

Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak

John BOUND, David A. JAEGER, and Regina M. BAKER*

We draw attention to two problems associated with the use of instrumental variables (IV), the importance of which for empirical work has not been fully appreciated. First, the use of instruments that explain little of the variation in the endogenous explanatory variables can lead to large inconsistencies in 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 endogenous explanatory variable approaches 0. To illustrate these problems, we reexamine the results of a recent paper by Angrist and Krueger, who used large samples from the U.S. Census to estimate wage equations in which quarter of birth is used as an instrument for educational attainment. We find evidence that, despite huge sample sizes, their IV estimates may suffer from finite-sample bias and may be inconsistent as well. These findings suggest that valid instruments may be more difficult to find than previously imagined. They also indicate that the use of large data sets does not necessarily insulate researchers from quantitatively important finite-sample biases. We suggest that the partial R^2 and the F statistic of the identifying instruments in the first-stage estimation are useful indicators of the quality of the IV estimates and should be routinely reported.

KEY WORDS: Compulsory attendance; Finite-sample bias; Inconsistency; Weak instrument.

1. INTRODUCTION

Empirical researchers often wish to make causal inferences about the effect of one variable on another. Doing so in nonexperimental settings is frequently difficult, because some of the explanatory variables are endogenous; that is, they are influenced by some of the same forces that influence the outcome under study. For example, economists examining the effect of education on earnings have long been concerned about the endogeneity of education. It seems quite plausible that the same unobserved factors might influence both individuals' educational attainment and their earnings (see, for example, Griliches 1977). "Ability" is often cited as one factor possibly correlated with earnings (those with higher ability earn more) and education (those with higher ability obtain more education).

When explanatory variables are endogenous, ordinary least squares (OLS) gives biased and inconsistent estimates of the causal effect of an explanatory variable on an outcome. A common strategy for dealing with this endogeneity is to use instrumental variables (IV) estimation, using as "instruments" variables thought to have no direct association with the outcome. The exogenous instruments allow the researcher to partition the variance of the endogenous explan-

atory variable into exogenous and endogenous components. The exogenous component is then used in estimation. More specifically, the IV estimator uses one or more instruments to predict the value of the potentially endogenous regressor. The predicted values are then used as a regressor in the original model. Under the assumptions that the instruments are correlated with the endogenous explanatory variable but have no direct association with the outcome under study, the IV estimates of the effect of the endogenous variable are consistent. Bowden and Turkington (1984) and Angrist, Imbens, and Rubin (forthcoming) have provided useful introductions to IV estimation.

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 article we draw attention to two other problems associated with the use of such instruments. First, if the instruments are only weakly correlated with the 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 IV estimates. Second, in finite samples, IV estimates are biased in the same direction as OLS estimates, with the magnitude of the bias approaching that of OLS as the R^2 between the instruments and the endogenous explanatory variable approaches 0. Though these results are known, their general importance for empirical work has not been fully appreciated.

To illustrate these issues, we reexamine the results of a provocative paper by Angrist and Krueger (1991; henceforth denoted by AK-91). This paper used the large samples available in the U.S. Census to estimate wage equations where quarter of birth is used as an instrument for educational

* John Bound is Associate Professor, Department of Economics, University of Michigan, Ann Arbor, MI 48109, and Faculty Research Fellow, National Bureau of Economic Research, Cambridge, MA 02138. David A. Jaeger is a Ph.D. candidate, Department of Economics, University of Michigan, Ann Arbor, MI 48109. Regina M. Baker is a Ph.D. candidate, Department of Political Science, University of Michigan, Ann Arbor, MI 48109. Computations for this article were funded by the Population Studies Center and the Computational Sciences Division, Information Technology Division at the University of Michigan. The authors thank Chris Cavanagh, Gary Chamberlain, Zvi Griliches, Guido Imbens, David Lam, Kevin M. Murphy, and seminar participants at the University of Michigan for helpful discussions. The authors also thank Joshua Angrist, Eli Berman, Charles Brown, Alan Krueger, Jeffrey Grogger, Kevin Lang, Lung-Fei Lee, Marianne Page, two anonymous referees, and especially Gary Solon for helpful comments on earlier drafts and Cynthia Glovinsky for assistance in preparing the manuscript. The work was stimulated in part by a question asked of the first author by Sanders Korenman.

attainment. Although quarter of birth is only weakly related to educational attainment—the R^2 in the regression of educational attainment on quarter of birth ranged between .0001 and .0002 in their samples—the authors obtained reasonable standard errors on their estimates of the effect of education on weekly earnings, due to the large samples they used.

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-91's estimates. We also present results that indicate that the finite-sample bias may be quantitatively significant for some of the estimates that AK-91 reported. Together these results suggest that the "natural experiment" afforded by the interaction between compulsory school attendance laws and quarter of birth does not give much usable information regarding the causal effect of education on earnings.

Our results illustrate the general significance of these issues. In particular, they suggest that it may be even harder than many have thought to find legitimate instruments for potentially endogenous variables. Although 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 estimates produced using OLS. In addition, these results indicate that even researchers working with very large data sets need to be more concerned about the finite-sample properties of IV estimators.

