Interpreting Instrumental Variables Estimates of the Returns to Schooling

Jeffrey R. Kling

Department of Economics and Woodrow Wilson School
Princeton University, Princeton, NJ 08544
kling@princeton.edu

October 2000

Journal of Business and Economic Statistics, forthcoming

Interpreting Instrumental Variables Estimates of the Returns to Schooling
October 2000
JEL No. J3

ABSTRACT

This paper synthesizes economic insights from theoretical models of schooling choice based on individual benefits and econometric work interpreting instrumental variables estimates as weighted averages of individual-specific causal effects. Linkages are illustrated using college proximity to instrument for schooling. After characterizing groups differentially affected by the instrument according to family background, I directly compute weights underlying estimation of the overall return. In analyzing the level of schooling at which individuals change their behavior in response to the instrument, I demonstrate that this instrument has its greatest impact on the transition from high school to college. Specification robustness is also examined.

Jeffrey R. Kling
Department of Economics and Woodrow Wilson School
Princeton University
Princeton, NJ 08544
and NBER
kling@princeton.edu

1. INTRODUCTION

Instrumental variables (IV) estimation provides strategies to solve the identification problem that arises when individual's choice of educational attainment is related to their potential earnings (Card 1999). These strategies often rely upon observing individuals influenced to acquire more schooling through some rule or incentive, which typically affect schooling decisions of a subgroup of the population. If the return to education is not constant across individuals, then equally valid identification strategies relying on different subgroups may generate different results when returns differ across groups.

It is extremely useful to have information on the composition of the subgroup affected by an instrumental variable for the purposes of applying results in new contexts. Consider assessing the likely impact of tuition subsidies, such as the tax expenditures discussed in the 2000 U.S. Presidential campaign. Some IV results, such as those based on college proximity as an instrument (Card 1995b), are likely to be more relevant than results based on other IV strategies, such as compulsory schooling laws (Angrist and Krueger 1991). As with a subsidy, college proximity can lower the cost of attending college. We might expect those who face particularly high borrowing costs to be more sensitive to costs lowered by college proximity. If college proximity is actually inducing college attendance, then we would expect the largest effects to occur between years 12 and 16 of schooling. The population and the decision process affected by a tuition subsidy are likely to be more similar to those affected by college proximity than those affected by compulsory schooling laws. To the extent that these factors are systematically related to individual differences in returns to schooling, college proximity may incorporate these same interactions and have more external validity for forecasting future effects of tuition subsidies (Meyer 1995).

This paper uses data on covariates to show who is actually affected by the college proximity instrument, and at what level of education. This type of analysis can enhance our confidence that an instrumental variable and a proposed policy affect schooling decisions through the same mechanisms.

To summarize the organization of this paper, Section 2 reviews a simple model of schooling choice that provides the framework for interpretation of empirical results. Section 3 replicates previous work using college proximity to study the returns of schooling. Section 4 links the theoretical model of schooling choice to recent econometric work on the interpretation of IV estimates as a weighted average of causal effects for particular subgroups of the population, and computes these weights for the college proximity example. Section 5 examines the impact of the college proximity instrument at different levels of schooling. Section 6 discusses the sensitivity of results to use of different age groups. Section 7 examines the robustness of this general identification strategy using more recent data. Section 8 concludes.

2. A MODEL OF SCHOOLING CHOICE

Much of the literature on returns to schooling has been concerned with "ability bias" (e.g. Griliches, 1977). The following discussion of a simple human capital model is an attempt to clarify what it is that I hope to estimate, and to categorize the types of biases that may complicate interpretation of the results. In a canonical model of schooling choice, individuals will invest in schooling until the point at which the marginal change in present discounted value of future income from an additional year of schooling is equal to the intertemporal rate of substitution. Below, I adapt a version of this model of Becker's supply and demand model (1975), based on Card (1995a).

Individuals earnings are a function of schooling, $y_i = g_i(s)$. Following the literature (e.g. Willis 1986), this model uses the utility function $U(y,s) = \log(y_i) - \varphi_i(s)$, based on an individual maximizing the present discounted value of income, discounting the future at a constant rate, and earning nothing while in school. The optimal level of schooling is determined implicitly by the first order condition, when the marginal benefit of schooling is equal to the marginal cost. Equation (1) establishes notation for the marginal benefits (b_i) that influence an individual's optimal schooling choice, parameterizing them as functions of observable characteristics, X, and unobservable components, η_1.

$$\frac{g_i'(s)}{g_i(s)} = b_i = X_i\pi_1 + \eta_{1i} \qquad \textbf{(1)}$$

Equation (2) does the same for marginal costs. I have set up the model to suggest that instrumental variables in Z are available that affect schooling through the discount rate (e.g. capacity to pay for schooling) and not the marginal benefit.

$$\varphi_i'(s) = r_i + ks \quad ; \qquad r_i = X_i\pi_2 + Z_i\gamma_{2i} + \eta_{2i} \qquad \textbf{(2)}$$

The assumption in (2) of marginal cost with an individual specific discount rate (r_i) and a component that is increasing at a constant positive rate (k) in the amount of schooling seems plausible, for example, when individuals can finance education first internally from family savings, then from federally subsidized sources, and finally from private sources. The parameter γ_{i2} is a random coefficient to allow for individual differences in response to the instrument.

