Economics
Hebrew University
Jerusalem 91905
Israel
011 972 2 8831
KEUJA@HUJIVMI
and NBER

Guido W. Imbens
Department of
Economics
Harvard University
Cambridge, MA 02138
USA
2(617) 495-4129
guido@haavelmo.harvard.edu

1. Introduction

For a variety of ethical and practical reasons, empirical researchers have long been interested in alternatives to evaluation designs based on random assignment. In medicine, the use of random assignment to evaluate drug efficacy and medical interventions may require that potentially beneficial treatments be denied to seriously ill patients. Some physicians argue that denial of a potentially beneficial treatment violates the "Patient Care Principle" in medical ethics (Royall 1991). In the social sciences, randomization of treatment raises some of the same ethical issues as in medicine, and may be impractical because of cost considerations, or because of political resistance to the randomization of social policy interventions (Manski and Garfinkel 1991.)

Although random assignment of treatment is often represented as an ideal research design, credible alternative designs and statistical methodologies may be available. Hearst, Newman, and Hulley (1986) used the randomly assigned *priority* for conscription generated by the Vietnam-era draft lottery to estimate the effects of military service on the subsequent mortality of veterans. Angrist (1990) used the lottery to study the effect of military service on the civilian earnings of veterans. In a study of the effects of test-preparation on Graduate Record Examination (GRE) scores, a randomly chosen group of test-takers was encouraged, but not compelled, to prepare in advance for the GRE (Powers and Swinton 1984.) A fourth example is a study of the effect of maternal smoking on birth weight, in which a randomly selected sample of pregnant smokers was enrolled in a course designed to encourage participants to reduce or quit smoking (Permutt and Hebel 1984, 1989.) Finally, in clinical trials, patients randomly assigned to the treatment group may decline treatment, and some members of the control group may be able to

obtain the treatment on their own.

A clinical trial with partial compliance is sometimes referred to as randomization of "intention-to-treat" (Efron and Feldman 1991.) In his discussion of the Powers and Swinton (1984) test preparation study, Holland (1988) refers to the randomization of encouragement or assistance as an "encouragement design." The smoking study by Permutt and Hebel also fits into this category. Each of these examples is actually a special case of the instrumental variables framework widely used in observational studies in econometrics. In estimation based on the draft lottery, the randomly assigned lottery number is an instrument for whether men born in certain years served in the military. More generally, instrumental variables techniques can be used to evaluate the effect of an intervention whenever a variable can be found (the "instrument") that is associated with the outcome of interest solely by virtue of its association with treatment assignment.

Whether couched in the econometrician's language of instrumental variables or not, most of the literature on causal effects in evaluation research is concerned with estimating the average effect of a binary treatment. Since Rubin's (1974, 1977) influential formulation of the problem of causal inference, causal effects in statistics have usually been defined as the average difference between the outcomes of the treated and what these outcomes would have been in the absence of treatment (Holland 1986).[1] In a recent paper (Imbens and Angrist 1991), we extend the definition of causal

[1]See, for example, Heckman (1990), Manski (1991), and Angrist and Imbens (1991). Rubin (1974, 1977) and Angrist (1991b) are concerned with the average causal effect of binary treatment in an entire population. Heckman and Robb (1985) discuss both of these types of causal effects in the context of linear econometric models. An important early formulation of the problem of causal inference is Roy (1951).

effects to include the notion of a Local Average Treatment Effect (LATE). LATE is the average effect of treatment for those whose treatment status is affected by exogenous variation in some third variable. For example, in the smoking study, LATE is the average effect of maternal smoking on birth weight for babies whose mothers quit or reduced their smoking as a consequence of counseling and assistance. In the test-preparation study, LATE is the effect of test preparation for students whose studying behavior was influenced by the randomly assigned encouragement intervention. Finally, in Angrist's (1990) study of the draft lottery, LATE is the effect of veteran status for men who served in the military as a consequence of their draft lottery number.

In our previous paper, we showed that LATE for a binary treatment is identified under a mild regularity condition satisfied in a wide range of models and circumstances. Essentially, this condition requires that the instrumental variable affect treatment status in a monotone way. In the draft lottery example, we require that men with lottery numbers putting them at risk of conscription are at least as likely to serve as they would have been had they had lottery numbers exempting them from conscription.[2] When the monotonicity condition is satisfied, LATE is identified and can be estimated using conventional linear instrumental variables and Two-Stage Least Squares (TSLS) estimators.

The purpose of this paper is to extend the definition of LATE and the corresponding identification and estimation results to the case of variable treatments. For example, instead of estimating the effect of a certain drug

[2]Cigarette smoking is not a binary treatment and so LATE is not defined in the Permutt and Hebel (1989) smoking study. But the monotonicity condition can still be defined; it requires that women given anti-smoking counseling smoke *no more* than they would have without the counseling. Permutt and Hebel suggest this condition in an informal discussion of their results.

treatment regime on health, we are interested in estimating features of the entire dose-response function. The methodological points are illustrated through two empirical examples. First, we briefly discuss the results in Powers and Swinton's (1984) study of encouraged preparation for the GRE. Our second illustration is based on a study of compulsory school attendance by Angrist and Krueger (1991), which showed how compulsory attendance laws interact with students' quarter of birth to induce exogenous variation in years of schooling. Angrist and Krueger used this exogenous variation in schooling to estimate the effect of schooling in econometric earnings functions. The causal effect identified in each of these examples is a weighted average of points along the response function that would be identified if treatment were randomly assigned.

2. Causal Effects

To fix ideas, we continue to refer to the study of the effect of test preparation on GRE scores. The intervention in this case was the mailing by the Educational Testing Service of test-preparation materials such as practice exams, and a strongly worded letter designed to encourage the use of these materials. The treatment of interest, however, was not the receipt of preparation materials or the encouragement to use them, but the actual number of hours of preparation.

Let $Y_j \in \mathcal{X}$ be the exam score given j hours of preparation, for $j = 0, 1, 2, \ldots, J$. We assume that the Y_j are well-defined and that a full set exists for each person, even though only one of the Y_j is actually observed. We also define $D_z \in \{0, 1, 2, \ldots, K\}$ for $Z \in \{0, 1\}$ to be the number of hours of preparation by a test candidate conditional on the indicator for

whether he or she received encouragement, Z. As with Y_j, D_Z is assumed to exist for each value of Z for each person even though only one D_Z is observed. This setup is an innovation to the framework outlined by Rubin (1974, 1977) because the Rubin framework is limited to counter-factual outcomes and binary treatments.

Note that Y_0 is the test score of someone who doesn't prepare for the GRE. One of the causal effects we are interested in is $Y_1 - Y_0$, the effect of preparing one hour rather than zero hours. Given J+1 levels of treatment, there are $(J+1 \times J)/2$ possible treatment effects, $Y_j - Y_i$, each of which can be expressed in terms of the J linearly independent treatment effects for a unit increase in treatment level, $Y_j - Y_{j-1}$. The sequence of $Y_j - Y_{j-1}$ defines the true causal response function for each individual.

If the treatment level, D, is randomly assigned, then $E[Y_j - Y_{j-1}]$ can be estimated by subtracting the average response for individuals with treatment level $j-1$ from the average response for individuals with treatment level j. We assume that the level of treatment (hours of preparation) is not randomly assigned, but determined at least partly on the basis of information unavailable to the researcher. Because this information may also be related to outcomes, comparisons of average outcomes for different treatment levels do not consistently estimate the effect of a unit increase in treatment.

