

NBER TECHNICAL PAPER SERIES

THE CURE CAN BE WORSE
THAN THE DISEASE: A
CAUTIONARY TALE REGARDING
INSTRUMENTAL VARIABLES

John Bound

David A. Jaeger

Regina Baker

Technical Paper No. 137

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 1993

We have benefited from valuable conversations with Chris Cavanagh, Gary Chamberlain, Zvi Griliches, Guido Imbens, David Lam and Kevin M. Murphy and seminar participants at the University of Michigan. Joshua Angrist, Eli Berman, Charles Brown, Alan Krueger, Kevin Lang, Lung-Fei Lee, Marianne Page, and Gary Solon provided extremely useful comments on an earlier draft. This work was stimulated, in part, by a question asked of the first author by Sanders Korenman. We alone, however, are responsible for the views expressed here. 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 #137
June 1993

THE CURE CAN BE WORSE THAN THE DISEASE: A CAUTIONARY
TALE REGARDING INSTRUMENTAL VARIABLES

ABSTRACT

In this paper we draw attention to two problems associated with the use of instrumental variables (IV) whose importance for empirical work has not been fully appreciated. First, using potential instruments that explain little of the variation in the endogenous explanatory variables can lead to large inconsistencies of the IV estimates even if only a weak relationship exists between the instruments and the error in the structural equation. Second, in finite samples, IV estimates are biased in the same direction as ordinary least squares (OLS) estimates. The magnitude of the bias of IV estimates approaches that of OLS estimates as the R^2 between the instruments and the potentially endogenous explanatory variable approaches 0. To illustrate these problems with IV estimation we reexamine the results of the recent provocative paper by Angrist and Krueger "Does Compulsory School Attendance Affect Schooling and Earnings?" and find evidence that their IV estimates of the effects of educational attainment on earnings are possibly both inconsistent and suffer from finite sample bias. To gauge the severity of both problems we suggest that both the partial R^2 and the F statistic on the excluded instruments from the first stage estimation be reported when using IV as approximate guides to the quality of the IV estimates.

John Bound
Population Studies Center
The University of Michigan
1225 South University Ave.
Ann Arbor, MI 48104-2509
(313) 998-7149
John.Bound@umich.edu
and NBER

David A. Jaeger
Population Studies Center
The University of Michigan
1225 South University Ave.
Ann Arbor, MI 48104-2590
(313) 998-7149
David.Jaeger@umich.edu

Regina Baker
Institute for Social Research
The University of Michigan
426 Thompson Street
Ann Arbor, MI 48109-1248
Regina.Baker@um.cc.umich.edu

When searching for plausible instruments for a potentially endogenous explanatory variable it is common to find that the candidates are only weakly correlated with the endogenous variable in question. It is well recognized that using such variables as instruments is likely to produce estimates with large standard errors. In this paper we draw attention to a number of other problems associated with the use of such instruments. First, if the potential instruments are only weakly correlated with an endogenous explanatory variable then even a weak correlation between the instruments and the error in the original equation can lead to a large inconsistency in instrumental variables (IV) estimates. Second, in finite samples, IV estimates are biased in the same direction as are ordinary least squares (OLS) estimates, with the magnitude of the bias approaching that of OLS as the R^2 between the instruments and the potentially endogenous explanatory variable approaches 0. While these results are known, we believe that their potential importance for empirical work has not been fully appreciated.

To illustrate these issues we reexamine the results of the recent provocative paper by Angrist and Krueger [1991] (henceforth AK), "Does Compulsory School Attendance Affect Schooling and Earnings?" In this paper AK use quarter of birth as an instrument for education in wage equations. While quarter of birth is only weakly related to educational attainment—the R^2 in the regression of educational attainment on quarter of birth ranges between 0.0001 and 0.0002 in their samples—AK obtain reasonable standard errors on their estimates of the effect of education on weekly earnings due to the large samples they use.

We present evidence suggesting that a weak correlation between quarter of birth and wages (independent of the effect of quarter of birth on education) exists and is sufficiently large to have quantitatively significant effects on AK's estimates. We also present results that indicate that the finite sample bias may be quantita-

tively significant for the estimates that AK report using quarter of birth \times state of birth interactions as instruments. Together these results suggest that the "natural experiment" afforded us by the interaction between compulsory school attendance laws and quarter of birth does not give us much usable information regarding the causal effect of education on earnings.

While these results are directly applicable only to an interpretation of AK's estimates, we believe they are of more general significance. In particular, they suggest that it may be even harder to find legitimate instruments for potentially endogenous variables than many have thought. While researchers may believe, *a priori*, that the variation in an instrument is largely unrelated to the process under study, this is not sufficient to imply that IV estimates will be less biased than those produced using OLS. In addition, these results suggest that economists working with large micro-level data sets need to be concerned more than they previously have been about the finite sample properties of IV estimators.