Equating (1) and (2) and substituting parameterizations of observables results in an explicit equation (3) for an individual's optimal schooling, s_i, where S is observed years of schooling. The instrument is assumed not to decrease schooling ($\gamma_{i3} \geq 0$).

$$\frac{b_i - r_i}{k} = s_i = S_i = X_i\pi_3 + Z_i\gamma_{3i} + \eta_{3i} \tag{3}$$

Integrating (1) and substituting (where Y is the observed log wage) results in the equation for earnings in (4), where ρ_i is a random coefficient.

$$\int_0^{s_i} \frac{g_i^{/}(s)}{g_i(s)} ds = \log(y_i) = a_i + b_i s_i = Y_i = X_i\beta + S_i\rho_i + \varepsilon_i \tag{4}$$

There are two types of "ability" in this model. First, individual differences in earnings capacity that do not interact with education are embodied in the individual specific earnings equation intercept a_i. If a_i is independent of the instrument variable affecting the level of schooling through the discount rate, then the instrument satisfies the exclusion restriction (Angrist, Imbens, and Rubin 1996). Specifically, I assume Z is independent of η_{3i} and ε_i, conditional on X. This model provides theoretical motivation for two stage least squares estimation (2SLS), in which the first stage is estimated by (3) and the second stage by (4).

The second type of ability is the marginal benefit to a year of schooling captured in b_i, which varies across individuals. Note that (4) is allowed to be linear in the endogenous schooling variable (as in Wooldridge 1997, Ashenfelter and Rouse 1998), and not quadratic (as in Card 1995a, Heckman and Vytlacil 1998) because b_i in (1) is modeled as invariant to schooling level. The 2SLS point estimate of the schooling coefficient is a weighted average of the heterogeneous marginal benefits for those whose schooling choice are affected by the instrument, conditional on X. The monotoncity implied in (3) when $\gamma_{i3} \geq 0$ allows results from IV estimates to be interpreted as an average causal response (Angrist and Imbens 1995), similar to the local average treatment effect derived for instrumenting a binary endogenous regressor

(Imbens and Angrist 1994). The weights imposed by 2SLS estimation are discussed in detail in sections 4 and 5.

3. REPLICATION OF PREVIOUS RESULTS

To provide background on this application, I replicate estimates reported by Card (1995b) on the return to schooling for those affected by college proximity. Card utilizes data from the original young men's cohort of the National Longitudinal Survey (NLSYM66), which began with youth aged 14-24 in 1966. Data on hourly wages and highest grade completed are taken from the 1976 wave. Card notes that proximity to two year colleges seems to have little impact on educational attainment. Perhaps only four year colleges have the visibility to influence schooling decisions, but it remains unclear why marginal consumers would not be influenced by proximity to two year colleges. Nevertheless this replication uses Card's instrumental variable, an indicator for whether or not there was a four year college in the local labor market (defined as the residential county) in 1966.

In Table I, I present a sample of results from Card (1995b) which I have replicated exactly. IV estimates of the return to schooling range from .094 to .132. Card's estimates include the traditional quadratic in potential experience (with age-squared as an additional excluded instrument), which complicates interpretation of the IV estimate for schooling in principle. As shown column 4, omitting this control has no substantive impact on the estimates. To simplify for the remainder of this paper, I use only a linear control for potential experience and instrument by age, so that the IV estimate of ρ in (4) effectively includes an additive constant from the potential experience coefficient.

There are of course several reasons why college proximity may not be a good instrument. Observably equivalent families may have different unobserved taste for education, and choose to live different distances from a college. Also, wages in labor markets for those who grew up near a college may simply reflect unobserved geographic wage premia, rather than difference in wages due to schooling. Card (1995b) shows that these estimates are not particularly sensitive to alternative strategies to control for local labor market conditions.

4. CHARACTERIZING THE AFFECTED GROUPS

With these replicated estimates as a frame of reference, this section presents new quantitative evidence characterizing the group affected by this instrumental variable. These estimates cannot be generalized to the larger population without additional assumptions--such as a constant marginal benefit of education across individuals. Without such an assumption, the external validity of the estimates depends upon the precision with which the affected subgroup can be characterized and on the policy interest generated by that group.

As a point of departure in attempting to characterize the affected group, Card (1995b) speculates that the effect of college proximity is more important for children of less wealthy households. He constructs an index of family background (predicted education from a linear regression of years of schooling on region and background characteristics, estimated on the sample that did not have a college nearby). I replicate this index and then define four quartiles based on it. Information on parental education provides the most significant set of variables predicting the index used to construct family background quartiles.

To see the differences in the underlying components used to create the background quartiles, I show descriptive statistics of selected characteristics in Table 2. In the lowest quartile, none of individuals report that either their mother or father graduated from high school. In the highest quartile, 97% of mothers are high school graduates, as are 84% of fathers. The presence of college in the 1966 county does vary somewhat across quartiles, from 59% for the lowest to 73% for the highest quartile. While this difference is much less dramatic than that of other characteristics of these individuals, I focus this analysis on differences in education and wages by college proximity conditional on having the same background characteristics. Card (1995b) reports that unconditional average education difference by college proximity is larger for the lowest background quartile (1.1 years) than the other quartiles (.2 to .4 years), which is suggestive of the important differences by quartile, but does not directly translate into the weight received by that group in IV estimation of the return to schooling