Initially we assume that Z can take on only two values, 0 and 1, indicating assignment to the encouragement intervention or not.[3] D_0 is the hours of preparation when not encouraged and D_1 is the hours of preparation if encouraged. For each person in the sample of test candidates, we observe the

[3]The test-preparation study involved the random assignment of 4 different types of preparation materials as well as encouragement to prepare for the test, in a 5×2 factorial design.

triple (Z,D,Y), where Z is the level of encouragement, $D = D_Z = Z \cdot D_1 + (1 - Z) \cdot D_0$ is the hours of preparation, and $Y = Y_D$ is the candidate's test score. Our principal identifying assumption (apart from assuming the existence of Y_j) is that Z is independent of all potential outcomes and potential treatment intensities. In the test-preparation example, this assumption is satisfied because encouragement is randomly assigned. Formally, we have:

Assumption 1 (Independence).
The random variables D_0, D_1, Y_0, Y_1, . . ., Y_J are jointly independent of Z.

It is important to note that this assumption alone is not enough to identify a meaningful average treatment effect. Example 1 in Imbens and Angrist (1991) shows that treatment effect heterogeneity can make comparisons of people by intervention status (Z) meaningless. In this example $E[Y|\ Z]$, $E[Y|D]$, and $E[Y|\ D, Z]$ are all constant even though $E[D|\ Z]$ varies with Z and the treatment effect for every individual is strictly positive. Therefore, there is no way to estimate an average treatment effect from the observed distribution of Y. The intuition for the result in this example is that while the instrument causes a large group of people with small treatment effects to shift from non-treatment to treatment, a small group with large treatment effects is induced to leave treatment. On average, effects in the two groups cancel each other out even though the instrument is correlated with treatment status and all treatment effects are positive.

The most common way to get around this problem is simply to assume a constant treatment effect, $Y_j - Y_{j-1} = \alpha$ for all j and all individuals.

This is the assumption underlying most econometric applications using linear regression models, as well as the application of instrumental variables techniques by Permutt and Hebel (1989). In his comment on Holland's (1988) discussion of causality, Leamer (1988) points out that given an independence assumption such as Assumption 1, the problem of causal inference is trivial in linear models with constant treatment effects. We believe the importance of the Rubin's "counter-factual" approach to causal inference is that in this framework, treatment effect heterogeneity arises naturally from the assumption of counter-factual individual outcomes. Use of a model with heterogeneous treatment effects therefore helps clarify the definition of causality that motivates evaluation research.[4]

Instead of restricting treatment effect heterogeneity, in this paper we impose a non-parametric restriction on the process determining D as a function of Z.[5] This restriction can be characterized in the test-preparation example as follows: We can allow the encouragement intervention to lead to different increases in test-preparation time for different people, and we can allow the intervention to have no effect for some people. But we assume that the intervention never leads to *fewer* hours of test preparation. More generally, we make the following monotonicity assumption:

---

[4]Aigner and Zellner (1988), Holland (1986), and Rubin (1990) survey alternative frameworks for causal inference.

[5]Elsewhere (Angrist and Imbens 1991), we discuss bounds on treatment effects attainable by imposing *a priori* restrictions on the difference between two alternative weighted average treatment effects. A variety of other results on non-parametric bounds for treatment effects are given by Manski (1992.)

Assumption 2 (Monotonicity).

$\Pr(D_1 - D_0 \geq 0) = 1$ or $\Pr(D_1 - D_0 \leq 0) = 1$.

This means that either $D_1 - D_0 \geq 0$ for each person or $D_1 - D_0 \leq 0$ for each person. Assumption 2 is not verifiable, since it involves unobserved variables (only one of D_1 or D_0 is observed.) Nevertheless, if $J > 1$, Assumption 2 has the testable implication that the cumulative distribution function (CDF) of D given $Z = 1$ and the CDF of D given $Z=0$ should not cross.[6] If $J = 1$, the CDF's cannot cross because the treatment is binary. In section 4, below, we compare empirical CDF's given Z in two examples.

The main theoretical result of the paper is given below for the case where $D_1 - D_0 \geq 0$:

Theorem 1. Suppose that assumptions 1 and 2 hold and that $\Pr(D_1 \geq j > D_0) > 0$ for at least one j. Then,

$$(1) \quad \frac{E[Y|Z=1]-E[Y|Z=0]}{E[D|Z=1]-E[D|Z=0]} = \sum_{j=1}^{J} \omega_j \cdot E[Y_j - Y_{j-1} | D_1 \geq j > D_0] = \beta$$

where

$$\omega_j = \frac{\Pr(D_1 \geq j > D_0)}{\sum_{i=1}^{J} \Pr(D_1 \geq i > D_0)}$$

which implies that $0 \leq \omega_j \leq 1$ and $\sum_{j=1}^{J} \omega_j = 1$,

so that β is a weighted average per-unit treatment effect that can be estimated from a sample of (Y, D, Z).

---

[6] If $D_1 \geq D_0$ then $\Pr(D_1 \geq j) \geq \Pr(D_0 \geq j)$ for all j. This implies $\Pr(D \geq j | Z=1) \geq \Pr(D \geq j | Z=0)$ or $F_D(j | Z=0) \geq F_D(j | Z=1)$ where F_D is the CDF of D.

Proof: Let I(A) be the indicator function for the event A. Define the following indicators: $\delta_{Zj} = I(D_Z = j)$ for $Z = 0,1$ and $j = 0,1,\ldots,J$; and $\lambda_{Zj} = I(D_Z \geq j)$ for $Z = 0,1$ and $j = 0,1,2,\ldots,J+1$. Note that $\lambda_{Z0} = 1$ and $\lambda_{ZJ+1} = 0$ for all Z. The indicators δ and λ are related by the equations $\lambda_{Zj} = \sum_{i=j}^{J} \delta_{Zj}$ and $\delta_{Zj} = \lambda_{Zj} - \lambda_{Zj+1}$ for $Z = 0,1$ and $j = 0,1,\ldots,J$.

In terms of the δ_{Zj}, Y can be written as:

$$Y = Z \cdot Y_{D_1} + (1 - Z) \cdot Y_{D_0} = \left(Z \cdot \sum_{j=0}^{J} Y_j \cdot \delta_{1j} \right) + \left((1 - Z) \cdot \sum_{j=0}^{J} Y_j \cdot \delta_{0j} \right)$$

Therefore,

$$E[Y|Z = 1] - E[Y|Z = 0]$$

$$= E\left(\sum_{j=0}^{J} Y_j \cdot \delta_{1j} \,\middle|\, Z = 1 \right) - E\left(\sum_{j=0}^{J} Y_j \cdot \delta_{0j} \,\middle|\, Z = 0 \right)$$

Using the independence assumption, this can be written

$$E\left(\sum_{j=0}^{J} Y_j \cdot [\delta_{1j} - \delta_{0j}] \right)$$

$$= E\left(\sum_{j=0}^{J} Y_j \cdot [\lambda_{1j} - \lambda_{1j+1} - \lambda_{0j} + \lambda_{0j+1}] \right)$$

$$= E\left(\sum_{j=1}^{J} [(Y_j - Y_{j-1}) \cdot (\lambda_{1j} - \lambda_{0j})] + Y_0 \cdot (\lambda_{10} - \lambda_{00}) \right)$$

which reduces to

$$= E\left(\sum_{j=1}^{J} (Y_j - Y_{j-1}) \cdot (\lambda_{1j} - \lambda_{0j}) \right)$$

because $\lambda_{Z0} = 1$ for $Z = 0,1$. Now, note that $\lambda_{1j} \geq \lambda_{0j}$ by assumption 2 and that λ_{1j} and λ_{0j} equal zero or one. Therefore, $\lambda_{1j} - \lambda_{0j}$ equals zero or one, and we

can write the previous expression as

$$\sum_{j=1}^{J} E[Y_j - Y_{j-1} \mid \lambda_{1j} - \lambda_{0j} = 1] \cdot \Pr(\lambda_{1j} - \lambda_{0j} = 1)$$

$$= \sum_{j=1}^{J} E[Y_j - Y_{j-1} \mid D_1 \geq j > D_0] \cdot \Pr(D_1 \geq j > D_0). \tag{2}$$