1. Potential Problems Using an Instrument that is Only Weakly Correlated with the Endogenous Variable

We are interested in estimating equation (1) from the following system (for notational simplicity individual observation subscripts have been suppressed and we assume all random variables have mean 0):

$$y = \beta x + \epsilon \quad (1)$$

$$x = Z\Pi + \nu, \quad (2)$$

where y , x , ϵ , and ν are $N \times 1$ vectors, and Z is a $N \times K$ matrix of random variables, while Π is a $K \times 1$ vector of constants and β is a scalar constant. We assume $E(\nu|Z) = 0$. Expanding (1) and (2) to include common exogenous variables would complicate the notation, but would not otherwise change the results.

A. Inconsistency Due to Correlation Between the Potential Instruments and the Error

We know that

$$\text{plim } \hat{\beta}_{\text{ols}} = \beta + \frac{\sigma_{x,\epsilon}}{\sigma_x^2}, \quad (3)$$

and

$$\text{plim } \hat{\beta}_{\text{iv}} = \beta + \frac{\sigma_{\hat{x},\epsilon}}{\sigma_{\hat{x}}^2}, \quad (4)$$

where $\sigma_{i,j}$ is the covariance between i and j , σ_i^2 is the variance of i , and \hat{x} represents the (population) projection of x onto Z . The relative inconsistency of IV is therefore

$$\frac{\text{plim } \hat{\beta}_{\text{iv}} - \beta}{\text{plim } \hat{\beta}_{\text{ols}} - \beta} = \frac{\sigma_{\hat{x},\epsilon}/\sigma_{x,\epsilon}}{R_{x,Z}^2}, \quad (5)$$

where $R_{x,Z}^2$ is the population R^2 from the regression of x on Z . When equations (1) and (2) include common exogenous variables, the relevant parameter is the partial R^2 , the population R^2 from the regression of x in Z once the common exogenous variables have been partialled out from both x and Z .

When $K = 1$ equation (5) can be rewritten as

$$\frac{\text{plim } \hat{\beta}_{\text{iv}} - \beta}{\text{plim } \hat{\beta}_{\text{ols}} - \beta} = \frac{\rho_{z,\epsilon}/\rho_{x,\epsilon}}{\rho_{x,z}} \quad (6)$$

where $\rho_{i,j}$ is the correlation between i and j . It is clear from (6) that a weak correlation between the potentially endogenous variable, x , and the instrument, z , will exacerbate any problems associated with a correlation between the instrument and the error, ϵ . If the correlation between the instrument and the endogenous explanatory variable is very weak even a very small correlation between the instrument and the error will produce a larger inconsistency in the IV estimate of β than in the OLS estimate.

B. Finite Sample Bias

We now assume that $E(\epsilon|Z) = 0$, implying that the columns of Z are legitimate instruments and that $\hat{\beta}_{iv}$ is a consistent estimator of β . In finite samples, however, $\hat{\beta}_{iv}$ is biased in the direction of the expectation of the OLS estimator of β . The magnitude of this bias depends on both the sample size (as the sample size increases the bias is reduced) and the multiple correlation between the instruments and the endogenous explanatory variable (as $R_{x,Z}^2$ increases the bias of $\hat{\beta}_{iv}$ decreases). The finite sample bias arises since we don't know Π but must instead use an estimate. Intuitively, this implies a certain amount of overfitting of the first stage equation leading to a bias in the direction of OLS. Consider the special case where the true value of each element of Π is 0; i.e. the instruments, Z , are completely unrelated to the endogenous explanatory variable, x . In any finite sample the estimates of the elements of Π will not be exactly equal to zero, however. It should not be too surprising that in this case the expectation of $\hat{\beta}_{iv}$ will be equal to the expectation of $\hat{\beta}_{ols}$.¹

Results on the magnitude of the finite sample bias of IV estimates extend back to the work of R. L. Basmann.² Under the assumption of joint normality, Richardson [1968] and Sawa [1969] independently derive expressions for the exact finite sample distribution of IV estimator in the case where there is only one endogenous explanatory variable but multiple instruments. In particular, Sawa shows that the finite sample bias of IV is in the same direction as the OLS bias and, in the limit as $R_{x,Z}^2$ approaches 0, is of the same magnitude as the OLS bias.

Alternatively, it is possible to derive approximations to the finite sample bias

¹ Nelson and Startz [1990a,b] present an interesting and intuitive discussion of the finite sample properties of the IV estimator for the special case of exact identification and one stochastic disturbance.