The IV estimate can be interpreted as a weighted average of the causal effect of a year of schooling within a subgroup (Card 1995a), as in equation (5). For a population subgroup Q, let $\Delta Y_q = E(Y|Z{=}1,Q) - E(Y|Z{=}0,Q)$ be the impact on earnings; let $\Delta S_q = E(S|Z{=}1,Q) - E(S|Z{=}0,Q)$ be the impact on schooling; let $\rho_q = E[\rho_i \mid Q]$ be the average return to schooling; let $w_q = P(Q)$ be weights. If the effect of the instrument on schooling is conditionally mean independent of the return, $E[\gamma_i \rho_i \mid Q] = E[\gamma_i \mid Q]\, E[\rho_i \mid Q]$, then the effect of the instrument on earnings within each subgroup is the product of the impact on educational attainment and the return to that education ($\Delta Y_q = \Delta S_q \rho_q$).

$$\frac{E[Y|Z=1] - E[Y|Z=0]}{E[S|Z=1] - E[S|Z=0]} = \frac{\sum_q w_q \Delta Y_q}{\sum_q w_q \Delta S_q} = \frac{\sum_q w_q \Delta S_q \rho_q}{\sum_q w_q \Delta S_q} \tag{5}$$

In this application using geographic proximity to college, however, the instrument is only valid conditional on X, and the main effect of X is controlled for using linear regression.

Consider the case where Q is a discrete indicator for family background quartile, so Q takes on four values. Let P(Z|X,Q) be the conditional probability of growing up with college-in-county. The instrumental variables estimate controlling for X and Q imposes weighting from the regression proportional to the conditional variance of Z, P(Z|X,Q)(1-P(Z|X,Q)), as shown by Angrist (1998). To focus on differences by Q while controlling for X, let $\lambda_{q|x}$ = E[P(Z|X,Q)(1-P(Z|X,Q)) | Q]. Note that if the instrument were independent of X and Q, then $\lambda_{q|x}$ would be a constant. With college proximity as the instrument, the IV estimates are also affected by the impact of the instrument on schooling conditional on X within each quartile. Let $\Delta S_{q|x}$ = E[E(S|Z=1,X,Q) - E(S|Z=0,X,Q) | Q]. The overall weight received by each quartile using 2SLS is $\omega_{q|x} = (w_q \lambda_{q|x} \Delta S_{q|x})/(\Sigma_q w_q \lambda_{q|x} \Delta S_{q|x})$.

Estimates of $\omega_{q|x}$ and its components are shown in Table 3. Column 1 is the sample proportion in each quartile. In column 2, $\lambda_{q|x}$ for each quartile is estimated using linear regression including X and Q to estimate P(Z|X,Q), and then the expectation is taken over the empirical distribution function of X for each value of Q. Linear regression is used to approximate full nonparametric conditioning in estimation of P(Z|X,Q) used by Angrist (1998). In order to correspond most directly to the 2SLS estimation in Table 1, $\Delta S_{q|x}$ is also computed using linear regression, controlling for X while estimating the coefficients on interactions between Z and Q. These conditional estimates of the impact of the instrument on schooling are somewhat lower than Card's unconditional estimates.

Column 4 contains the weight $\omega_{q|x}$ that would be used to form a weighted average of the marginal return to schooling for that group for those affected by the instrument, across the four quartiles. This calculation provides direct evidence that the lowest background quartile receives 53% of the weight in the overall IV estimate, and the highest two groups receive only 23% of the weight. If the average marginal benefit of schooling were higher for the lowest background quartile, this would be also consistent with the large estimates of return to schooling using distance as an instrument.

These IV estimates will represent the average marginal benefit from an additional year of education for the subgroup affected by the college proximity instrument . This subgroup turns out to be of policy interest both because it is composed largely of children from disadvantaged families, and because it represents individuals "on the margin" whose schooling decisions are affected by small changes in education costs.

Unfortunately, the sample size is too small to obtain reliable estimates of each ρ_q itself. In an attempt to increase estimation efficiency, I have incorporated additional years of panel data from 1976-1980 for individuals. However, I find that the effect on the precision of estimates of the return to schooling is negligible after accounting for the covariance structure of individual wages over time.

5. CHARACTERIZING THE RESPONSE FUNCTION

In addition to affecting a particular group as in Section 4, an instrumental variable may affect different transitions between levels of schooling. For example, college proximity may have the greatest impact on the transition from twelfth grade to the first year of college if it is affecting liquidity-constrained individuals. The response function traces out the impact of the instrument at each level of schooling, providing a richer description of the process through which the instrument influences behavior which can be useful for assessing external validity and relating IV estimates to proposed policy initiatives.

This response function can be estimated from the cumulative distribution functions (CDFs) of schooling at different values of the instrument. The difference in the CDFs is equivalent to the fraction of the population who received at least one more year of schooling due to the instrument. The CDF difference at each level of schooling is shown by Angrist and Imbens (1995) to integrate to the average difference in schooling caused by the instrument and to be proportional to the Average Causal Response (ACR) weighting function used by IV estimation. The average marginal benefit of a year of schooling for the group affected by the instrument can then be interpreted as the weighted average of per-unit changes in schooling.

Figure 1 shows the difference in the CDFs with a 95% pointwise confidence interval. The CDF difference for each schooling level j is estimated conditional on X and then integrated over the empirical distribution function of X, $E[P(S \geq j|Z=1,X) - P(S \geq j|Z=0,X)]$, where j is the number of years of schooling, X represents background characteristics, and Z=1 when a college is nearby. Note that the ACR weighting function is derived for nonparametric conditioning; I approximate using linear probability models for each integer schooling level conditioning on age, location and background.