2. POTENTIAL PROBLEMS USING AN INSTRUMENT THAT IS WEAKLY CORRELATED WITH THE ENDOGENOUS EXPLANATORY VARIABLE

We are interested in estimating β , the causal effect of x on y in Equation (1) from the following system (in which, for simplicity, we assume that all random variables have mean 0):

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

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

where x , ε , and ν are $N \times 1$ vectors of independent realizations of the random variables x , ε , and ν , respectively, y is an $N \times 1$ vector, Z is an $N \times K$ matrix in which the rows are independent realizations of the vector z , composed of random variables $z_1, \dots, z_k$, Π is a $K \times 1$ vector of constants, and β is a scalar constant. Note that (1) differs from the formulation familiar to statisticians in that we do not assume that x is uncorrelated with the error term ε , and therefore βx may not be the conditional mean of y given x . We assume $E(\nu|z) = 0$. The IV estimator of β is

$$\hat{\beta}_{iv} = (x'P_Z x)^{-1} x'P_Z y, \quad (3)$$

where $P_Z = Z(Z'Z)^{-1}Z'$, the projection matrix for Z . This is numerically equivalent to estimating (1) and (2) by two-stage least squares, where the first stage [equation (2)] is estimated by OLS and the predicted values from this estimation, $\hat{x} = Z\hat{\Pi}$, are used in place of x in the second-stage estimation of Equation (1) by OLS. Expanding (1) and (2)

to include common exogenous variables would complicate the notation but would not otherwise change the results.

It is straightforward to show that

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

and

$$\text{plim } \hat{\beta}_{iv} = \beta + \frac{\sigma_{\hat{x},\varepsilon}}{\sigma_{\hat{x}}^2} \quad (5)$$

where σ_{ij} 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 .

Equation (4) indicates that $\sigma_{x,\varepsilon}^2$ must be nonzero and x must be uncorrelated with ε for $\hat{\beta}_{ols}$ to be a consistent estimator for β . Similarly, Equation (5) makes clear that $\sigma_{\hat{x}}^2$ must be nonzero (i.e. that there must be an association between x and z) and that z (and therefore $\hat{x}$) must be uncorrelated with ε for $\hat{\beta}_{iv}$ to be a consistent estimator of β . That is, $\hat{\beta}_{iv}$ is consistent only if there is no direct association between z and y .

2.1 Inconsistency

Equations (4) and (5) imply that the inconsistency of IV relative to OLS is

$$\frac{\text{plim } \hat{\beta}_{iv} - \beta}{\text{plim } \hat{\beta}_{ols} - \beta} = \frac{\sigma_{\hat{x},\varepsilon} / \sigma_{x,\varepsilon}}{R_{x,z}^2}, \quad (6)$$

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 on z once the common exogenous variables have been partialled out of both x and z .

When $K = 1$, Equation (6) can be rewritten as

$$\frac{\text{plim } \hat{\beta}_{iv} - \beta}{\text{plim } \hat{\beta}_{ols} - \beta} = \frac{\rho_{z,\varepsilon} / \rho_{x,\varepsilon}}{\rho_{x,z}}, \quad (7)$$

where $\rho_{i,j}$ is the correlation between i and j . It is clear from Equation (7) that a weak correlation between the potentially endogenous variable, x , and the instrument, z_1 , 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 weak, then even a small correlation between the instrument and the error can produce a larger inconsistency in the IV estimate of β than in the OLS estimate.

It is instructive to examine the special case where $K = 1$ and z is dichotomous, partitioning the sample into two distinct subpopulations. Let $\bar{y}_1$, $\bar{y}_2$, $\bar{x}_1$, and $\bar{x}_2$ represent the subpopulation means of y and x , and define $\Delta\bar{y} \equiv \bar{y}_2 - \bar{y}_1$ and $\Delta\bar{x} \equiv \bar{x}_2 - \bar{x}_1$. The IV estimator of β can then be written as

$$\hat{\beta}_{iv} = \frac{\Delta\bar{y}}{\Delta\bar{x}}, \quad (8)$$

which is the Wald estimator of β (Durbin 1954; Wald 1940).

It is easy to see the possible inconsistency of the Wald estimator. If we take the difference across the two groups of z then Equation (1) becomes

$$\Delta \bar{y} = \beta \Delta \bar{x} + \Delta \bar{\varepsilon}, \quad (9)$$

where $\Delta \bar{\varepsilon} = \bar{\varepsilon}_2 - \bar{\varepsilon}_1$. Dividing by $\Delta \bar{x}$ and taking the probability limit gives

$$\text{plim } \hat{\beta}_{iv} = \beta + \frac{\text{plim } \Delta \bar{\varepsilon}}{\text{plim } \Delta \bar{x}}. \quad (10)$$

$\hat{\beta}_{iv}$ will be inconsistent if $\text{plim } \Delta \bar{\varepsilon} \neq 0$. The magnitude of the inconsistency will depend on the extent to which $\bar{y}_1$ and $\bar{y}_2$ differ for reasons unrelated to differences between $\bar{x}_1$ and $\bar{x}_2$ (i.e., $\text{plim } \Delta \bar{\varepsilon} \neq 0$) and on the magnitude of $\text{plim } \Delta \bar{x}$. Even very small direct effects of z on y will matter if $\text{plim } \Delta \bar{x}$ is small.