² For an introductory discussion of these issues see Bowden and Turkington [1984].

of the IV estimator without assuming normality using power series approximation methods. Buse [1992], building on earlier work by Nagar [1959], derives an expression for the approximate bias of $\hat{\beta}_{iv}$:

$$\frac{\sigma_{\epsilon,\nu}}{\Pi' Z' Z \Pi} (K - 2), \quad (7)$$

where K is the number of instruments.³ A little rearranging gives the approximate bias as

$$\frac{\sigma_{\epsilon,\nu}}{\sigma_\nu^2} \times \frac{\sigma_\nu^2}{\Pi' Z' Z \Pi} (K - 2). \quad (8)$$

Define τ^2 , the *concentration parameter* [Basmann, 1963], as $\tau^2 \equiv \frac{\Pi' Z' Z \Pi}{\sigma_\nu^2}$. Equation (8) implies that for $K > 2$ the IV bias is in the same direction as the OLS bias, and is approximately inversely proportional to $\frac{\tau^2}{K}$. This is the population analog to the F statistic on the instruments, Z , in the OLS estimation of equation (2).⁴ It should be noted, however, that the F statistic estimated from any particular (finite) sample will tend to overestimate $\frac{\tau^2}{K}$ for the same reason that the sample R^2 is an upward biased estimate of the population R^2 . Even so, equation (8) suggests that examining the F statistic on the excluded instruments in the first stage regression of IV is useful in gauging the finite sample bias of IV relative to OLS. An F statistic at all close to unity should be cause for concern.

It is possible to call into question the validity of employing power series approximation methods to study the finite sample properties of IV if one believes that the higher order terms in the expansion are quantitatively important. Under the assumption of normality, however, the exact distribution of the IV estimator can be derived. Drawing on Sawa's [1969] work, we derive the bias of the IV estimator using this assumption in the appendix. It again turns out to be the case that

³ When equations (1) and (2) include common exogenous variables, K is the number of excluded instruments.

⁴ When equations (1) and (2) include common exogenous variables, the relevant statistic would be analogous to the F statistic on the excluded instruments.

magnitude of the bias depends on the parameter $\frac{\tau^2}{K}$. It is worth noting that the exact sample results do not show the same knife edge at $K = 2$ as do the results based on power series methods. For moderately large K 's and small values of $\frac{\tau^2}{K}$ the power series methods show somewhat larger biases than do the results based on the assumption of normality. For moderately large K 's and values of $\frac{\tau^2}{K}$ larger than 2, the two methods show biases of similar magnitude. Details can be found in the appendix.

2. A Reexamination of Angrist and Krueger's Results

AK argue that compulsory school attendance laws account for the statistically significant relationship between quarter of birth and educational attainment. The typical law requires a student to remain in school until he or she turns 16. An individual born in the early months of the year will usually enter first grade when he or she is close to 7 and will reach the age of 16 in the middle of tenth grade. Individuals born in the third or fourth quarter will typically start school either just before or just after turning 6 and will finish tenth grade before reaching their 16th birthday. As a result, to the extent that compulsory school attendance laws are binding, those born late in the calendar year should be somewhat more likely to finish tenth grade than those born earlier in the year.

AK present several tabulations to support their claim that compulsory attendance laws are part of the mechanism generating a relationship between quarter of birth and educational attainment. First, educational attainment was higher for those born in the third and fourth quarter than for those born in the first and second quarter for cohorts born in the 1930's and 1940's. Second, the relationship between quarter of birth and educational attainment is weaker for more recent cohorts that would have been less likely to have been constrained by the law. Third, the relationship between quarter of birth and education is weaker for the better educated.

Lastly, the relationship between quarter of birth and educational attainment varies across states depending on when the state requires children to start school (see Angrist and Krueger [1992]). Each of these patterns is consistent with the assertion that compulsory school attendance laws are responsible for the association between quarter of birth and educational attainment. We are left with little doubt that compulsory attendance laws are working to induce a correlation between quarter of birth and educational attainment. Moreover, it is hard to imagine that they would induce any direct correlation between quarter of birth and wages. Quarter of birth dummy variables would therefore seem to be legitimate instruments for education in a wage equation.

Equation (5) suggests, however, that even a minimal direct correlation between quarter of birth and wages could seriously bias the IV estimates, given the weak correlation between quarter of birth and educational attainment. At issue, therefore, is not only whether compulsory school attendance laws induce a relationship between quarter of birth and educational attainment, but also whether these laws are the only reason for the relationship that apparently does exist between quarter of birth and wages.