Now we turn to the denominator of (1):

$$D = Z \cdot D_1 + (1 - Z) \cdot D_0 = (Z \cdot \sum_{j=0}^{J} j \cdot \delta_{1j}) + ((1 - Z) \cdot \sum_{j=0}^{J} j \cdot \delta_{0j}).$$

Therefore,

$$E[D|Z = 1] - E[D|Z = 0] = E\left(\sum_{j=0}^{J} j \cdot \delta_{1j} \mid Z = 1\right) - E\left(\sum_{j=0}^{J} j \cdot \delta_{0j} \mid Z = 0\right).$$

Again, using the independence assumption, this equals

$$= E\left(\sum_{j=0}^{J} j \cdot (\delta_{1j} - \delta_{0j})\right).$$

Substituting for $\lambda_{zj} - \lambda_{zj+1}$ for δ_{zj} this can be rewritten as

$$E\left\{\sum_{j=0}^{J} j \cdot (\lambda_{1j} - \lambda_{1j+1} - \lambda_{0j} + \lambda_{0j+1})\right\} = E\left\{\sum_{j=1}^{J} (\lambda_{1j} - \lambda_{0j})\right\}$$

$$= \sum_{j=1}^{J} \Pr(D_1 \geq j > D_0). \qquad \blacksquare$$

The requirement that $\Pr(D_1 \geq j > D_0) > 0$ for some j means that the instrument must affect the level of treatment, D. Also, note that in the proof of Theorem 1, D is assumed to take on only integer values between 0 and J. The only restriction necessary, however, is that D be bounded and take on a finite number of rational values. Then one can always use a linear transformation to ensure that D takes on only integer values between 0 and J.

A linear transformation of D does not have any effect on the numerator of the ACR. The denominator is multiplied by a constant. The linear transformation therefore amounts to changing the units in which treatment intensity is measured.

Theorem 1 is important because it shows that in a wide variety of models and circumstances, it is possible to identify features of the distribution of $Y_j - Y_{j-1}$. For example, the monotonicity assumption appears plausible in research designs based on the draft lottery, and in designs based on randomly assigned encouragement or intention-to-treat; the monotonicity assumption is also mechanically satisfied in the latent index models commonly employed in econometrics (Imbens and Angrist 1991.) We refer to the parameter β as the Average Causal Response (ACR.) This parameter captures a weighted average of causal responses to a unit change in treatment, *for those whose treatment status is affected by the instrument*. Note that this group need not be representative of the population.

The weight attached to the average of $Y_j - Y_{j-1}$ is proportional to the number of people who, because of the instrument, change their treatment from less than j units to j or more units. This proportion is $Pr(D_1 \geq j > D_0)$. In the test-preparation example, this is the proportion of people who study at least j hours when encouraged, but would study less than j hours if not encouraged. These weights can be estimated using a random sample of (Y,D,Z) because

$$Pr(D_1 \geq j > D_0) = E(\lambda_{1j} - \lambda_{0j})$$

$$= Pr(D_1 \geq j) - Pr(D_0 \geq j) = Pr(D_0 < j) - Pr(D_1 < j)$$

$$= Pr(D < j \mid Z=0) - Pr(D < j \mid Z=1).$$

Thus, the weighting function is just the difference between the empirical CDF's of D given Z.

A natural estimator of β is the sample analog of the left hand side of (1). This estimator is an application of Wald's (1940) grouping method of fitting straight lines, where the data have been grouped by the instrument. Durbin (1954) appears to have been the first to point out that the Wald estimator is also an instrumental variables estimator.

We conclude this section with a corollary that can be used to interpret parameter estimates in models where a variable treatment is incorrectly parameterized as a binary treatment. For example, Permutt and Hebel (1989) discuss conditions sufficient to identify the effect of smoking when it is assumed that all that matters for health is whether any cigarettes are smoked. Similarly, econometricians sometimes estimate the effect of college and/or high school graduation on earnings, ignoring the fact that dummy variables indicating graduation are nonlinear functions of an underlying years-of-schooling variable (e.g., Rosen and Willis 1979.)

The corollary is based on the smoking example, and shows that Wald estimates constructed by treating cigarette-smoking as a binary treatment have a probability limit proportional to the ACR. The factor of proportionality is greater than or equal to one.

Corollary (Mis-specified binary treatment.) Suppose that the treatment of interest is assumed to be an indicator function of D, say $d = \lambda_{Z1} = I(D_Z \geq 1)$. Then, given Assumptions 1 and 2,

$$(3) \qquad \frac{E[Y|Z=1]-E[Y|Z=0]}{E[d|Z=1]-E[d|Z=0]} = \phi \cdot \beta = \beta_b.$$

where

$$\phi = \frac{E[D|Z=1] - E[D|Z=0]}{E[d|Z=1] - E[d|Z=0]} = \frac{\sum_{j=1}^{J} Pr(D_1 \geq j > D_0)}{Pr(D_1 \geq 1 > D_0)},$$

so that $\phi \geq 1$.

Proof: To establish the formula for β_b, note that the numerator is the same as in Theorem 1. The denominator can be written

$$E[\lambda_{11}|\ Z=1] - E[\lambda_{01}|\ Z=0] = E[\lambda_{11} - \lambda_{01}] = Pr(D_1 \geq 1 > D_0).$$

That $\phi \geq 1$ is immediate from the formula for ϕ. In fact, the only situation where $\phi = 1$ is when then the instrument has no effect other than to cause people to switch from $D = 0$ to $D = 1$. ■

Thus, when a variable treatment is incorrectly parameterized as binary, the resulting estimate tends to be too large relative to the average per-unit effect along the length of the response function. On the other hand, the sign of the ACR is still identified. This result is similar to the conventional omitted-variables bias formula in a regression where the omitted variables are actually functions of the treatment intensity other than the indicator function, d.

3. Multiple Instruments

In many empirical applications, a number of instrumental variables are available. For example, the experiment designed to encourage test preparation

involved the random assignment of 5 different types of test preparation material (the fifth type was no material), as well as a letter encouraging the use of these materials. The act of sending materials without a letter of encouragement also led to an increase in hours of exam preparation. Assuming that both encouragement and the sending of materials have no effect other than to increase the number of hours of preparation, the interaction of 5 types of preparation material with the encouragement letter in a factorial design generates 9 potential instrumental variables.

The typical econometric application of instrumental variables techniques imposes a constant-treatment-effect model, in which $Y_j - Y_{j-1} = \alpha$ for all j and all individuals. In this case, alternative instrumental variables estimates of the same α can be combined into a single more efficient estimate using Two-Stage Least Squares (TSLS.) In fact, one interpretation of TSLS in the constant treatment effect model is that it is an instrumental variables estimator where the instrument being used is the fitted value from a regression of D on all the possible instruments.

The discussion in the previous section suggests that estimates of β constructed using different instruments should be expected to differ. This is because different instruments are associated with different weighting schemes in the definition of the ACR. What does the TSLS estimator -- which combines alternative instrumental variables estimates -- produce when it is applied to the heterogeneous-treatment-effects model outlined in Section 2? We explore this question for the case where K mutually orthogonal binary instruments are combined to form a single TSLS estimate. This is a fairly general example because any set of discrete instruments can be recoded as a set of mutually exclusive indicator variables. Alternately, TSLS using K orthogonal

indicators can be thought of as a means of exploiting a single K+1-valued instrument, W. For example, in the test-preparation experiment, W indexes the 10 treatment and control groups.