Although these results are not new (see Angrist, Imbens, and Rubin 1993; Bartels 1991, and Shea 1993 for recent discussions that mirror our own), their importance has been largely ignored by empirical researchers.

2.2 Finite-Sample Bias

We now assume that $E(\varepsilon|z) = 0$, implying that the instruments z are legitimate 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 $R^2_{x,z}$ increases, the bias of $\hat{\beta}_{iv}$ decreases). The finite-sample bias arises because we do not know the first-stage coefficients, Π , but instead must use estimates. 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 zero; that is, the instruments, z , are completely unrelated to the endogenous explanatory variable, x . But in any finite sample, the estimates of the elements of Π will not be exactly equal to zero. The decomposition of x into components $\hat{x}$ and $\hat{v}$ (where $\hat{v} = x - \hat{x}$) is arbitrary in this case. It thus seems quite natural that in this case, the expectation of $\hat{\beta}_{iv}$ would equal the expectation of $\hat{\beta}_{ols}$. The sampling variability of the two estimators will not be the same, of course. Interesting and intuitive discussions of the finite-sample properties of the IV estimator for the special case of exact identification and one stochastic disturbance have been presented by Nelson and Startz (1990a,b).

Results on the magnitude of the finite-sample bias of IV estimates extend back to the work of R. L. Basman. Under the assumption of joint normality, Richardson (1968) and Sawa (1969) independently derived expressions for the exact finite-sample distribution of the IV estimator in the case of only one endogenous explanatory variable but multiple instruments. In particular, Sawa showed that the finite-sample bias of IV is in the same direction as the OLS bias and, in the limit as $R^2_{x,z}$ approaches 0, is of the same magnitude as the OLS bias.

Alternatively, it is possible to derive approximations to the finite-sample bias of the IV estimator without assuming normality using power series approximation methods. Buse (1992), building on earlier work by Nagar (1959), derived an expression for the approximate bias of $\hat{\beta}_{iv}$ in samples of size N ,

$$\frac{\sigma_{\varepsilon, v}}{\Pi'Z'Z\Pi} (K - 2), \quad (11)$$

where K is the number of instruments; when Equations (1) and (2) include common exogenous variables, K is the number of excluded instruments. A little rearranging gives the approximate bias as

$$\frac{\sigma_{\varepsilon, v}}{\sigma_v^2} \cdot \frac{\sigma_v^2}{\Pi'Z'Z\Pi} (K - 2). \quad (12)$$

Note that $\sigma_{\varepsilon, v}/\sigma_v^2$ is approximately equal to the asymptotic bias of $\hat{\beta}_{ols}$ when z explains little of the variation of x . Define τ^2 , the *concentration parameter* (Basman 1963), as $\tau^2 \equiv (\Pi'Z'Z\Pi)/\sigma_v^2$. Equation (12) implies that for $K > 2$, the bias of the IV estimator of β relative to OLS is approximately inversely proportional to τ^2/K . This is the population analog to the F statistic on the instruments, Z , in the OLS estimation of Equation (2). When Equations (1) and (2) include common exogenous variables, the relevant statistic is analogous to the F statistic on the excluded instruments. It should be noted that the F statistic estimated from any particular (finite) sample will tend to overestimate τ^2/K for the same reason that the sample R^2 is an upward-biased estimate of the population R^2 . Even so, Equation (12) 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.

It is possible to call into question the validity of using power series approximation methods to study the finite-sample properties of IV. At issue is the potential importance of the higher-order terms in the expansion. 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 relative bias of the IV estimator using this assumption in the Appendix. When $K = 1$ (i.e., the system is just identified), the expectation of $\hat{\beta}_{iv}$ does not exist. When $K > 1$, it again turns out to be true that magnitude of the bias depends on the parameter τ^2/K . It is worth noting that unlike the results based on power series methods, the exact finite-sample results do not show a knife edge at $K = 2$. For moderately large K 's and small values of τ^2/K , the power series methods show somewhat larger biases than the results based on the assumption of normality. The two methods show relative biases of similar magnitude for moderately large K 's and values of τ^2/K larger than 2. Details can be found in the Appendix.

In a recent paper, Staiger and Stock (1994) took a different approach to analyzing the finite-sample properties of IV estimates. They developed an asymptotic distribution theory that does not rely on approximation or on the assumption of normality for IV estimates with weak instruments. Their approach holds the first-stage coefficients, Π , in an $N^{-1/2}$

neighborhood of zero as the sample size increases. In this context, they showed that $F - 1$ is an asymptotically unbiased estimator of τ^2/K and that $1/(1 + \tau^2/K)$ approximates the magnitude of the finite-sample bias of IV relative to OLS. This implies that $1/F$ is an approximate estimate of the finite-sample bias of $\hat{\beta}_{iv}$ relative to $\hat{\beta}_{ols}$.