The response function indicates that the college proximity instrument does indeed have larger effects for individuals who were caused to attend at least some college. At 13 years of schooling, I interpret the estimates in Figure 1 to indicate that 7% of individuals with similar demographics were induced to obtain 13 or more years of schooling due to college proximity, when they would have obtained 12 or fewer years of schooling without a college nearby. Note that the CDF difference captures the number of person-years of additional schooling, so an individual who was induced to complete 13 years of schooling by college proximity but who would have otherwise only finished tenth grade contributes to the CDF difference at 11, 12 and

13 years of schooling. This multi-year effect of the instrument rationalizes at least part of the positive CDF difference over the range of 9 to 12 years of schooling. The significant difference in CDFs at grade 10 implies that some individuals were induced to receive four additional years of schooling by college proximity, or that the anticipation of college kept potential dropouts in high school for one or two extra years even if they did not eventually attend college.

To identify an average causal effect using IV when marginal benefits of schooling vary across individuals, Angrist and Imbens (1995) show that an instrument must affect all individuals in a monotonic manner. In this case, college proximity should not decrease educational attainment for any individuals. This assumption has the testable implication that the difference in the cumulative distribution functions of schooling for the two values of a binary instrument should not be negative, which is satisfied in Figure 1.

The model of schooling choice in Section 2 has an additional simple prediction that the density of the weighting function should be greater over lower years of schooling for individuals from families with higher discount rates. This prediction is derived from (6), which shows that the covariance of discount rates and optimal schooling is nonpositive in this model.

$$Cov(r_i\ ,\ s_i) = \frac{\sigma_{br} - \sigma_r^2}{k} \leq 0 \tag{6}$$

For higher discount rates, schooling will be lower regardless of the distribution of ability, unless discount rates and ability are perfectly positively correlated. Thus, on average I can expect that a group with a high discount rate will be induced by college proximity to increase their education from a lower initial level (under the model's assumption that proximity affects all discount rates equally). For example, proximity to college will tend to cause high discount rate individuals to graduate from high school and attend some college when they might otherwise have dropped out of high school, while proximity will tend to cause low discount rate individuals who would have attended college anyway to increase the number of years they attend college.

Assuming that disadvantaged backgrounds have less capacity to pay for schooling -- higher r_i in equation (2) -- I use background quartile as a proxy for the discount rate. The prediction that lower quartiles, or higher r_i, will be affected at lower ranges of the response function is verified in Figure 2, where I compute the CDF difference by background quartile, estimated using linear regression and including a full set of quartile main effects and interactions with the college proximity instrument. There appears to be substantial response in the two lower background quartiles from those who otherwise would not have attended college. The response for the upper two quartiles is concentrated among those who would have attended at least some college, but attend additional years due to college proximity. In assessing the statistical variability, a simple rule of thumb for Figure 2 is that point estimates greater than .05 can be distinguished from zero at a p-value of .05.

Within the lowest background quartile, 13% received 13 or more years of schooling in Figure 2 when a college was nearby in 1966, while comparable members of that quartile who did not have a college nearby completed twelve years of schooling or less. There also appears to be some effect of inducing 9th-12th graders to graduate high school and attend some college. One interpretation is that the schooling levels of those in the lowest quartile differ for reasons that have nothing to do with college proximity itself, which would cast doubt on the validity of the instrument. This behavior could be consistent, however, with the role of college proximity in affecting schooling choices. For instance, individuals may be forward looking but have imperfect foresight (e.g. college proximity caused a potential tenth grade dropout to complete eleventh grade in anticipation of being able to afford to go to college, but this individual had to quit high school and start working after an unexpected death in the family). Alternatively, college proximity may simply have had large effects on these individuals (e.g. causing a potential tenth grade dropout to attend at least some college).

Overall, these patterns suggest that for the least wealthy, being near a college helps motivate individuals to stay in school longer and makes it more affordable to at least try out college. For individuals from moderately wealthy families, college proximity allows individuals to complete additional years of college.

6. SENSITIVITY OF ESTIMATION

While Sections 4 and 5 focused largely on the interpretation of the external validity of existing estimates, an important aspect in interpreting IV estimates is investigation of internal validity -- examining the sensitivity of estimates to violations of key identifying assumptions. One important issue in these data is that college proximity is constructed based on residential location at the time of first interview in 1966. At that time, older sample members (say, ages 20-24) may have relocated to a new county precisely in order to complete their postsecondary education. Card (1995b) notes this possibility, and presents estimates based only on sample members ages 14-19 in 1966, shown in the fourth column of Table 1. For these younger sample members, the estimated return to schooling is somewhat lower than the estimate for the full sample. Although imprecisely estimated, this is consistent with an upward bias in the coefficient induced by endogenous location of older sample members.

A more direct test of differences in the effect of the distance instrument by age examines the additional schooling obtained by members of different age groups. The results in Table 4 show that individuals ages 14-19 in 1966 received about 0.26 additional years of schooling when a college was nearby in 1966 (conditional on age, past location, and background), while those ages 20-24 in 1966 appear to have received 0.66 additional years of schooling when a college is nearby. The significantly higher schooling level of the older group suggests that this group contains some individuals with higher education who moved to a local area that has a college, and that the college proximity variable is not a legitimate instrument for this group.