In general, a K+1-valued instrument can be used to form $(K+1)\times K/2$ ACR's, defined as:

$$\beta_{k\ell} = \frac{E[Y|\ W=k] - E[Y|\ W=\ell]}{E[D|\ W=k] - E[D|\ W=\ell]}; \quad \text{for } k \neq \ell.$$

We assume that each underlying binary instrument affects treatment the so that the denominators are non-zero. Only K of the $\beta_{k\ell}$ are linearly independent and the different ACR's are related as follows:

$$\beta_{k\ell} = \frac{E[D|\ W=k] - E[D|\ W=m]}{E[D|\ W=k] - E[D|\ W=\ell]}\beta_{km} + \frac{E[D|\ W=m] - E[D|\ W=\ell]}{E[D|\ W=k] - E[D|\ W=\ell]}\beta_{m\ell}.$$

Theorem 2 below shows that the TSLS estimator constructed by using K linearly independent dummy variables, $\delta_k = I(W = k)$, plus a constant as instruments is a weighted average of the K linearly independent ACR's, $\beta_{k,k-1}$. Since each of the $\beta_{k,k-1}$ is a weighted average of points on the causal response function, the TSLS estimate also converges to a weighted average of points on the causal response function.

Let the points of support of W be ordered such that $\ell < m$ implies $E[D|\ W=\ell] < E[D|\ W=m]$. Finally, note that using K dummies, $\delta_k = I(W = k)$, plus a constant in TSLS estimation is the same thing as instrumental variables estimation using $E[D|\ W]$ plus a constant as instruments. Then we have:

Theorem 2. Suppose that $E[D|\ W]$ and a constant are used as instruments to construct instrumental variables estimates of β_W in the equation

$$Y = \gamma + \beta_W D + \epsilon. \tag{4}$$

The resulting estimate has probability limit

$$\beta_W = \frac{E(Y \cdot (E[D|\ W] - E[D]))}{E(E[D|W] \cdot (E[D|W] - E[D]))} = \sum_{k=1}^{K} \mu_k \beta_{k,k-1}, \tag{5}$$

where $\mu_k = (E[D|W=k] - E[D|W=k-1]) \cdot \dfrac{\sum_{\ell=k}^{K} \pi_\ell\ (E[D|W=\ell] - E[D])}{\sum_{\ell=0}^{K} \pi_\ell\ E[D|W=\ell](E[D|W=\ell] - E[D])}$

and $\pi_\ell = \Pr[W=\ell]$. Moreover, $1 \geq \mu_k \geq 0$ and $\sum_{k=1}^{K} \mu_k = 1$.

Proof: The denominator of the formula for μ_k is the same as the denominator of the expression for β_W. To evaluate the numerator, we can write

$$E[Y|\ W=\ell] = \beta_{\ell,\ell-1}\ (E[D|W=\ell] - E[D|W=\ell-1]) + E[Y|\ W=\ell-1] \tag{6}$$

$$= \sum_{k=1}^{\ell} \beta_{k,k-1}\ (E[D|W=k] - E[D|W=k-1]) \ + \ E[Y|\ W=0]$$

and

$$E(Y \cdot (E[D|\ W] - E[D])) = E(E[Y|\ W=\ell] \cdot (E[D|\ W=\ell] - E[D])). \tag{7}$$

Substituting (6) for $E[Y|\ W=\ell]$ in (7), the numerator is

$$\sum_{\ell=0}^{K} \sum_{k=1}^{\ell} \pi_\ell (E[D|\ W=\ell] - E[D])\ \beta_{k,k-1} (E[D|W=k] - E[D|W=k-1])$$

$$= \sum_{\ell=0}^{K} \sum_{k=1}^{K} I(1 \leq k \leq \ell) \cdot \pi_\ell (E[D|\ W=\ell] - E[D])\ \beta_{k,k-1} (E[D|W=k] - E[D|W=k-1])$$

$$= \sum_{k=1}^{K} \sum_{\ell=k}^{K} \pi_\ell (E[D|\ W=\ell] - E[D])\ \beta_{k,k-1} (E[D|W=k] - E[D|W=k-1]).$$

This establishes the right hand side of (5). The weights, μ_j, are non-negative because the points of support of W are ordered such that $E[D|\ W=k] > E[D|\ W=k-1]$. To show that the weights sum to one, note that the sum of the numerator of the μ_k's is

$$- \sum_{k=1}^{K} \sum_{\ell=k}^{K} \pi_\ell (E[D|\ W=\ell]-E[D])\ (E[D|W=k]-E[D|W=k-1]).$$

Reversing the order of summation as before, this equals

$$\sum_{\ell=1}^{K} \sum_{k=1}^{\ell} \pi_\ell (E[D|\ W=\ell]-E[D])\ (E[D|W=k]-E[D|W=k-1]). \tag{8}$$

Now,

$$\sum_{k=1}^{\ell} (E[D|W=k]-E[D|W=k-1]) = E[D|\ W=\ell] - E[D|\ W=0],$$

so that (8) can be written

$$\sum_{\ell=1}^{K} \pi_\ell (E[D|\ W=\ell]-E[D])\ (E[D|\ W=\ell] - E[D|\ W=0]). \tag{8}$$

This equals

$$\sum_{\ell=0}^{K} \pi_\ell\ (E[D|\ W=\ell]-E[D])\ E[D|W=\ell] \tag{9}$$

because

$$\sum_{\ell=1}^{K} \pi_\ell (E[D|\ W=\ell]-E[D])\ E[D|\ W=0] = -\pi_0(E[D|\ W=0]-E[D])\ E[D|\ W=0]$$

Expression (9) is the same as the denominator of μ_k. ■

Theorem 2 provides a useful interpretation for conventional TSLS estimates. Just as the simple Wald estimator of Theorem 1 provides a weighted average effect along the length of the causal response function, TSLS estimates provide one way of combining a set of different weighted average

effects into a new weighted average.

One reason for reporting TSLS estimates as well as Wald estimates is that the TSLS estimate may have lower sampling variance than any single Wald estimate. The TSLS estimate also provides a summary statistic that combines estimates based on different weighting schemes. However, standard errors for TSLS estimates in the model outlined here should take account of the fact that there is a different treatment effect for each instrument. In practice, this means that the TSLS residual ($Y - \gamma - \beta_w D$) is likely to be heteroscedastic (conditional on Z). White (1982) provides a heteroscedasticity-consistent covariance-matrix estimator that can be used in this case.

TSLS estimators are usually associated with an over-identification test statistic that equals the objective function implicitly minimized by the estimates (Newey 1985). In a constant-treatment-effect model estimated by TSLS, the statistic provides an over-identification test for the null hypothesis that all the instruments are orthogonal to the regression error term. The constant treatment effect is over-identified because any single instrument would be sufficient for identification. But in the model outlined here, it no longer makes sense to talk about over-identification; in principle, each instrument can lead to a different estimate even though all the instruments satisfy the independence assumption. In fact, Theorem 1 provides one explanation for why estimates of causal effects such as the return to schooling may differ in different studies.

The conventional TSLS instrument-error orthogonality test statistic may still be worth computing, however, because it provides a summary measure of how much different weighting schemes affect estimates of the ACR. Angrist (1991) has shown that when the instruments are a set of mutually exclusive

dummy variables as in Theorem 2, then the over-identification test statistic is the same as a Wald statistic for the equality of a full set of linearly independent Wald estimates. In other words, the statistic provides a test of the null hypothesis H_0: $\beta_{K,K-1} = \beta_{K-1,K-2} = \ldots \beta_{1,0}$. The Wald statistic combines the differences between pairs of Wald estimates in a quadratic form, with weighting matrix equal to the inverse of the covariance matrix of the estimates.[7] In the context of the model outlined here, the test statistic should be large when there is substantial treatment effect heterogeneity. But it is important to note that the test statistic may also be large for the same reason a conventional over-identification test is failed: some of the instruments do not satisfy the independence assumptions.