All three approaches suggest that the bias of IV relative to OLS is a function of τ^2/K . If the relationship between the instruments and the endogenous explanatory variable is weak enough, even enormous samples do not eliminate the possibility of quantitatively important finite-sample biases. Each approach suggests that the first-stage F statistic contains valuable information about the magnitude of the finite-sample bias and that F statistics close to 1 should be cause for concern.

3. A REEXAMINATION OF ANGRIST AND KRUEGER'S RESULTS

3.1 Inconsistency

AK-91 used quarter of birth as an instrument for educational attainment in wage equations. In a subsequent article (Angrist and Krueger 1992, henceforth denoted by AK-92), they used quarter of birth as an instrument for age at school entry in educational attainment equations. For quarter of birth to be a legitimate instrument in the first case, its effect on educational attainment must be the only reason for its association with earnings. Similarly, for quarter of birth to be a legitimate instrument in the second case, the only reason for its association with educational attainment must be its effect on age at school entry.

AK-91 and AK-92 documented significant associations between quarter of birth and age at school entry, educational attainment, and earnings for cohorts of men born during the 1930s and 1940s. Individuals born during the first quarter of the year start school later, have lower educational attainment, and earn less than those born during the rest of the year.

Angrist and Krueger argued that compulsory school attendance laws account for these associations. The typical law requires a student to start first grade in the fall of the calendar year in which he or she turns age 6 and to continue attending school until he or she turns 16. Thus an individual born in the early months of the year will usually enter first grade when he or she is close to age 7 and will reach age 16 in the middle of tenth grade. An individual born in the third or fourth quarter will typically start school either just before or just after turning age 6 and will finish tenth grade before reaching age 16.

AK-91 and AK-92 presented 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, 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. Second, the relationship between quarter of birth and education is weaker for better-educated individuals. Third, the relationship between quarter of birth and educational attainment varies across states, de-

pending on when each state requires children to start school. 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.

Given the evidence that Angrist and Krueger presented, we are left with little doubt that compulsory attendance laws are working to induce a correlation between quarter of birth and educational attainment. We question, however, whether these laws are the *only* reason for this correlation, and, therefore, whether quarter of birth is a legitimate instrument in estimating either wage or educational attainment equations. The relationship between quarter of birth and age at school entry must be the *only* reason for the association between quarter of birth and educational attainment for quarter of birth to be a legitimate instrument for age at school entry in educational attainment equations. Similarly, there cannot be any direct association between quarter of birth and wages for quarter of birth to be a legitimate instrument for educational attainment in wage equations. Because the association between quarter of birth and education is very weak, Equation (6) indicates that even a small direct association between quarter of birth and wages is likely to badly bias the estimated coefficient on education in wage equations.

Although we know of no indisputable evidence on the direct effect of quarter of birth on education or earnings, it seems quite plausible that such effects exist. First, quarter of birth may affect a student's performance in school. There is some evidence that quarter of birth is related to school attendance rates (Carroll 1992), the likelihood that a student will be assessed as having behavioral difficulties (Mortimore, Sammons, Stoll, Lewis, and Ecob 1988), the likelihood that a student will be referred for mental health services (Tarnowski, Anderson, Drabman, and Kelly 1990), and performance in reading, writing, and arithmetic (Mortimore et al. 1988; Williams, Davies, Evans, and Ferguson 1970 for a summary of earlier literature). Second, there are identifiable differences in the physical and mental health of individuals born at different times of the year. There is substantial evidence that individuals born early in the year are more likely to suffer from schizophrenia (see, for example, O'Callaghan et al. 1991, Sham et al. 1992, and Watson, Kucala, Tilleskjor, and Jacobs 1984). There is also evidence of variation by quarter of birth in the incidence of mental retardation (Knoblock and Pasamanick 1958), autism (Gillberg 1990), dyslexia (Livingston, Adam, and Bracha 1993), multiple sclerosis (Templer et al. 1991), and manic depression (Hare 1975), as well as somewhat mixed evidence regarding differences in IQ among children born at different times of the year (Whorton and Karnes 1981). Third, there are clear regional patterns in birth seasonality (Lam and Miron 1991). Finally, there is some evidence suggesting that those in families with high incomes (Kestenbaum 1987) are less likely to be born in the winter months. Although the evidence for some of these associations is weaker than for others, and not all of the effects are in the same direction, the existing literature casts doubt on the notion of negligible direct associations between quarter of birth and either educational attainment or earnings.