For those ages 18-19, it is not entirely clear whether address information was based on location of household at screening or address at first interview, where college students may have been temporarily away from their permanent residence at the time of the interview. (Center for Human Resources Research 1994, p. 228). The results in Table 2 suggest that any effect of this phenomenon is not overwhelmingly strong, as the effect of college proximity for those between ages 14-17 those ages 18-19 is very similar and well within sampling error.

7. A COMPARISON OF 1976 TO 1989

To further assess the robustness of the college proximity identification strategy, this section analyses data virtually identical to that in the NLSYM66 described above, but collected thirteen years later. I use a cohort of 3213 young men surveyed in the National Longitudinal Survey of Youth (NLSY79). I matched the confidential county level geocode information on residential location at age 14 to data on the location of four-year colleges (National Center for Education Statistics, 1978). The age range is restricted to those 14-19 in 1979, for comparability with the preferred NLSYM66 sample based on the analysis in section 6.

In Table 5, I show the first stage estimates of the effect of college proximity on the educational attainment of men ages 24-29. Using specifications similar to Table 3, I interact the College-In-County indicator with background quartile. For the NLSYM66, column 1 shows that restricting the sample to 24-29 year olds leads to results across quartiles substantively similar to the results for ages 24-34. The response function for this group corresponding to Figure 2 is also similar. As a type of specification check for omitted variable bias, the results in columns 2 also show that the strong effect in the NLSYM66 is attenuated only slightly (a decrease from 1.04 to .96) when additional family background controls are included for availability of print media, immigrant status, and family size – even though these variables are jointly very significant. The corresponding returns to schooling IV estimate is .131 (.064).

For the NLSY79, column 3 in the first row of Table 5 shows that the lowest background quartile has the largest interaction, but the effect of college proximity on educational attainment is less than half as large in the NLSY79 as in the NLSYM66. Moreover, the effect is nearly cut in half again in column 4 with additional background controls; in a test of the joint statistical significance of the four college-in-county interactions, I accept the null hypothesis of no difference from zero among quartiles (p-value .11). Other recent research using the NLSY79 has come to the same substantive conclusion about these interactions (Cameron and Taber 2000). For the overall first stage relationship, the indicator for college-in-county is only associated with 0.08 years of additional schooling in the NLSY79 (with a standard error of .08), which is statistically insignificant and about one-third the magnitude of the first stage relationship reported in for the NLSYM66 cohort in Section 3. As a consequence, the corresponding IV estimate of the return to schooling is extremely imprecise, with a standard error of 0.46 (and a point estimate of 0.44) using 1989 hourly wage data in the NLSY79.

Is this decline over time in the effect of college-in-county on educational attainment to be expected? Recent work shows that the market for higher education has become geographically more integrated. Hoxby (1997) presents evidence on several factors that may have led to this integration between 1960s and the 1980s, including the deregulation of the airline and telecommunications industries (reducing the transportation and communication costs of attending college further from home), and tuition reciprocity among states' public college systems. Earlier innovations gradually took hold and also helped integrate the market, such as standardized admissions testing and financial need analysis, and the information exchange system initiated by the National Merit Scholarship program. While the costs of going away to college appear to have decreased in absolute terms, they also seem to have decreased even more in relative terms. Travel, communication, and room & board costs are have become a relatively smaller share of the cost of college as tuition costs have risen faster over time (Kane 1999). Students have also become much more likely to apply to out-of-state colleges over time (Hoxby 1997). This may be related to a type of network effect, where it is more feasible to conceive of going away to college if one knows others who are doing so. In general, this evidence is consistent with a reduction over time in the importance of proximity as an inducement for college attendance.

8. CONCLUSION

In this paper I offer an interpretation of an IV estimate of the return to schooling as a weighted average of causal effects, using the example of college proximity as an instrumental variable. I replicate earlier work in which Card (1995b) concludes that the returns to education are about 10-14% for individuals whose schooling is affected by the instrumental variable of distance to college, as opposed to a typical OLS estimate of 8%.

One of the most important questions raised by this instrumental variable is its external validity: what subgroups of the population do these estimates tell us about? I find that most of the individuals affected were from more disadvantaged family backgrounds, particularly with lower parental education. I quantify the weights implicit in the IV estimates, and show that 53% of the weight is on the lowest family background quartile, and only 23% is on the top two quartiles. Also, using family background quartiles as a proxy for discount rate differences, my results show that proximity to college tends to cause high discount rate individuals to graduate from high school and attend some college, while proximity tends to cause low discount rate individuals to increase the number of years they attend college.

A second important question for interpretation is the internal validity of this instrumental variable: does it identify a causal effect? The evidence here is mixed. Two rationalizable but puzzling facts are that proximity to a two year college has no effect on schooling and that the presence of a college nearby appears to affect the educational attainment of some individuals who would otherwise have only completed ninth or tenth grade. However, restricting the age range to a more valid subset and introducing additional background controls does not substantively change the original results using the NLSYM66 (in Card 1995b), and the reduced impact of the instrument in more recent data is consistent with increasing geographic integration of the higher education market.

Although analysis of the internal validity of this particular application suggests that estimates of the returns to education using college proximity as an instrument should be viewed with some caution, the synthesis of modeling and methodology used to examine the external validity is generally applicable to future studies of schooling using instrumental variables to identify causal effects.