4. Empirical Examples

4.1 Test-Preparation

In this section we discuss estimates of the ACR in two examples. The first is the Powers and Swinton (1984) test-preparation study. Assuming that the randomly assigned encouragement intervention satisfies independence and monotonicity restrictions, the experimental data can be used to estimate features of the causal relationship between test preparation and test scores. For example, an estimate of the ACR for the effect of test preparation on the GRE Analytical test score can be computed from data reported in the Powers and Swinton article. The mean Analytical score for those who received the letter encouraging preparation was 531.8 and the mean score for those not encouraged

---

[7]Multiple use of the label "Wald" may be confusing here. A *Wald estimate* is the sample analog of equation (1). A *Wald statistic* for the null hypothesis H_0: $\theta = \theta_0$ is the quadratic form: $m(\theta^*) = n[\theta^*-\theta_0]'\Phi^{-1}[\theta^*-\theta_0]$, where θ^* is an estimate of the parameter θ in a sample of size n, and Φ is the limiting covariance matrix of the estimate.

was 509.7. The mean hours of preparation for the Analytical section was 3.37 for those encouraged and 2.8 for those not encouraged. The ratio of the difference in scores to the difference in hours of preparation is 38.8, that is, an average causal score response of 38.8 points per hour of preparation.[8]

The anatomy of this estimate can be explored using data from Table 2 in Powers and Swinton (1984), which shows the frequency distribution of hours of preparation for the Analytical GRE according to whether subjects were assigned to receive encouragement or not. To simplify discussion of the ACR in this example, we have assumed that actual hours of preparation can be described by a discrete variable, D, taking 5 values corresponding to the five intervals reported by Powers and Swinton. The Powers and Swinton intervals are for the number preparing 0 hours, positive hours less than 1, 1-2 hours, 3-5 hours, and 6 or more hours, and are listed in column 1 of Table 1 here. Column 2 of Table 1 lists the discrete treatment intensities assumed to correspond to these intervals; D can be 0 hours, .5 hours, 1.5 hours, 4 hours, or 6 hours of test preparation.

Column 3 reports the cumulative frequency distributions of D by encouragement status. For example, where Powers and Swinton report the number preparing 0 hours, columns 3 and 4 record the number preparing less than .5 hours. Note that the empirical CDF of D given no encouragement, $Pr(D_0<j)$, always exceeds the CDF given encouragement, $Pr(D_1<j)$. This is a necessary (but not sufficient) condition for the monotonicity assumption to be satisfied

---

[8]Under the null hypothesis of no treatment effect, the asymptotic standard error for the ACR is given by the standard error of the numerator divided by the denominator (Angrist 1990.) This is 6.96 for the estimate of 38.8 points of GRE-score improvement per hour of study. Standard errors for the general case must be computed using conventional TSLS formulas requiring micro data.

by the encouragement intervention. The ACR weighting function is reported in column 5. This is simply the difference between columns 3 and 4, normalized to sum to 1.

The last two columns of Table 1 show which increases in treatment intensity the ACR weights apply to, and which part of the sample is contributing to the weights. The ACR weights act to combine four separate causal effects: the effect of moving from zero to .5 hrs of study, the effect of moving from .5 to 1.5 hours of study, the effect of moving from 1.5 to 4 hours of study, and the effect of moving from 4 to 6 hours of study. The first effect is weighted by 16.4 percent, and, as shown in column 7, this represents the fraction of the sample caused by encouragement to move from zero to .5 or more hours of study. The bulk of the weight falls on the middle two treatment effects, for moving from .5 to 1.5 hours of study and from 1.5 to 4 hours of study. The weights here represent the fraction of the sample induced to move from less than 1.5 to 1.5 or more hours of study (39.7 percent) and the fraction induced to move from less than 4 to 4 or more hours of study (33.1 percent.) Only 10.8 percent of those whose treatment status was affected by the encouragement letter moved from less than 6 to 6 or more hours of study. This fraction weights the effect of moving from 4 to 6 hours of study in the computation of the ACR.

Finally, note that Powers and Swinton also compute the regression of the average GRE Analytical score on the average hours of preparation for each of the 10 treatment and control groups underlying the test-preparation experimental design. The coefficient on average hours of test preparation in this 10 observation bivariate regression is a version of the TSLS estimator

described in Theorem 2.[9] The slope estimate is 30.74 points of GRE-score improvement. This estimate implicitly combines the Wald estimate of 38.8 points in a weighted average with the 8 other estimates that can be computed from pairwise comparisons of treatment and control groups.

4.2 Compulsory School Attendance

In two recent papers, Angrist and Krueger (1991, 1992) show that students' quarter of birth interacts with compulsory attendance laws and age at school entry to generate exogenous variation in years of completed schooling. State compulsory attendance laws typically require students to enter school in the Fall of the year in which they turn six, but allow students to drop out of school when they reach their 16th birthday. This induces a relationship between quarter of birth and educational attainment because students born in the first quarter of the year enter school at an older age than students born in later quarters. Students who enter school at an older age are allowed to drop out of school after having completed less schooling than students who enter school at a younger age.

Angrist and Krueger (1991) estimate the coefficient ρ in the following equation:

$$Y = \gamma + \rho E + \epsilon, \tag{10}$$

[9]If the grouped regression residual is homoscedastic, then the coefficient estimated from the bivariate grouped regression is the same as the TSLS estimate described in Theorem 2. More generally, weighted least squares estimation using data grouped by the value of discrete instruments is the same as TSLS estimation in micro data using these instruments (Angrist 1991.) The regression weights should equal the reciprocal of the residual variance in the grouped regression.

where E is years of schooling and Y is the log of weekly wages.[10] The coefficient ρ can be interpreted as a percentage "return" to a year of schooling and is usually on the order of 6-8 percent in econometric studies. But naive estimation procedures such as comparisons of average earnings by schooling level or Ordinary Least Squares (OLS) do not necessarily generate estimates with a causal interpretation. This is because those more educated may be people who, perhaps because they are more able, would have earned more even if they had not gotten more schooling.

Theorem 1 shows that even though schooling is not randomly assigned, the average causal response of earnings to schooling can be estimated if there are instruments available satisfying exclusion and monotonicity conditions. Even in the absence of a true experiment, a "natural experiment" may generate instruments satisfying these conditions. The premise underlying the estimation strategy in Angrist and Krueger (1991) is that age at school entry and compulsory attendance laws interact to generate variation in schooling that is likely to be unrelated to determinants of labor market outcomes other than education. This exogenous variation is then used to construct instrumental variables estimates of the effect of schooling on earnings. The instruments are dummy variables indicating quarter of birth.

A simple application of this idea compares the education and earnings of men born in the first quarter to the education and earnings of men born in the fourth quarter. Calculations underlying Wald estimates based on a first quarter/fourth quarter comparison are laid out in Table 2. Panel A of the Table shows results tabulated from data on the wages and earnings of men in

---

[10]Angrist and Krueger (1991) estimate variations on this equation that include additional covariates.

the 1970 Census and Panel B shows results tabulated using data from the 1980 Census. In both data sets, men born in the first quarter earn slightly less and have slightly less schooling than men born in later quarters. The ratio of differences in earnings to differences in schooling generates a Wald estimate of the return to schooling of 5.3 percent using the 1970 Census and 8.9 percent using the 1980 Census. These estimates are within sampling error of the OLS estimates of 8 percent and 7 percent in the two Census data sets.

Angrist and Krueger use linear regression models with constant coefficients (like equation 10), to interpret estimates of the return to schooling based on quarter of birth. In constant coefficient models, the independence assumption requires only that the regression error term be mean-independent of quarter of birth. Monotonicity (Assumption 2) is not required because there is no treatment effect heterogeneity.

In the context of the model outlined in Section 2 of this paper, the Wald estimates in Table 2 should be interpreted as the average effect of a one-year increase in schooling, for people whose schooling is influenced by quarter of birth. This is a small group, not necessarily representative of the entire population. To identify the ACR for this group, the monotonicity condition requires that men born in the fourth quarter get at least as much schooling as they would have had they been born in the first quarter. If this condition is satisfied, we can get some idea of the size and characteristics of the group contributing to the ACR through the ACR weighting function.