Are any of these seasonal effects large enough to cause large biases on the coefficient of interest in the estimation of either educational attainment or wage equations? As noted earlier, the family income of those born early in the year tends to be lower than the family income of those born later in the year. Our calculations using the 1980 U.S. Census show that the difference in mean log per capita family income between those born in the first quarter of the year and those born in the second through fourth quarters of the year is $-.0238$ among children age 0 to 3 years. Ideally, we would like an estimate of this difference for the samples of men born in the 1930s and 1940s that Angrist and Krueger analyzed. But because seasonal variation in fertility has declined over the last 50 years (Lam and Miron 1991; Seiver 1985), our estimate of differences in family income by children's season of birth is likely to underrepresent the difference for men born in that period. Using data from the Panel Study of Income Dynamics (unreported results using the same sample as in Solon 1992) found that in a regression of children's educational attainment on the log of fathers' earnings, a 1% rise in fathers' earnings was associated with a .014-year rise in children's educational attainment. The difference in family income between those born in the first quarter and those born in subsequent quarters thus can explain approximately a .03 grade difference in educational attainment between these groups. AK-91 reported that for men born between 1930 and 1939, those born in the first quarter had on average .1 year less of educational attainment than those born in the second through fourth quarters. Thus differences in family income across those born in different quarters would seem to be capable of explaining about one-third of the association between quarter of birth and educational attainment.

In terms of wages, the weak association between educational attainment and quarter of birth indicates that even if other seasonal effects are weak, they could still have large effects on the estimated coefficients. AK-91 actually presented results suggesting the effects of quarter of birth on wages independent of the effect of quarter of birth on educational attainment. It reported IV estimates that controlled and did 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, in AK-91 when these controls were added to the OLS results reported in column (3) of Table V, the coefficient on education dropped 11%, from .071 to .063. In comparison, when the same controls were added to the IV estimates reported in column 4, the coefficient on education dropped 21%, from .076 to .060. This result implies an association between quarter of birth and the control variables. For example, blacks are .7% more likely than whites to have been born during the winter quarter. Because blacks on average have lower educational attainment and earnings than whites, race partially accounts for the lower educational attainment and earnings among individuals born during the winter quarter. In the samples used in AK-91, the associations between quarter of birth and other covariates (race, marital status, and location of residence) are all quite small. But even these

small differences in the seasonal birth pattern have substantial effects on the estimated coefficient on education.

The earlier calculation using differences in per capita family income across quarter of birth suggests that omitted factors could easily account for all of the small association between earnings and quarter of birth. As noted previously, we found that the difference in mean log per capita family income among children age 0 to 3 years between those born in the first quarter of the year and those born in the second through fourth quarters of the year is $-.0238$. Solon (1992) and Zimmerman (1992) both found an intergenerational correlation in long-run income of at least .4. Given their results, the differences in family incomes between those born in the first quarter and the rest of the year would lead us to expect that men born in the winter to earn about .95% lower wages than those born during the rest of the year. Data from the 1980 U.S. Census show that among men born during the 1930s, those born during the first quarter earn 1.1% lower wages (AK-91). Thus differences in family income at time of birth would seem to account for virtually all of the association between quarter of birth and wages.

3.2 Finite-Sample Bias

Because quarter of birth and educational attainment are only very weakly correlated, AK-91's estimates may be subject to finite-sample bias even with its enormous samples. To investigate this possibility, we reexamined these data drawn from the U.S. Census. AK-91 used three 10-year cohorts, emphasizing results for men born between 1930 and 1939. We use this sample to illustrate our points. We were able to replicate exactly AK-91's samples and results using the information in its Appendix 1, and we refer the reader there for details regarding sample and variable creation.

Table 1 presents estimates of the effects of education on the logarithm of men's weekly earnings. Columns (3) through (6) replicate AK-91's Table V, columns (5) through (8). Columns (3) and (5) present OLS estimates, and columns (4) and (6) present IV estimates, with quarter of birth and quarter of birth interacted with year of birth as the excluded instruments. Columns (3) and (4) differ from columns (5) and (6) in the age controls used. Columns (3) and (4) use single year of birth dummies, whereas in columns (5) and (6) age and age squared, measured in quarter years, are added. In addition to these replications, in columns (1) and (2) we report results from a simpler specification than those used in AK-91. These models use only age and age squared as controls for age, and column (2) uses quarter of birth dummies without year-of-birth interactions as instruments. We report the coefficient and standard error on education from each OLS and second-stage IV regression. In addition, we report the F statistic for the test of the joint statistical significance of the excluded instruments and the partial R^2 of the excluded instruments from the first-stage regression of each IV specification. We also report Basman's (1960) F test for overidentification for each IV specification.

The F statistic on the excluded instruments decreases across the different IV specifications in Table 1. In the simplest specification that includes only three quarter-of-birth dummies as instruments, the F statistic suggests negligible

Table 1. *Estimated Effect of Completed Years of Education on Men's Log Weekly Earnings*
(standard errors of coefficients in parentheses)

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Coefficient	.063 (.000)	.142 (.033)	.063 (.000)	.081 (.016)	.063 (.000)	.060 (.029)
<i>F</i> (excluded instruments)		13.486		4.747		1.613
Partial R^2 (excluded instruments, $\times 100$)		.012		.043		.014
<i>F</i> (overidentification)		.932		.775		.725
<i>Age Control Variables</i>						
Age, Age ²	x	x			x	x
9 Year of birth dummies			x	x	x	x
<i>Excluded Instruments</i>						
Quarter of birth		x		x		x
Quarter of birth $\times$ year of birth				x		x
Number of excluded instruments		3		30		28

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509. 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 R^2 are for the instruments in the first stage of IV estimation. *F* (overidentification) is that suggested by Basman (1960).

finite-sample bias. Because quarter of birth is related, by definition, to age measured in quarters within a single year of birth, and because age is an important determinant of earnings, we find the specification using within-year age controls [column (6)] to be more sensible than the specification that does not [column (4)]. The *F* statistic on the excluded instruments in column (6) indicates that quantitatively important finite-sample biases may affect the estimate. Comparing the partial R^2 in columns (2) and (6) shows that adding 25 instruments does not change the explanatory power of the excluded instruments by very much, explaining why the *F* statistic deteriorates so much between the two specifications.