While we know of no solid evidence on the direct effect of quarter of birth on earnings, it does not seem implausible that such effects might exist. There are a number of specific reasons that this might be true. First, quarter of birth is related to age at entry in the school system. If age at school entry affects performance in school (as many parents and educators believe) quarter of birth could have a small effect on earnings independent of its effect on educational attainment. Second, there is some evidence suggesting that there are identifiable differences between individuals born at different times of the year. There is evidence, for example, that individuals born early in the year are more likely to be mentally retarded [Knoblock

and Pasamanick, 1958], and somewhat mixed evidence regarding difference in IQ across children born at different times of the year [Whorton and Karnes, 1981]. There are also clear regional patterns in birth seasonality, with those living in the south less likely to give birth during the winter months and some evidence suggesting the same is true for those with low incomes [Lam and Miron, 1991]. While, as AK note, much of the evidence on the association between quarter of birth and these characteristics is suggestive rather than conclusive, the existing evidence does make it hard to assert with any confidence that there is absolutely no direct association between quarter of birth and earnings.

Are any of these seasonal effects large enough to matter? Equation (5) implies that even if these effects are weak they could still have large effects on the estimated coefficients. AK actually present results that are suggestive of the effects of quarter of birth on wages independent of the effect of quarter of birth on educational attainment. They report IV estimates that both control and do not control for race, urban status, marital status, and region of residence. In each case including these variables as controls reduces the IV estimates substantially more than their inclusion reduces the OLS estimates. For example, when AK add these controls to the OLS results reported in column (3) of Table V, the coefficient on education drops 11 percent from 0.071 to 0.063. In comparison, when they add the same controls to the IV estimates reported in column (4), the coefficient on education drops 18 percent from 0.076 to 0.060.⁵ This result implies that there is an association between quarter of birth and the control variables. Thus, for example, blacks are 0.7 percentage points more likely than whites to have been born during the winter quarter. Because blacks, on average, have lower educational attainment and

⁵ We experimented with deleting these controls one at a time. Each variable worked in the same direction with its inclusion lowering the estimate of the effect of education on weekly wages for both the OLS and IV estimates. In each case the effect on the IV estimates was greater than that on the OLS estimates.

earnings than whites, race, in part, accounts for the lower educational attainment and earnings among individuals born during the winter quarter. In the samples AK use the associations between quarter of birth and race, marital status and location of residence are all quite small. But even these small differences in the seasonal pattern of births have substantial effects on the estimated coefficient on education.

Because there is a strong association between age and earnings and because, in a cross section, quarter of birth and age are related, it is important to carefully control for age when using quarter of birth as an instrument. The first two panels of Table 1 present estimates of the effects of education on the logarithm of weekly earnings for men born between 1930 and 1939, and between 1940 and 1949, respectively. Columns (1) and (3) contain OLS estimates while columns (2), (3) and (4) contain results from a variety of IV specifications that include both different age controls and different instruments for education.⁶ We were able to replicate exactly AK's samples and results using the information in their Appendix 1 and refer the reader to that appendix for details regarding the samples. In addition to the coefficient and standard error⁷ on education, for each IV specification we report the *F* statistic for the test of the joint statistical significance of the excluded instruments in the first stage regression, the partial R^2 of the instruments in the first stage regression and Basmann's [1960] *F* test for overidentification.⁸

Substantial changes in the estimated magnitude of the coefficient on education when using different age controls should be cause for concern. Comparing the

⁶ Our column (3) replicates AK's column (7) and our column (5) replicates their column (8) from their tables V and VI.

⁷ We calculated heteroskedastic-consistent standard errors using the methods suggested by White [1980] and White [1982] for columns (1) and (2) of Table 1. In no instance did the White standard errors differ from the reported standard errors in the three decimal places reported.

⁸ This test is asymptotically equivalent to the Lagrange Multiplier test for overidentification that AK report [Spanos, 1986].

Table 1.
Estimated Effect of Years of Education on Men's Log Weekly Earnings

Standard errors of coefficients are in parentheses (). P-values of F statistics are in brackets [].