The CDF's of schooling by quarter of birth for men in the 1970 and 1980 Censuses are graphed in Figures 1 and 2. Both figures show that the CDF for men men born in the fourth quarter lies below the CDF for men born in the first quarter. The weighting function underlying estimates of the ACR in

Table 2 is proportional to the difference between the CDF of schooling for men born in the first quarter and the CDF of schooling for men born in the fourth quarter. For each level of schooling, j, this difference is the fraction of the population whose schooling is switched by quarter of birth from less than j years to at least j years.

Figures 3 and 4 show differences in the CDF of schooling by quarter of birth. In each figure, the difference between the CDF of schooling for men born in the 1st and fourth quarters is plotted, along with 95 percent confidence bands at each point.[11] ACR weighting functions for estimates based on comparisons between first and fourth quarter births are the CDF differences plotted in the figures, normalized to sum to one.

The figures show that the groups contributing most to estimates of the ACR based on quarter of birth are those with 8-12 years of schooling. There is a sharp decline in the weighting function at 12 years of schooling. A maximum of a little over 2 percent of the sample was induced by being born in the fourth quarter to complete 11th grade, but much smaller fractions were induced to complete higher grades. This is not surprising since compulsory attendance laws affect young students and cannot compel students to go to college. Some weight is contributed by college attenders, however, perhaps because some students forced by accident of birth to graduate high school later decide to go on to college after all.

Figures 5 and 6 plot the contrast between schooling CDF's for birth quarters 1-3 relative to fourth-quarter births. The figures show that schooling CDF's are ordered by quarter of birth. This is evidence that any

[11]The difference between CDF's by quarter of birth is the difference between two independent sample proportions. The confidence bands are calculated using the conventional formula for a difference in proportions.

adjacent pair of quarters can be used to define a binary instrumental variable that satisfies the monotonicity assumption. TSLS using three quarter of birth dummies is a weighted average of the three possible Wald estimates based on adjacent quarters of birth. TSLS estimates of the return to schooling in this case are .062 (standard error = .016) in the 1970 Census and .103 (standard error = .020) in the 1980 Census. These are similar to the Wald estimates based on a comparison of first and fourth-quarter births.

The TSLS over-identification test statistics take on the values 2.35 and 2.85 in the two Census data sets. Both statistics have chi-square distributions with 2 degrees of freedom under the null hypothesis of constant treatment effects and instrument-error orthogonality. These values therefore cast little doubt on the constant treatment effect and independence assumptions. Finally, the fact that the various instrumental variables and TSLS estimates are so close to the OLS estimates suggest that naive comparisons of earnings by the level of educational attainment may have a causal interpretation after all.

5. Summary and Conclusions

This paper defines the average causal response to variable treatments such as drug dosage, cigarettes smoked, hours of study, and years of schooling. The definition is motivated by Rubin's notion of counter-factual outcomes in evaluation research, and by our previous definition of Local Average Treatment Effects for binary treatments. We show here that a weighted average of per-unit casual responses to a change in treatment intensity is identified in a wide variety of models and circumstances. The average response we can identify is for those individuals whose treatment status is

affected by an instrumental variable that is independent of potential outcomes and potential treatment intensities. The monotonicity condition imposed when deriving this result requires only that the instrumental variable affect treatment intensity in the same direction for each unit of observation.

We have presented a number of formulas for the weighting function that underlies instrumental variables estimates of average causal effects. These formulas can help empirical researchers understand which observations are contributing to a particular estimate. But we have not presented new estimators, and, for researchers already using exogenous variation to estimate treatment effects and causal responses, there is little here that should affect empirical practice. Rather, our results provide a useful interpretation for some of the simple estimators commonly employed in applied research. We also hope these results help build a bridge between the econometric literature on evaluation and the evaluation literature in biometrics, sociology, and other disciplines.

Finally, the most important issue in evaluation research is probably not treatment effect heterogeneity, but whether the source of identifying information -- be it an intervention involving experimental random assignment, or a natural experiment -- is really associated with the outcome of interest solely because of association with the treatment. After having made the case for this link, however, it is important to recognize that data can only be informative about the effect of treatment on those whose treatment status is affected by the intervention. The Average Causal Response discussed in this paper does this by setting out specific formulas for the anatomy of the causal response to a variable treatment.

Table 1: Encouraged Test Preparation

Hours of Preparation for the Analytical Section of the GRE[a]

Hours of preparation		Cumulative dsn. function		ACR weights	Treatment effect for:	Weighting for:
		Encouraged:				
Actual hours	j	No $Pr(D_0<j)$	Yes $Pr(D_1<j)$			
(1)	(2)	(3)	(4)	(5)	(6)	(7)
	0	0.0	0.0			
0	.5	8.32	2.53	16.4	0 -> .5	$D_0 < .5 \leq D_1$
< 1	1.5	28.5	14.5	39.7	.5 -> 1.5	$D_0 < 1.5 \leq D_1$
1-2	4	57.1	45.4	33.1	1.5 -> 4	$D_0 < 4 \leq D_1$
3-5	6	80.2	76.4	10.8	4 -> 6	$D_0 < 6 \leq D_1$
> 6		100.0	100.0			
mean score GRE analytical:[b]		509.7 (2.9)	531.8 (2.7)			
mean hours of preparation:		2.80	3.37			
sample:		2,127	1,865			

[a]Column (1) shows the intervals for which hours of preparation are reported in Powers and Swinton (1984, Table 2.) Column (2) shows the value of the variable treatment intensity assumed when reporting the cumulative distribution function. Columns 3 and 4 of the table show the cumulative distribution of hours of preparation by encouragement-group. Column (5) shows the ACR weights, equal to (3)-(4) normalized to sum to one. Column (6) shows the increase in treatment intensity to which the weight applies to. Column (7) shows the sub-population the ACR weight refers to.

[b]Standard errors in parentheses.

Table 2: Compulsory School Attendance

Panel A: Wald Estimates for 1970 Census -- Men Born 1920-1929[a]

	(1) Born in 1st Quarter of Year	(2) Born in 2nd, 3rd, or 4th Quarter of Year	(3) Difference (Std. Error) (1) - (2)
ln (Wkly. Wage)	5.1485	5.1578	-0.00935 (0.00374)
Education	11.3996	11.5754	-0.1758 (0.0192)
Wald est. of return to education			0.0531 (0.0196)
OLS return to education[b]			0.0797 (0.0005)

Panel B: Wald Estimates for 1980 Census -- Men Born 1930-1939

	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year	(3) Difference (std. error) (1) - (2)
ln (Wkly. Wage)	5.8916	5.9051	-0.01349 (0.00337)
Education	12.6881	12.8394	-0.1514 (0.0162)
Wald est. of return to education			0.0891 (0.0210)
OLS return to education			0.0703 (0.0005)

[a]The sample size is 122,223 in Panel A, and 162,515 in Panel B. Each sample consists of males born in the U.S. who had positive earnings in the year preceding the survey. The 1980 Census sample is drawn from the 5% sample, and the 1970 Census sample is from the State, County and Neighborhoods 1% samples.

A detailed description of the data sets is provided in the Appendix to Angrist and Krueger (1991.)

[b]The OLS return to education was estimated from a bivariate regression of log weekly earnings on years of education.