Compulsory attendance laws, and the degree to which these laws are enforced, vary by state. In AK-91 the authors used this cross-state variation to help identify the coefficient on education by including state of birth $\times$ quarter of birth interactions as instruments in some of their specifications. Besides improving the precision of the estimates, using variation across state of birth should mitigate problems of multicollinearity between age and quarter of birth. In Table 2 we report replications of AK-91's Table VII, columns (5) through (8). These models use quarter of birth $\times$ state of birth interactions in addition to quarter of birth and quarter of birth $\times$ year of birth interactions as instruments for educational attainment.

Including the state of birth $\times$ quarter of birth interactions reduces the 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 do not improve, however. These *F* statistics indicate that although including state of birth $\times$ quarter of birth interactions improves the precision and reduces the instability of the estimates, the possibility that small-sample bias may be a problem remains.

To illustrate that second-stage results do not give us any indication of the existence of quantitatively important finite-sample biases, we reestimated Table 1, columns (4) and (6),

and Table 2, columns (2) and (4), using randomly generated information in place of the actual quarter of birth, following a suggestion by Alan Krueger. The means of the estimated standard errors reported in the last row are quite close to the actual standard deviations of the 500 estimates for each model. Moreover, the distribution of the estimates appears to be quite symmetric. In these cases, therefore, the asymptotic standard errors give reasonably accurate information on the sampling variability of the IV estimator. This is specific to these cases, however. Nelson and Startz (1990a) showed, in the context of a different example, that asymptotic standard errors can give very misleading information about the actual sampling distribution of the IV estimator when the correlation between the instrument and the endogenous variable is weak.

Table 2. *Estimated Effect of Completed Years of Education on Men's Log Weekly Earnings, Controlling for State of Birth*
(standard errors of coefficients in parentheses)

	(1) OLS	(2) IV	(3) OLS	(4) IV
Coefficient	.063 (.000)	.083 (.009)	.063 (.000)	.081 (.011)
<i>F</i> (excluded instruments)		2.428		1.869
Partial R^2 (excluded instruments, $\times 100$)		.133		.101
<i>F</i> (overidentification)		.919		.917
<i>Age Control Variables</i>				
Age, Age ²			x	x
9 Year of birth dummies	x	x	x	x
<i>Excluded Instruments</i>				
Quarter of birth		x		x
Quarter of birth $\times$ year of birth		x		x
Quarter of birth $\times$ state of birth		x		x
Number of excluded instruments		180		178

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509. All specifications 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 R^2 are for the instruments in the first stage of IV estimation. *F* (overidentification) is that suggested by Basman (1960).

It is striking that the second-stage results reported in Table 3 look quite reasonable even with no information about educational attainment in the simulated instruments. They give no indication that the instruments were randomly generated. As the analytic results in Section 2.2 would suggest, the mean of the estimated coefficients in each case is close to the comparable OLS estimate. The estimated standard errors look encouraging and are only somewhat larger than the ones reported for the comparable IV estimates reported in Tables 1 and 2 where actual quarter of birth is used as the instrument. On the other hand, the F statistics on the excluded instruments in the first-stage regressions are always near their expected value of essentially 1 and do give a clear indication that the estimates of the second-stage coefficients suffer from finite-sample biases.

We conclude that it is likely that some of the results reported in AK-91 are affected by quantitatively large finite-sample biases. Our results imply that if the correlation between the instruments and the endogenous variable is small, then even the enormous sample sizes available in the U.S. Census do not guarantee that quantitatively important finite-sample biases will be eliminated from IV estimates. They also indicate that the common practice of adding interaction terms as excluded instruments may exacerbate the problem, even while reducing the standard error of the coefficient on the endogenous explanatory variable.

4. CONCLUSION

These results illustrate that the use of instruments that jointly explain little of the variation in the endogenous variable can do more harm than good. The example we chose to analyze is noteworthy, because Angrist and Krueger would have seemed to be on strong ground in choosing a valid instrument. They produced 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. We have shown, however, that these conclusions are not sufficient to ensure that the use of quarter of birth as an instrument for educational attainment will reduce the magnitude of the inconsistency inherent in the use of an endogenous variable as a regressor.

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. Although standard errors may be large, we imagine that we have eliminated most of bias inherent in the OLS estimates. But Equation (6) indicates that this may not be true even with instruments that seem reasonably exogenous to the process under study. If a set of potential instruments is weakly correlated with the endogenous explanatory variable (as is often the case), then even a small correlation between the potential instruments and the error can seriously bias estimates.