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) IV
Full Sample					
Men Born 1930-1939 (N = 329,509)					
Education	0.063 (0.000)	0.142 (0.033)	0.063 (0.000)	0.677 (1.536)	0.060 (0.029)
F (first stage)		13.486 [0.000]		0.088 [0.916]	1.613 [0.021]
Partial R ² (x 100)		0.012		0.000	0.014
F (overidentification)		0.932 [0.394]		0.082 [0.774]	0.725 [0.849]
Men Born 1940-1949 (N = 486,926)					
Education	0.052 (0.000)	0.201 (0.059)	0.052 (0.000)	-0.116 (0.255)	0.078 (0.024)
F (first stage)		6.256 [0.000]		0.546 [0.579]	2.736 [0.000]
Partial R ² (x 100)		0.004		0.000	0.016
F (overidentification)		4.359 [0.013]		6.775 [0.009]	1.873 [0.004]
Education ≤ 12 Years Sample					
Men Born 1930-1939 (N = 198,346)					
Education	0.056 (0.001)	0.164 (0.048)	0.056 (0.001)	0.176 (0.096)	0.062 (0.045)
F (first stage)		15.231 [0.000]		5.904 [0.003]	1.704 [0.012]
Partial R ² (x 100)		0.023		0.006	0.024
F (overidentification)		0.096 [0.908]		0.151 [0.698]	0.891 [0.628]
Men Born 1940-1949 (N = 233,780)					
Education	0.070 (0.001)	0.165 (0.049)	0.070 (0.001)	0.439 (0.213)	0.050 (0.051)
F (first stage)		20.484 [0.000]		3.008 [0.049]	1.875 [0.003]
Partial R ² (x 100)		0.026		0.003	0.022
F (overidentification)		6.488 [0.002]		5.158 [0.023]	2.076 [0.001]
Age Controls					
Age, Age ²	x	x	x	x	x
9 Year of Birth Dummies			x	x	x
Instruments					
Quarter of Birth	x			x	x
Quarter of Birth × Year of Birth				x	x

Notes:

Calculated from the 5% Public Use Micro Sample of the 1980 U.S. Census.

All specifications include Race (1=black), SMSA (1=central city), Married (1=married, living with spouse), and 8 Regional dummies as control variables.

F (first stage) and partial R² are for the instruments in the first stage of IV estimation.

F (overidentification) is that suggested by Basmann [1960].

coefficient estimates across the columns of the first two panels of Table 1 reveals substantial instability in those estimated using IV. Results from the simplest specification that includes controls age and age-squared (column (2)) reveal effects that are implausibly high. Adding year of birth dummies to this specification (column (4)) has large effects on both the coefficient and standard error estimates, implying that much of the identification for the coefficients in column (2) derives from variation in the effects of quarter of birth across single-year cohorts. Adding the quarter of birth \times year of birth interactions as instruments (column (5)) increases the precision of the estimates substantially. While there is some theoretical justification for including these interaction terms, we are concerned that most of the identifying information on the effect of education comes from these interactions. With these interactions included, the reported *F* statistics on the excluded instruments in the first stage are small enough to suggest that quantitatively important finite sample biases may affect the estimates. In addition, the standard errors on the specifications that include a quadratic in age as control variables are large enough to imply that the IV estimates have very little power to distinguish between plausible alternative point estimates.⁹

While AK do not impose restrictions on the educational attainment of the samples that they use, their argument for the effects of compulsory schooling laws on educational attainment (and therefore on wages) would seem to be strongest for individuals with no more than a high school education. Indeed, if the direct effect of compulsory school attendance laws on educational attainment identifies AK's results we should expect that most, if not all, of the identification would come from individuals with no more than 12 years of education. Limiting the samples

⁹ In addition to the specifications we report, AK also report results from a specification that includes controls for single year of birth cohort but does not include either a linear or quadratic term in age. Within single year birth cohorts, age (measured in quarters) and quarter of birth are perfectly collinear, however. This specification does not seem to us to be very sensible.

AK use to such individuals should not substantially influence the magnitude of the estimates and should actually improve their precision.¹⁰

The third and fourth panels of Table 1 present results for samples in which individuals received no more than 12 years of education. AK's methodology seems to work somewhat better for the full samples than it does for the restricted sample, however. The point estimates of the effects of education are somewhat more stable for the full samples, and the estimated standard errors are, in almost all cases, larger for the restricted sample. We conclude that at least some of the identification of the estimates for the full sample that AK report derives from men that would not have been constrained by the compulsory attendance laws. While we have no doubt that compulsory attendance laws did influence educational attainment and that these effects are part of what identifies the coefficients on education that AK report, there appear to be other forces at work as well. Without understanding what these forces are, it is impossible to know whether the IV estimates AK report could be expected to be asymptotically less biased than OLS.

Compulsory attendance laws as well as the degree to which they are enforced vary by state. AK use this cross-state variation to help identify the coefficient on education by including in their list of instruments state of birth \times quarter of birth interactions as instruments in some of their specifications. In addition to improving the precision of the estimates, using the variation across states of birth variation should mitigate problems of multicollinearity between age and quarter of birth. In Table 2 we report replications of some of AK's results using quarter of birth \times state of birth interactions for the full samples together with results for those with no

¹⁰ That limiting the sample should not substantially change the estimates is quite intuitive, and is explicitly proved within the context of their model, by Angrist and Imbens [1992]. The precision of the estimates with the limited sample should be greater because adding irrelevant observations is simply adding noise. It is possible to explicitly prove this using the Angrist and Imbens framework. We thank Imbens for pointing this out to us.