ACKNOWLEDGMENTS

I thank John Abowd, Josh Angrist, Sue Dynarski, Matt Eichner, Jerry Hausman, Bo Honoré, Caroline Hoxby, Guido Imbens, George Jakubson, Larry Kahn, Larry Katz, Whitney Newey, Steve Pischke, Ceci Rouse, Chris Taber, and John Tyler for helpful comments, and David Card both for comments and for providing data from his research. I gratefully acknowledge support during this research from a National Science Foundation Graduate Fellowship and from the National Bureau of Economic Research.

REFERENCES

Angrist, J. D. (1998), "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, 66:2, 249-288.

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

Angrist, J. D., Imbens, G. W, and Rubin, D. B. (1996), "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444-455.

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

Ashenfelter, O. C., and Rouse, C. E. (1998), "Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins," *Quarterly Journal of Economics,* 113, 253-284.

Becker, G. S. (1975), *Human Capital.* Chicago: University of Chicago Press.

Center for Human Resources Research, (1994), "NLS User's Guide," Columbus: Ohio State University.

Cameron, S., and Taber, C. R. (2000) "Borrowing Constraints and the Returns to Schooling," NBER Working Paper No. 7761.

Card, D. E. (1995a), "Earnings, Schooling, and Ability Revisited," *Research in Labor Economics*, 14, 23-48.

----- (1995b), "Using Geographic Variation in College Proximity to Estimate the Returns to Schooling," in *Aspects of Labour Market Behaviour: Essays in Honor of John Vanderkamp*, eds. L. N. Christofides *et al.* Toronto: University of Toronto Press. 201-221.

----- (1999), "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics, Volume 3A*, eds. O. C. Ashenfelter and D. E. Card. Amsterdam: North-Holland, 1801-1863.

Griliches, Z. (1977), "Estimating the Returns to Schooling: Some Econometric Problems." *Econometrica,* 45, 1-22.

Heckman, J. J., and Vytlacil, E. J. (1998) "Instrumental Variables Methods for the Correlated Random Coefficient Model," *Journal of Human Resources*, 33:4, 974-1002.

Hoxby, C. M. (1997) "How the Changing Market Structure of U.S. Higher Education Explains College Tuition," NBER Working Paper No. 6323.

Imbens, G. W., and Angrist, J. D. (1994) "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-476.

Kane, T. J. (1999), *The Price of Admission*, Washington, D.C.: The Brookings Institution.

Meyer, B. D. (1995), "Natural and Quasi-experiments in Economics," *Journal of Business and Economic Statistics*, 13:2, 151-161.

National Center for Education Statistics, (1978), *ICPSR 7647: Higher Education Institutional Characteristics, 1977,* Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

Willis, R. J. (1986), "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions," in *Handbook of Labor Economics, Volume 1*, eds. O. C. Ashenfelter and R. Layard, Amsterdam: North-Holland, 525-602.

Wooldridge, J. M. (1997), "On two stage least squares estimation of the average treatment effect in a random coefficients model," *Economic Letters*, 56, 129-133.

Table 1. Replicated IV Estimates of Returns to Schooling.

Estimation technique	OLS (1)	IV (2)	IV (3)	IV (4)
Schooling coefficient	.073 (.003)	.132 (.049)	.094 (.064)	.133 (.049)
Quadratic Experience	yes	yes	yes	no
Sample (Age in 1966)	14-24	14-24	14-19	14-24
N = number of observations	3010	3010	2037	3010
Card (1995b): Table-col/row	T2-c2	T3-c5B	T4-r7	–

NOTE: Regression of log hourly wage in 1976 on years of schooling, with standard errors in parentheses. All specifications also include Experience, Race, Original Location, Current Location, and Background. Experience = Age - Schooling - 6. In columns 1-3, schooling, experience, and experience squared are treated as endogenous, with an indicator for college in county, age, and age-squared as excluded instruments; the quadratic in experience (and age) is omitted in column 4. Race is an indicator for black. Original Location variables are indicators for nine regions of residence in 1966 and an indicator of residence in an SMSA in 1966. Current Location variables are an indicator for living in an SMSA in 1976, and for living the Southern U.S in1976. Family Background variables are highest grade attained by mother and by father and 12 indicators: residence with both natural parents at age 14, residence with mother only at age 14, missing mother's and missing father's education, and eight interactions of parental education. The sample includes all observations with reporting highest grade completed in 1976. Replicating Card's sample exactly requires including only observations with wages between $1 and $30 per hour. 49 additional observations were excluded for whom the interview month was missing in 1966 (D. Card, personal communication, April 20, 1995). This appears to have been an attempt to omit those not interviewed, however those 49 individuals were actually interviewed even though the interview month is not recorded in the data.

Table 2. Probability of Characteristics by Family Background Quartile

Quartile	Lowest	3rd	2nd	Highest
Mother High School Graduate	0	.05	.55	.97
Father High School Graduate	0	.06	.31	.84
Lived with Both Parents at Age 14	.54	.73	.86	.93
Race is Not Black	.41	.67	.88	.96
Lived in SMSA at Age 14	.46	.64	.70	.74
College in County in 1966	.59	.67	.71	.73

NOTE: Background Quartile is computed as follows. Following Card (1995b), a predicted value is estimated from a regression of schooling on indicators for age and race as well as original location and background variables defined in Table 1 for the sample of 1163 observations with no college in their county in 1966. The 25^{th}, 50^{th}, and 75^{th} percentiles of the predicted values from this sample were used to group all 3010 observations with valid wage and education data into four quartiles. Each background quartiles includes one quarter of the observations with no college in their 1966 county, 239-240 observations per quartile. There are 339 observations that do have a college in county in the lowest quartile, 483 in the third quartile, 593 in the second quartile, and 638 in the highest quartile. Those with missing parental education are coded as not reporting that parent was a high school graduate, since both groups go on to have low educational attainment.