We also show that working with large samples does not insulate us from quantitatively important finite-sample biases. Although we have no way of knowing to what extent this issue is empirically important for those working with such data, our results suggest that even those working with

Table 3. Estimated Effect of Completed Years of Education on Men's Log Weekly Earnings, Using Simulated Quarter of Birth (500 replications)

Table (column)	*	1 (4)	1 (6)	2 (2)	2 (4)
<i>Estimated Coefficient</i>					
Mean		.062	.061	.060	.060
Standard deviation of mean		.038	.039	.015	.015
5th percentile		-.001	-.002	.034	.035
Median		.061	.061	.060	.060
95th percentile		.119	.127	.083	.082
<i>Estimated Standard Error</i>					
Mean		.037	.039	.015	.015

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509.

large cross-sectional samples should be cautious about adding instruments to increase precision.

Our results emphasize the importance of examining characteristics of the first-stage estimates. The standard errors reported in AK-91 appear reasonable, and overidentification tests give no indication that the authors' models are misspecified. But despite these observations, we have shown that there is good reason to doubt the validity of inferences they drew about the effect of educational attainment on earnings when quarter of birth is used as an instrument. More generally, our results suggest that the partial R^2 and F statistic on the excluded instruments in the first-stage regression are useful as rough guides to the quality of IV estimates. We suggest that both statistics be routinely reported when IV estimates are presented.

APPENDIX: THE EXACT FINITE-SAMPLE BIAS OF IV

IV 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 (A.1) from the following system (in which, for simplicity, we assume that all random variables have mean 0):

$$\mathbf{y} = \beta \mathbf{x} + \varepsilon, \quad (\text{A.1})$$

$$\mathbf{x} = \mathbf{Z}\boldsymbol{\Pi} + \nu, \quad (\text{A.2})$$

where $\mathbf{x}$, ε , and ν are $N \times 1$ vectors of independent realizations of the random variables x , ε , and ν , respectively, $\mathbf{y}$ is an $N \times 1$ vector, $\mathbf{Z}$ is an $N \times K$ matrix in which the rows are independent realizations of the vector $\mathbf{z}$, composed of random variables $z_1, \dots, z_K$, $\boldsymbol{\Pi}$ is a $K \times 1$ vector of constants, and β is a scalar constant. We assume that $E(\varepsilon|\mathbf{z}) = 0$ and $E(\nu|\mathbf{z}) = 0$. In addition, because $E(\hat{\beta}_{iv})$ does not exist when $K = 1$, we assume that $K > 1$.

Let the reduced form of equation (A.1) be

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\Pi}_0 + \nu_0. \quad (\text{A.3})$$

Define

$$\tau^2 \equiv \frac{\boldsymbol{\Pi}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\Pi}}{\sigma_\nu^2} \quad (\text{A.4})$$

Table A.1. Bias of IV Estimates Relative to OLS Estimates

K	τ^2/K					
	.5	1.0	2.0	4.0	10.0	100.0
2	.61	.37	.14	.02	.00	.00
3	.62	.41	.21	.09	.03	.00
10	.65	.47	.30	.17	.08	.01
20	.66	.49	.32	.19	.08	.01
100	.67	.50	.33	.20	.09	.01
200	.67	.50	.33	.20	.09	.01

NOTE: K is the number of excluded instruments and τ^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 $[1 - (\tau^2/K)_1 F_1(1, (K+2)/2; -\tau^2/2)]$, 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.

and

$$\rho \equiv \frac{\sigma_{v,v_0}}{\sigma_v^2}. \quad (A.5)$$

Sawa showed that under the assumptions that v and v_0 are distributed as jointly normal, the OLS and IV biases can be written as

$$E(\hat{\beta}_{ols}) - \beta = (\beta - \rho) \left[\frac{\tau^2}{N-1} {}_1F_1\left(1, \frac{N+1}{2}; -\frac{\tau^2}{2}\right) - 1 \right], \quad (A.6)$$

$$E(\hat{\beta}_{iv}) - \beta = (\beta - \rho) \left[\frac{\tau^2}{K} {}_1F_1\left(1, \frac{K+2}{2}; -\frac{\tau^2}{2}\right) - 1 \right], \quad (A.7)$$

where ${}_1F_1(\cdot, \cdot; \cdot)$ is the confluent hypergeometric function. Note that $\tau^2/(N-1) \simeq R_{x,z}^2$, the population R^2 from the regression of x on z , and that τ^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 , ${}_1F_1(1, (N+1)/2; -\tau^2/2) \simeq 1$. 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 (A.7).

Although the implication of Equations (A.6) and (A.7) is far from obvious, it is possible to approximate ${}_1F_1(1, \gamma; \zeta)$ for various values of γ and ζ . In Table A.1 we present the magnitude of the bias of IV relative to OLS for various values of τ^2/K and K . Clearly, the bias of IV relative to OLS depends on the τ^2/K , the population analog to the F statistic on the excluded instruments.