Table 2.
Estimated Effect of Years of Education on Men's Log Weekly Earnings
With State Effects

Standard errors of coefficients are in parentheses (). P-values of F statistics are in brackets [].

	(1) OLS	(2) IV	(3) IV	(4) OLS	(5) IV	(6) IV
Full Sample						
Men Born 1930-1939 (N = 329,509)						
Education	0.063 (0.000)	0.140 (0.032)	0.093 (0.011)	0.063 (0.000)	0.054 (0.029)	0.081 (0.011)
F (first stage)		14.772 [0.000]	2.144 [0.000]		1.706 [0.011]	1.869 [0.000]
Partial R ² ($\times 100$)		0.013	0.100		0.015	0.101
F (overidentification)		0.895 [0.409]	0.951 [0.654]		0.720 [0.853]	0.917 [0.779]
Men Born 1940-1949 (N = 486,926)						
Education	0.052 (0.000)	0.208 (0.060)	0.075 (0.012)	0.052 (0.000)	0.069 (0.024)	0.067 (0.011)
F (first stage)		6.319 [0.000]	1.883 [0.000]		2.682 [0.000]	1.929 [0.000]
Partial R ² ($\times 100$)		0.004	0.059		0.015	0.071
F (overidentification)		3.989 [0.019]	1.123 [0.142]		1.865 [0.004]	1.140 [0.098]
Education \leq 12 Years Sample						
Men Born 1930-1939 (N = 198,346)						
Education	0.057 (0.001)	0.164 (0.049)	0.116 (0.022)	0.057 (0.001)	0.060 (0.045)	0.092 (0.022)
F (first stage)		15.385 [0.000]	1.368 [0.002]		1.740 [0.009]	1.192 [0.041]
Partial R ² ($\times 100$)		0.023	0.105		0.025	0.107
F (overidentification)		0.069 [0.933]	1.061 [0.289]		0.892 [0.626]	1.064 [0.266]
Men Born 1940-1949 (N = 233,780)						
Education	0.070 (0.001)	0.167 (0.049)	0.108 (0.024)	0.070 (0.001)	0.045 (0.050)	0.078 (0.024)
F (first stage)		20.068 [0.000]	1.631 [0.000]		1.965 [0.002]	1.375 [0.001]
Partial R ² ($\times 100$)		0.026	0.107		0.024	0.105
F (overidentification)		6.318 [0.002]	0.941 [0.687]		1.993 [0.002]	1.036 [0.355]
Age Controls						
Age, Age ²	x	x	x	x	x	x
9 Year of Birth Dummies				x	x	x
Instruments						
Quarter of Birth	x	x		x	x	
Quarter of Birth \times Year of Birth				x	x	
Quarter of Birth \times State of Birth	x			x	x	

Notes:

Calculated from the 5% Public Use Micro Sample of the 1980 U.S. Census.

All specifications also include Race (1=black), SMSA (1=central city), Married (1=married, living with spouse), 8 Regional dummies, and 50 State of Birth dummies as control variables.

F (first stage) and Partial R² are for the instruments in the first stage of IV estimation.

F (overidentification) is that suggested by Baumann (1960).

more than 12 years of education.

Including the state of birth \times quarter of birth interactions reduces standard errors on the IV results by more than a factor of two and stabilizes the point estimates considerably. The *F* statistics on the excluded instruments in the first stage of IV deteriorate, however. In particular, for the 0 to 12 years of education sample the *F* statistics on the instruments in the first stage reported in Table 2 using the state of birth \times quarter of birth interactions as instruments (columns (3), (5), and (6)) range from 1.192 to 1.631. These *F* statistics would seem to indicate that while including state of birth \times quarter of birth interactions improves the precision and reduces the instability of the estimates, it magnifies the finite sample biases of these estimates substantially.

To illustrate this point, following a suggestion made by Alan Krueger, we reestimated column (3) of Table 2 for the 1930-1939 cohort using randomly generated information in place of the actual quarter of birth. We repeated this procedure 100 times. For the full sample, using the simulated quarters gave us a mean estimate of the coefficient on education of 0.060. The mean estimate of the standard error of the coefficient was 0.016. Limiting the sample to individuals with 0 to 12 years of education gave us a mean coefficient estimate of 0.053 with a mean estimated standard error of 0.026. Not surprisingly, the *F* statistics on the quarter of birth and quarter of birth \times state interactions were always close to their expected value of 1 in both samples. The similarity of the average point estimates to the OLS results and of the average estimated standard errors to those using the actual quarter of birth data is striking. Despite the fact that no information about individuals' educational attainment is contained in the simulated data, the computer output from the second stage regressions gives us no indication that this is true.