Table 3. Decomposition of IV weighting by Family Background Quartile

Q	w_q (1)	$\lambda_{q\|x}$ (2)	$\Delta S_{q\|x}$ (3)	$\omega_{q\|x}$ (4)
Lowest quartile	.19	.18	.93 (.18)	.53
3rd quartile	.24	.16	.35 (.18)	.23
2nd quartile	.28	.15	.20 (.19)	.14
Highest quartile	.29	.15	.13 (.20)	.09

NOTE: $w_q = P(Q)$. $\lambda_{q|x} = E[P(Z|X,Q)(1-P(Z|X,Q)) \mid Q]$. $\Delta S_{q|x} = E[E(S|Z=1,X,Q) - E(S|Z=0,X,Q) \mid Q]$. $\omega_{q|x} = (w_q\, \lambda_{q|x}\, \Delta S_{q|x})/(\Sigma_q\, w_q\, \lambda_{q|x}\, \Delta S_{q|x})$. $\lambda_{q|x}$, $S_{q|x}$, are computed linear regression as described in the text. X includes age, region and background characteristics (defined in Table 1). Since background quartile is positively correlated with college in county, this implies that there is a larger share of data in the upper quartiles with a college in their county in 1966, as is shown in column 1. This rule differs from Card's original formulation of equal size quartiles based on the overall population with valid education data, and was adopted to improve estimation efficiency by ensuring that the upper quartiles do not include only a small number of observations with no college in county. Using Card's definition, the point estimates for the specification in column 3 are .87 (lowest), .16 (3rd), .39 (2nd), and .02 (highest).

Table 4. Difference in Highest Grade Completed at age 25
by College Proximity in 1966 for Five Age groups

Age in 1966:	14-15	16-17	18-19	20-21	22-24
Interaction of Age and College Proximity	.25 (.18)	.24 (.18)	.29 (.21)	.66 (.25)	.65 (.21)

NOTE: Additional regressors are 11 indicators for Age, as well as Race, Original Location and Background as defined in Table 1. Standard errors in parentheses.

Table 5. Regression of Highest Grade Completed on Interaction of College-In-County (CIC) with Background Quartile and Other Covariates, Men ages 24-29

	NLSYM66 in 1976		NLSY79 in 1989	
	(1)	(2)	(3)	(4)
CIC * Lowest quartile	1.04	.96	.47	.28
	(.24)	(.24)	(.16)	(.15)
CIC * 3rd quartile	.48	.42	-.06	-.10
	(.22)	(.22)	(.16)	(.15)
CIC * 2nd quartile	.27	.19	.08	.03
	(.21)	(.20)	(.15)	(.15)
CIC * Highest quartile	-.09	-.04	-.21	-.29
	(.19)	(.19)	(.16)	(.16)
Age, Race	x	x	x	x
Original Location	x	x	x	x
Background	x	x	x	x
Background Quartile	x	x	x	x
Print Media		x		x
Immigrant/Family Size		x		x
R-squared	.26	.31	.27	.31
F-stat [p-value] of CIC * quartile interactions	5.79 [.00]	4.98 [.00]	2.80 [.02]	1.87 [.11]

NOTE: Age, Race, Original Location and Background are defined as in Table 1. Original Location for NLSY79 is region at age 14, and residence in an urban area at age 14. Background quartile is computed as Table 2. Print media includes indicators for presence at age 14 of magazines, of newspapers, and a library card. Immigrant/Family Size includes indicators for immigrant status and 2-3, 4-5, 6-8, and 9+ siblings. F-statistic is for the joint significance of the CIC by quartile interactions in rows 1-4. Sample sizes are 2037 (NLSYM66) and 3213 (NLSY79). Standard errors are in parentheses.