[Received June 1993. Revised June 1994.]

REFERENCES

- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (forthcoming), "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, forthcoming.
- Angrist, J. D., and Krueger, A. B. (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.
- Bartels, L. M. (1991), "Instrumental and 'Quasi-Instrumental' Variables," *American Journal of Political Science*, 35, 777-800.
- Basman, 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, 943-976.
- Bowden, R. J., and Turkington, D. A. (1984), *Instrumental Variables*, Cambridge, U.K.: Cambridge University Press.
- Buse, A. (1992), "The Bias of Instrumental Variable Estimators," *Econometrica*, 60, 173-180.
- Carroll, H. C. M. (1992), "Season of Birth and School Attendance," *British Journal of Educational Psychology*, 62, 391-396.
- Durbin, J. (1954), "Errors in Variables," *Review of the International Statistical Institute*, 22, 23-32.
- Gillberg, C. (1990), "Do Children with Autism Have March Birthdays?," *Acta Psychiatrica Scandinavica*, 82, 152-156.
- Griliches, Z. (1977), "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45, 1-22.
- Hare, E. H. (1975), "Manic-Depressive Psychosis and Season of Birth," *Acta Psychiatrica Scandinavica*, 52, 69-79.
- Kestenbaum, B. (1987), "Seasonality of Birth: Two Findings from the Decennial Census," *Social Biology*, 34, 244-248.
- Knoblock, H., and Pasamanick, B. (1958), "Seasonal Variation in the Births of the Mentally Deficient," *American Journal of Public Health*, 48, 1201-1208.
- Lam, D., and Miron, J. A. (1991), "Seasonality of Births in Human Populations," *Social Biology*, 38, 51-78.
- Livingston, R., Adam, B. S., and Bracha, H. S. (1993), "Season of Birth and Neurodevelopmental Disorders: Summer Birth is Associated with Dyslexia," *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, 612-616.
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D., and Ecob, R. (1988), *School Matters: The Junior Years*, Somerset, U.K.: Open Books.
- 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. R., and Startz, R. (1990a), "Some Further Results on the Exact Small-Sample Properties of the Instrumental Variable 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.
- O'Callaghan, E., Gibson, T., Colohan, H. A., Walshe, D., Buckley, P., Larkin, C., and Waddington, J. L. (1991), "Season of Birth in Schizophrenia: Evidence for Confinement of an Excess of Winter Births to Patients Without a Family History of Mental Disorder," *British Journal of Psychiatry*, 158, 764-769.
- Richardson, D. H. (1968), "The Exact Distribution of a Structural Coefficient Estimator," *Journal of the American Statistical Association*, 63, 1214-1226.
- Sawa, T. (1969), "The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators," *Journal of the American Statistical Association*, 64, 923-937.
- Seiver, D. A. (1985), "Trend and Variation in the Seasonality of United States Fertility, 1947-1976," *Demography*, 22, 89-100.
- Sham, P. C., O'Callaghan, E., Takei, N., Murray, G. K., Hare, E. H., and Murray, R. M. (1992), "Schizophrenia Following Prenatal Exposure to Influenza Epidemics Between 1939 and 1960," *British Journal of Psychiatry*, 160, 461-466.
- Shea, J. (1993), "Instrument Relevance in Linear Models: A Simple Measure," unpublished manuscript, University of Wisconsin-Madison.
- Solon, G. (1992), "Intergenerational Income Mobility in the United States," *American Economic Review*, 82, 393-408.
- Staiger, D., and Stock, J. H. (1994), "Instrumental Variables Regression with Weak Instruments," Technical Working Paper 151, National Bureau of Economic Research, Cambridge, MA.
- Tarnowski, K. J., Anderson, D. F., Drabman, R. S., and Kelly, P. A. (1990), "Disproportionate Referrals for Child Academic/Behavior Problems: Replication and Extension," *Journal of Consulting and Clinical Psychology*, 58, 240-243.
- Templer, D. I., Trent, N. H., Spencer, D. A., Trent, A., Corgiat, M. D., Mortensen, P. B., and Gorton, M. (1991), "Season of Birth in Multiple Sclerosis," *Acta Neurologica Scandinavica*, 85, 107-109.
- Wald, A. (1940) "The Fitting of Straight Lines if Both Variables are Subject to Error," *Annals of Mathematical Statistics*, 11, 284-300.
- Watson, C. G., Kucala, T., Tilleskjor, C., and Jacobs, L. (1984), "Schizophrenic Birth Seasonality in Relation to the Incidence of Infectious Diseases and Temperature Extremes," *Archives of General Psychiatry*, 41, 85-90.
- Whorton, J. E., and Karnes, F. A. (1981), "Season of Birth and Intelligence in Samples of Exceptional Children," *Psychological Reports*, 49, 649-650.
- Williams, P., Davies, P., Evans, R., and Ferguson, N. (1970), "Season of Birth and Cognitive Development," *Nature*, 228, 1033-1036.
- Zimmerman, D. J. (1992), "Regression Toward Mediocrity in Economic Stature," *American Economic Review*, 82, 409-429.