These results imply that if the correlation between the instruments and the

endogenous variable is low, even the enormous samples sizes available in the U.S. Census do not guarantee that quantitatively important finite sample biases will be eliminated from IV estimates. Because of these finite sample biases, the IV point estimates are biased in the same direction as are the OLS estimates.

3. Conclusion

These results illustrate that using instruments which jointly explain little of the variation in the endogenous variable them can do more harm than good. The example we chose to analyze is noteworthy because AK would have seemed to be on strong ground in choosing a valid instrument. They produce evidence supporting the notion that compulsory attendance laws induce a correlation between quarter of birth and educational attainment. Moreover, it seems implausible that there would be any very strong direct association between quarter of birth and wages. Still, as we have seen, these conclusions are not sufficient to ensure that using quarter of birth to instrument education will reduce the magnitude of the inconsistency inherent in the use of an endogenous variable. For this to be the case there would have to be essentially no direct association between quarter of birth and wages.

Having become acutely aware of the endogeneity of many of the variables whose impact we wish to study, we tend to believe that the use of plausible instruments will improve the validity of our inferences. While standard errors may be large, we imagine that we have eliminated most of bias inherent in the OLS estimates. However, equation (5) indicates that even with instruments that seem to be reasonably exogenous to the process under study, this may not be true. If, as is often the case, a set of potential instruments is relatively weakly correlated with the endogenous variable that needs instrumenting, even a weak correlation between the potential instruments and the error can seriously bias estimates.

Our results also show that working with large cross-sectional samples does not insulate us from quantitatively important finite sample biases. We have no way of knowing the extent to which this issue is empirically important for those working with such data, but our results suggest that even those working with large cross-sectional samples should be cautious about adding instruments to increase precision.

In practical terms, we suggest that when IV estimates are reported both the partial R^2 and F statistic on the excluded instruments in the first stage regression be routinely calculated and reported as rough guides to the quality of the IV estimates.

Appendix: The Exact Finite Sample Bias of IV

Instrumental variables estimates are biased in finite samples. As we have shown in the text, the magnitude of that bias can be approximated using power series expansion methods. It is also possible to derive the exact magnitude of the bias under the assumption of normality. In this appendix we expand on Sawa's [1969] work on the exact magnitude of the finite sample bias of IV to show that the two methods yield comparable results.

We are interested in estimating equation (A1) from the following system (for notational simplicity individual observation subscripts have been suppressed and we assume all random variables have mean 0):

$$y = \beta x + \epsilon \quad (\text{A1})$$

$$x = Z\Pi + \nu, \quad (\text{A2})$$

where y , x , ϵ , and ν are $N \times 1$ vectors, and Z is a $N \times K$ matrix of random variables, while Π is a $K \times 1$ vector of constants and β is a scalar constant. We assume that $E(\epsilon|Z) = 0$ and $E(\nu|Z) = 0$. In addition, because $E(\hat{\beta}_{iv})$ does not exist when $K = 1$, we assume $K > 1$.

Let the reduced form of equation (A1) be

$$y = Z\Pi_0 + \nu_0. \quad (\text{A3})$$

Define

$$\tau^2 \equiv \frac{\Pi' Z' Z \Pi}{\sigma_\nu^2} \quad (\text{A4})$$

and

$$\rho \equiv \frac{\sigma_{\nu, \nu_0}}{\sigma_\nu^2}. \quad (\text{A5})$$

Sawa shows that under the assumptions that ν and ν_0 are distributed as jointly normal, the OLS and IV biases can be written as

$$E(\hat{\beta}_{ols}) - \beta = (\beta - \rho) \left[1 - \frac{\tau^2}{N-1} {}_1F_1 \left(1, \frac{N+1}{2}; -\frac{\tau^2}{2} \right) \right] \quad (\text{A6})$$

$$E(\hat{\beta}_{iv}) - \beta = (\beta - \rho) \left[1 - \frac{\tau^2}{K} , F_1 \left(1, \frac{K+2}{2}; -\frac{\tau^2}{2} \right) \right] \quad (A7)$$

where $F_1(\cdot, \cdot, \cdot)$ is the confluent hyper-geometric function, defined as $F_1(\alpha, \gamma; \zeta) = \frac{\Gamma(\gamma)}{\Gamma(\alpha)} \sum_{j=0}^{\infty} \frac{\Gamma(\alpha+j)}{j! \Gamma(\gamma+j)} \zeta^j$. Note that $\frac{\tau^2}{N-1} \simeq R_{x,z}^2$, the population R^2 from the regression of x on Z , and that $\frac{\tau^2}{K}$ is population analog to F statistic for the regression of x on Z . For large values of N and small values of τ^2 , $F_1(1, \frac{N+1}{2}; -\frac{\tau^2}{2}) \simeq 1$. For large N and small $R_{x,z}^2$, therefore, the OLS bias approaches $\beta - \rho$ and the relative bias of IV approaches the expression in square brackets in (A7).