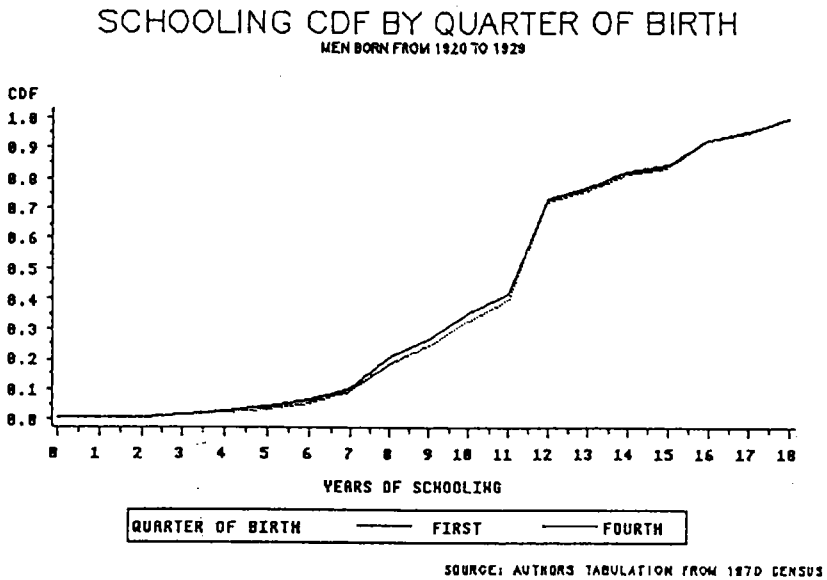
SCHOOLING CDF BY QUARTER OF BIRTH
MEN BORN FROM 1920 TO 1929
CDF
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
YEARS OF SCHOOLING
QUARTER OF BIRTH
FIRST
FOURTH
SOURCE: AUTHORS TABULATION FROM 1970 CENSUS

Figure 1

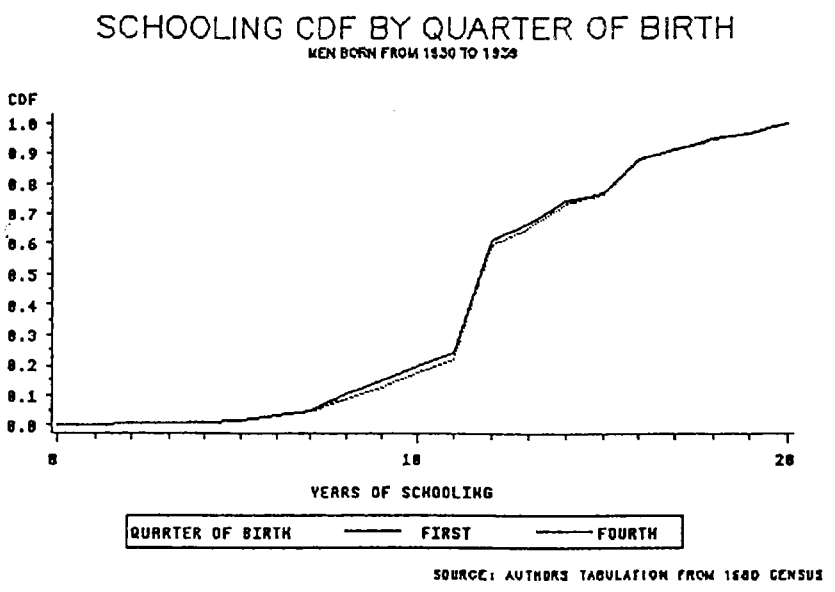
SCHOOLING CDF BY QUARTER OF BIRTH
MEN BORN FROM 1930 TO 1939
CDF
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
8
10
20
YEARS OF SCHOOLING
QUARTER OF BIRTH
FIRST
FOURTH
SOURCE: AUTHORS TABULATION FROM 1980 CENSUS

Figure 2

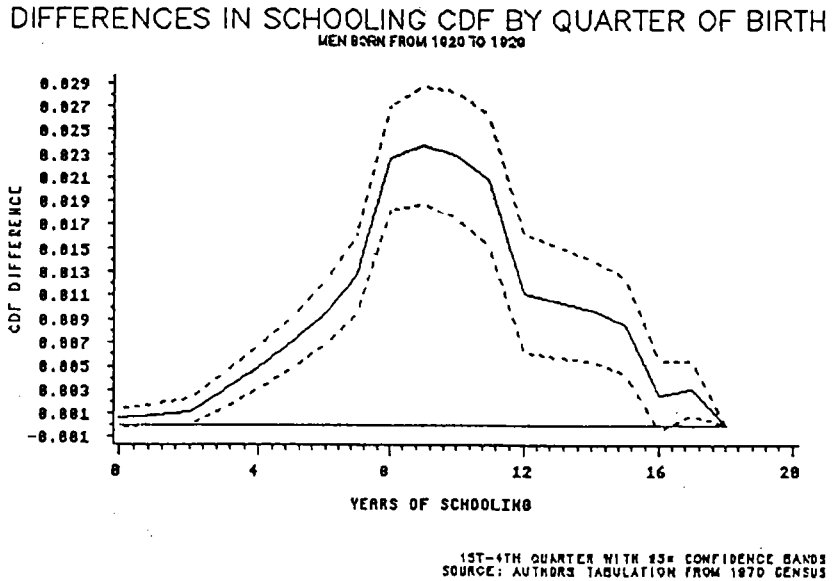
DIFFERENCES IN SCHOOLING CDF BY QUARTER OF BIRTH
MEN BORN FROM 1920 TO 1929
CDF DIFFERENCE
0.029
0.027
0.025
0.023
0.021
0.019
0.017
0.015
0.013
0.011
0.009
0.007
0.005
0.003
0.001
-0.001
0
4
8
12
16
20
YEARS OF SCHOOLING
1ST-4TH QUARTER WITH 95% CONFIDENCE BANDS
SOURCE: AUTHORS TABULATION FROM 1970 CENSUS

Figure 3

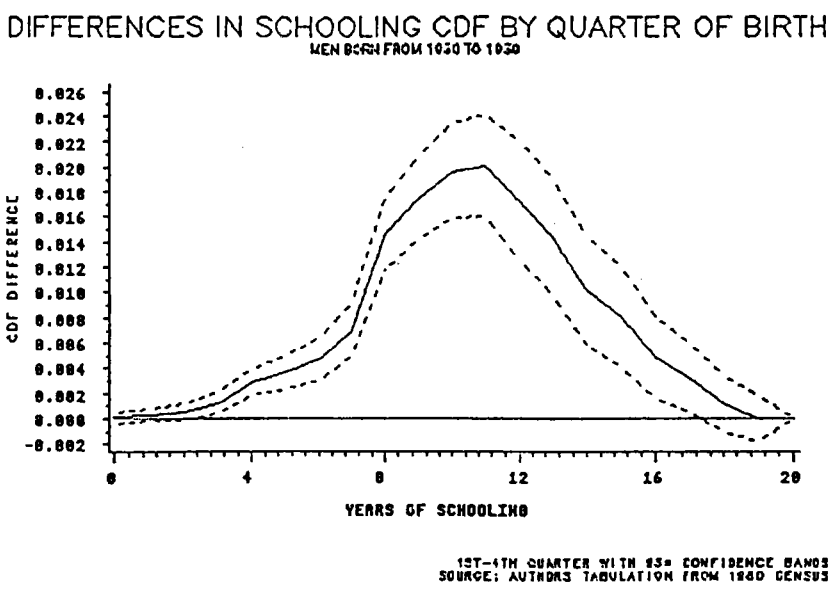
DIFFERENCES IN SCHOOLING CDF BY QUARTER OF BIRTH
MEN BORN FROM 1930 TO 1939
CDF DIFFERENCE
0.026
0.024
0.022
0.020
0.018
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
-0.002
0
4
8
12
16
20
YEARS OF SCHOOLING
1ST-4TH QUARTER WITH 95% CONFIDENCE BANDS
SOURCE: AUTHORS TABULATION FROM 1980 CENSUS