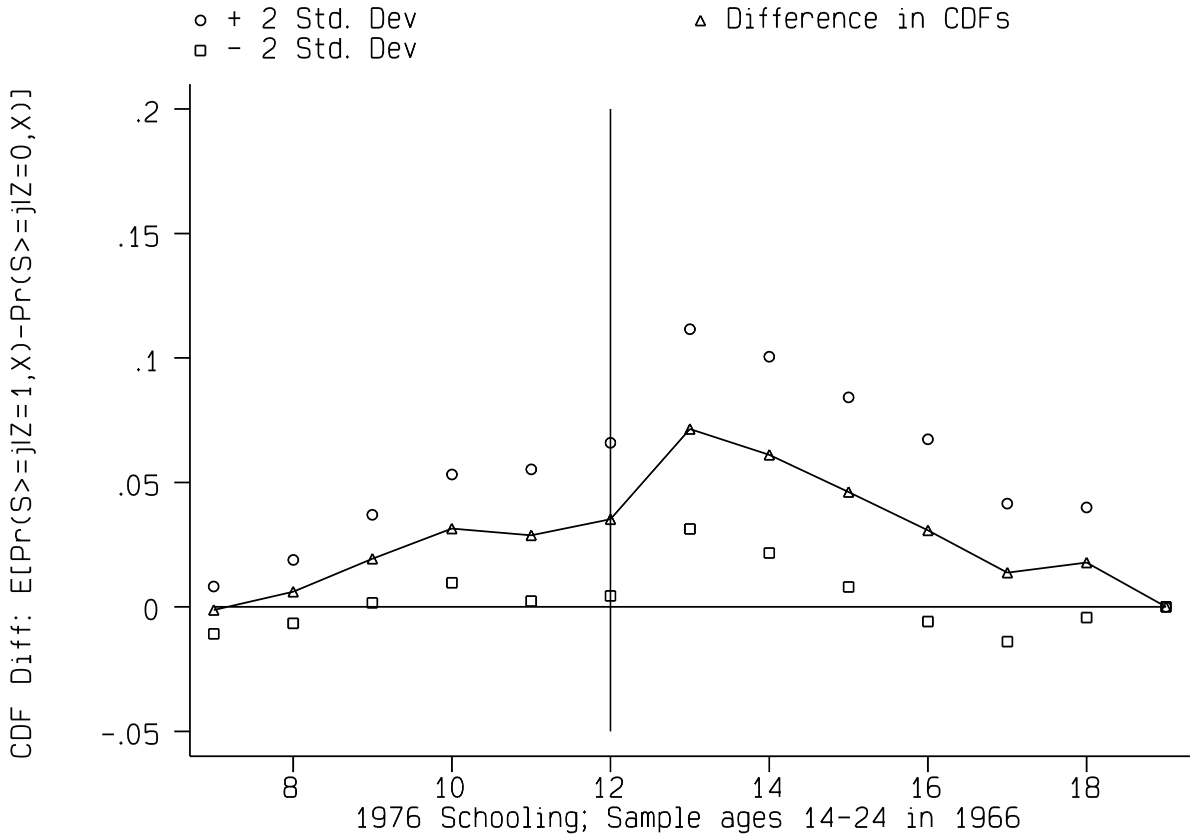

Figure 1: CDF Difference

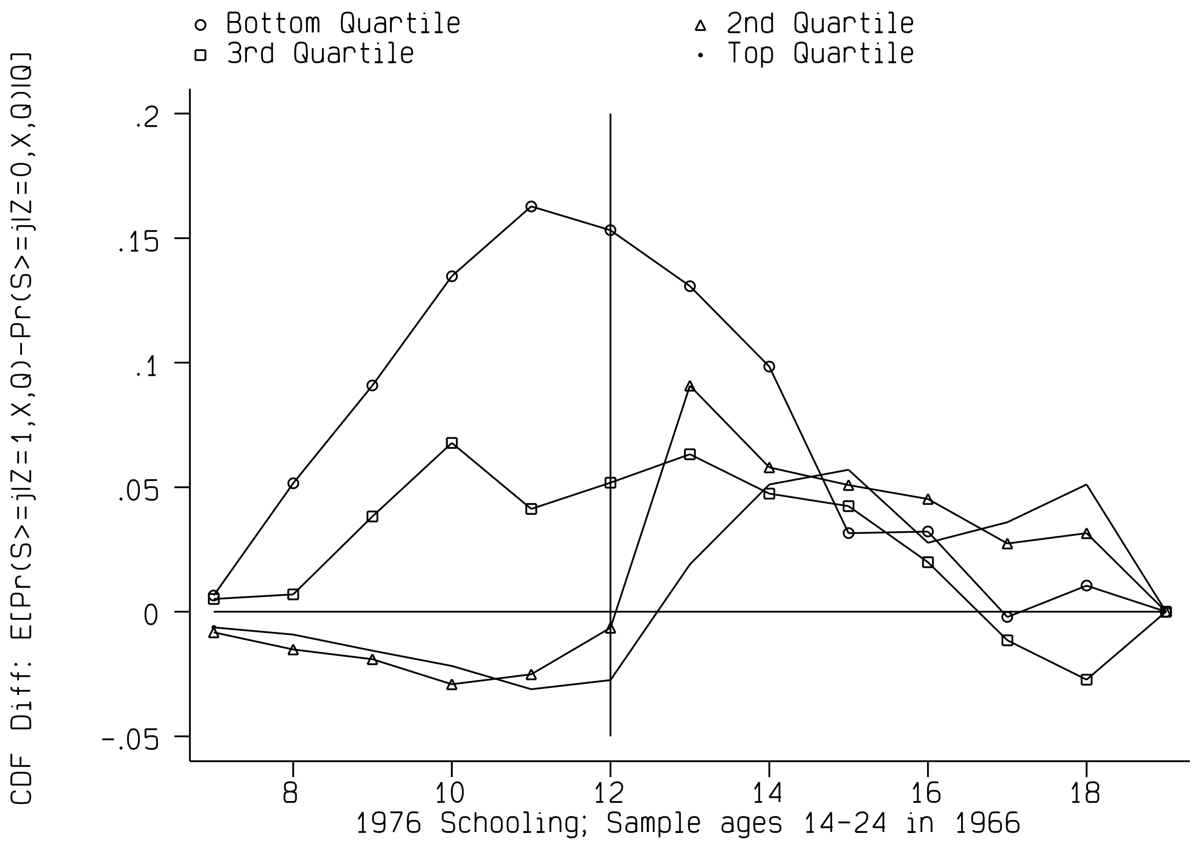

Figure 2: CDF Difference by Family Background Quartile