While the implication of equations (A6) and (A7) is far from obvious it is possible to approximate $F_1(1, \gamma; \zeta)$ for various values of γ and ζ . In Table A1 we present the magnitude of the bias of IV relative to OLS for various values of $\frac{\tau^2}{K}$ and K . Clearly, the bias of IV relative to OLS depends on the $\frac{\tau^2}{K}$, the population analogue to the F statistic on the excluded instruments.

It is worth noting that these exact sample results do not show the same knife edge at $K = 2$ as do the results based on power series methods. For moderately large K 's and small τ^2 's power series methods show somewhat smaller biases, while with large τ^2 's the two results agree.

Table A1.
Bias of IV Estimates
Relative to OLS Estimates
for Various Values of $\frac{r^2}{K}$ and K

		$\frac{r^2}{K}$					
		0.5	1.0	2.0	4.0	10.0	100.0
K	2	0.61	0.37	0.14	0.02	0.00	0.00
	3	0.62	0.41	0.21	0.09	0.03	0.00
	10	0.65	0.47	0.30	0.17	0.08	0.01
	20	0.66	0.49	0.32	0.19	0.08	0.01
	100	0.67	0.50	0.33	0.20	0.09	0.01
	200	0.67	0.50	0.33	0.20	0.09	0.01

Note: K is the number of excluded instruments and $\frac{r^2}{K}$ is the population analog to the F statistic for the joint statistical significance of the instruments in the first stage regression. Entries are $\left[1 - \frac{r^2}{K} F_1\left(1, \frac{K+2}{2}; -\frac{r^2}{2}\right)\right]$, which is the approximate bias of $\hat{\beta}_{iv}$ relative to $\hat{\beta}_{ols}$ when the R^2 between the instruments and the endogenous explanatory variable is small. Details are contained in the text.

References

- Angrist, Joshua D., and Guido W. Imbens** (1992), "Average Causal Response with Variable Treatment Intensity," NBER Technical Working Paper, No. 127, September.
- Angrist, Joshua D., and Alan B. Krueger** (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106, 979-1014.
- (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, 328-336.
- Basmann, R. L.** (1960), "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics," *Journal of the American Statistical Association*, 55, 650-659.
- (1963), "Remarks Concerning the Application of Exact Finite Sample Distribution Functions of GCL Estimators in Econometric Statistical Inference," *Journal of the American Statistical Association*, 58, 943-76.
- Bowden, Roger J. and Darrell A. Turkington** (1984), *Instrumental Variables*, Cambridge: University Press, 137-144.
- Buse, A.** (1992), "The Bias of Instrumental Variable Estimators," *Econometrica*, 60, 173-180.
- Knoblock, Hilda, and Benjamin Pasamanick** (1958), "Seasonal Variation in the Births of the Mentally Deficient," *American Journal of Public Health*, 48, 1201-1208.
- Lam, David and Jeffrey A. Miron** (1991) "Seasonality of Births," *Social Biology*, 38, 51-78.
- Nagar, A. L.** (1959), "The Bias and Moment Matrix of the General k -Class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27, 575-595.
- Nelson, C. E. , and R. Startz** (1990a), "Some Further Results on the Exact Small Sample Properties of the IV Estimator," *Econometrica*, 58, 967-976.
- (1990b), "The Distribution of the Instrumental Variable Estimator and its t-Ratio when the Instrument is a Poor One," *Journal of Business*, 63 part 2, S125-S140.
- Richardson, David H.** (1968), "The Exact Distribution of a Structural Coefficient Estimator," *Journal of the American Statistical Association*, 63, 1214-1226.
- Sawa, Takamitsu** (1969), "The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators," *Journal of the American Statistical Association*, 64, 923-937.

Spanos, Aris (1986), *Statistical Foundations of Econometric Modeling*, Cambridge: University Press, 651–653.

Whorton, James E., and Frances A. Karnes (1981), "Season of Birth and Intelligence in Samples of Exceptional Children," *Psychological Reports*, 49, 649–650.

White, Halbert (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

— (1982), "Instrumental Variables Estimation with Independent Observations," *Econometrica*, 50, 483–499.