Figure 4

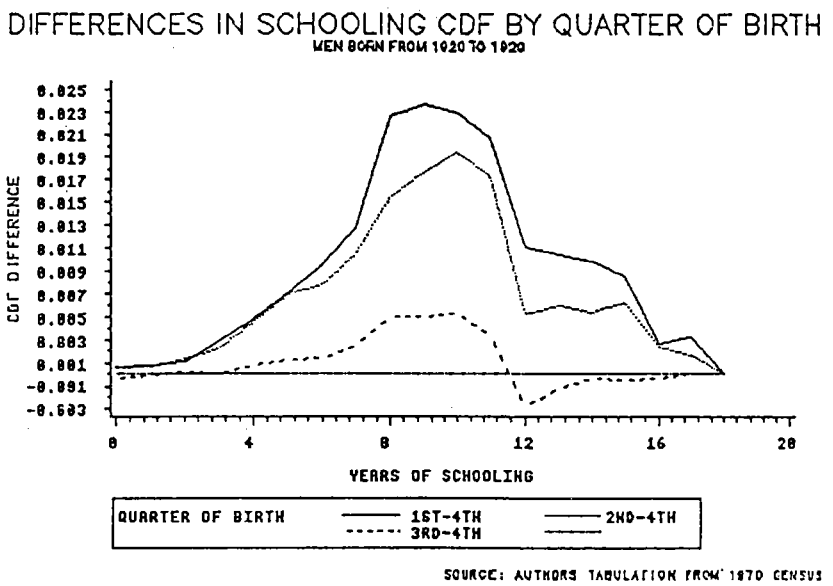
DIFFERENCES IN SCHOOLING CDF BY QUARTER OF BIRTH
MEN BORN FROM 1920 TO 1929
CDF DIFFERENCE
0.025
0.023
0.021
0.019
0.017
0.015
0.013
0.011
0.009
0.007
0.005
0.003
0.001
-0.001
-0.003
0
4
8
12
16
20
YEARS OF SCHOOLING
QUARTER OF BIRTH
1ST-4TH
2ND-4TH
3RD-4TH
SOURCE: AUTHORS TABULATION FROM 1970 CENSUS

Figure 5

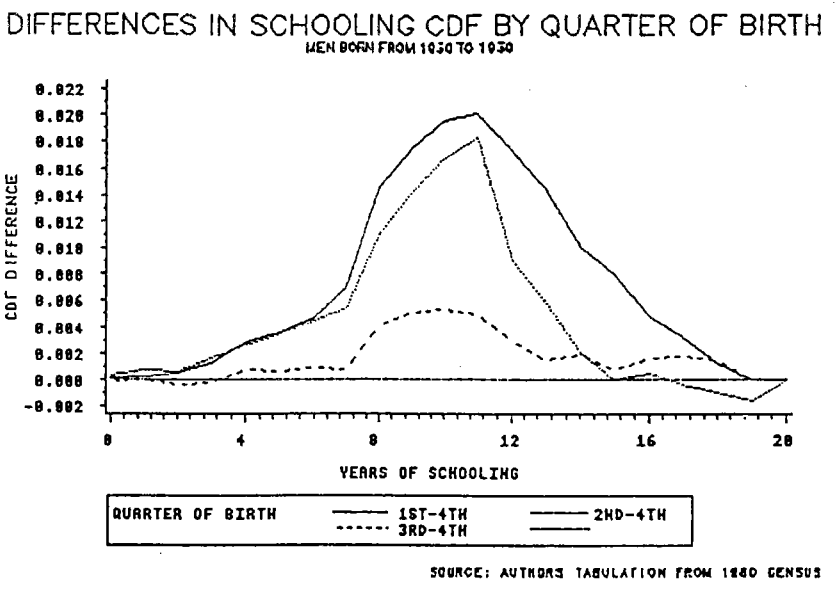
DIFFERENCES IN SCHOOLING CDF BY QUARTER OF BIRTH
MEN BORN FROM 1930 TO 1939
CDF DIFFERENCE
0.022
0.020
0.018
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
-0.002
0
4
8
12
16
20
YEARS OF SCHOOLING
QUARTER OF BIRTH
1ST-4TH
2ND-4TH
3RD-4TH
SOURCE: AUTHORS TABULATION FROM 1980 CENSUS

Figure 6

References

Aigner, D., and A. Zellner, (1988), Special Supplement on Causality in the Annals of the Journal of Econometrics (1988-3) 39.

Angrist, J., (1990), "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," American Economic Review, 80, 313-335.

_____ (1991a), "Grouped-Data Estimation and Testing in Simple Labor Supply Models", Journal of Econometrics, 47, 243--266.

_____ (1991b), "Instrumental Variables Estimation of Average Treatment Effects in Econometrics and Epidemiology," NBER Technical Working Paper No. 115, November.

_____ and A. Krueger, (1991), "Does Compulsory School Attendance Affect Schooling and Earnings", Quarterly Journal of Economics, 106, 979-1014.

_____ and _____ (1992), "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples," Journal of the American Statistical Association 87, June.

_____ and G. Imbens, (1991), "Sources of Identifying Information in Evaluation Models", NBER Technical Working Paper No. 117, December.

Durbin, J., (1954), "Errors in Variables", Review of the International Statistical Institute, 22, 23--32.

Efron, B., and D. Feldman, (1991), "Compliance as an Explanatory Variable in Clinical Trials", Journal of the American Statistical Association, 86, 9--26.

Hearst, N., Newman, T., and S. Hulley (1986), "Delayed Effects of the Military Draft on Mortality: A Randomized Natural Experiment," New England Journal of Medicine 314 (March 6), 620-624.

Heckman (1990), "Varieties of Selection Bias," American Economic Review, 80, 313-318.

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

Holland, P., (1986), "Statistics and Causal Inference," Journal of the American Statistical Association 81, 945-970.

_____ (1988). "Causal Inference, Path Analysis, and Recursive Structural Equations Models," Chapter 13 in Sociological Methodology, Washington: American Sociological Association.

Imbens, G.,and J. Angrist, (1991), "Identification and Estimation of Local Average Treatment Effects", NBER Technical Working Paper no 118, December.

Leamer, E.E. (1988), "Discussion," Chapter 14 in Sociological Methodology, Washington: American Sociological Association.

Manski, C. F., (1992), "The Selection Problem," in Advances in Econometrics, edited by C. Sims, New York: Cambridge University Press.

_____, and Irwin Garfinkel (1991), eds., Evaluating Welfare and Training Programs, Cambridge: Harvard University Press.

Newey, W., (1985), "Generalized Method Of Moments Estimation and Testing", Journal of Econometrics, 29, 229--256.

Permutt, T., and J. Hebel, (1984), "A Clinical Trial of the Change in Maternal Smoking and its Effect on Birth Weight," Journal of the American Medical Association, 251(7), (February 17), 911-915.

Permutt, T., and J. Hebel, (1989), "Simultaneous--Equation Estimation in a Clinical Trial of the Effect of Smoking on Birth Weight", Biometrics, 45, 619--622.

Powers, D.E., and S.S. Swinton, (1984), "Effects of Self-Study for Coachable Test Item Types," Journal of Educational Psychology 76, 266-78.

Rosen, S., and R.J. Willis, "Education and Self-Selection," Journal of Political Economy 87, S7-S36.

Roy, A. (1951), "Some Thoughts on the Distribution of Earnings," Oxford Economic Papers 3, 135-46.

Royall, M., (1991), "Ethics and Statistics in Randomized Clinical Trials", Statistical Science, 6, 52-88.

Rubin, D. (1974), "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," Journal of Educational Psychology, 66, 688-701.

_____ (1977), "Assignment to a Treatment Group on the Basis of a Covariate", Journal of Educational Statistics, 2, 1-26.

_____ (1990), "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies," Statistical Science 5, 472-480.

Wald, A., (1940), "The Fitting of Straight Lines if Both Variables are Subject to Error", Annals of Mathematical Statistics, 11: 284--300.

White, Halbert (1982), "Instrumental Variables Estimation with Independent Observations," Econometrica.