 Journal of Business & Economic Statistics, April 1995, Vol. 13, No. 2

Split-Sample Instrumental Variables Estimates of the Return to Schooling

Joshua D. Angrist
Department of Economics, Mount Scopus, Hebrew University, Jerusalem 91905, Israel

Alan B. Krueger
Department of Economics, Princeton University, Princeton, NJ 08544

This article reevaluates recent instrumental variables (IV) estimates of the returns to schooling in light of the fact that two-stage least squares is biased in the same direction as ordinary least squares (OLS) even in very large samples. We propose a split-sample instrumental variables (SSIV) estimator that is not biased toward OLS. SSIV uses one-half of a sample to estimate parameters of the first-stage equation. Estimated first-stage parameters are then used to construct fitted values and second-stage parameter estimates in the other half sample. SSIV is biased toward 0, but this bias can be corrected. The splt-sample estimators confirm and reinforce some previous findings on the returns to schooling but fail to confirm others.

KEY WORDS: Finite-sample bias; Human capital and wages; Two-stage least squares.

There has been longstanding interest in the finite-sample properties of instrumental variables (IV) estimators. In an influential early article, Nagar (1959) used an approximation argument to show that two-stage least squares (2SLS) estimates are biased toward the probability limit of ordinary least squares (OLS) estimates in finite samples with normal disturbances. Buse (1992) generalized this result to cases with nonnormal disturbances. Other things equal, the bias of 2SLS is greater if the excluded instruments explain a smaller share of the variation in the endogenous variable. Nelson and Startz (1990) and Maddala and Jeong (1992) showed that, in samples of the size typically used in time series analyses, IV estimates and their t ratios have highly nonnormal distributions if the first-stage R-square is low and the correlation between reduced-form and structural errors is large.

Recently, Bound, Jaegar, and Baker (in press) (henceforth BJB), Staiger and Stock (1994), and Bekker (1994) argued that finite-sample bias may also be a problem in cross-sectional studies that use very large samples with many excluded instruments. BJB and Staiger and Stock (1994) presented replication studies of Angrist and Krueger (1991) (henceforth AK-91), who reported the results of using quarter of birth to construct instruments for years of schooling in log-wage equations estimated with Census data. Quarter of birth is correlated with schooling because of a mechanical interaction between compulsory attendance laws and age at school entry. Both BJB and Staiger and Stock explored the possibility that the empirical finding that IV estimates of schooling coefficients are similar to OLS estimates is largely attributable to bias in the IV estimates.

In a second application (Angrist and Krueger 1992a; henceforth AK-92), we used draft lottery numbers to construct IV estimates of the returns to schooling for men at risk of being drafted during the Vietnam era. The idea behind AK-92 was to exploit the possibility that draft avoidance via college deferment generated a relationship between randomly assigned draft-lottery numbers and the educational attainment of men at risk of being drafted. As in AK-91, IV estimates of schooling coefficients in AK-92 were also similar to the OLS estimates.

The case for finite-sample bias in these two applications begins by noting that the first-stage equations explain little of the variance of the endogenous regressor. To see how a weak first stage can lead to bias, we experimented with randomly drawn fictitious instruments, which naturally generate a very weak first-stage relationship (similar experiments were reported by BJB). It turns out that by drawing instruments from a uniform random-number generator it is possible to generate 2SLS estimates that are quite close to the OLS estimates arising from specifications reported by AK-92, as well as some of those reported by AK-91. More importantly, the reported second-stage standard errors give the impression of a "statistically significant" structural coefficient estimate using the conventional normal approximation to the sampling distribution of 2SLS estimates.

The possibility of such misleading inferences highlights the importance of developing IV estimators that are not biased toward OLS. In this article, we propose a new estimator that we call split-sample instrumental variables (SSIV). SSIV works by randomly splitting the sample in half and using one half of the sample to estimate parameters of the first-stage equation. These estimated first-stage parameters are then used to construct fitted values and second-stage parameter estimates from data in the other half of the sample. This estimator is a special case of the two-sample instrumental variables (TSIV) estimator presented by Angrist and Krueger (1992b). (Altonji and Segal [1994] also discussed sample splitting to reduce the bias of generalized method of moments estimators.)

Unlike conventional IV estimates, SSIV estimates are biased toward 0 regardless of the degree of covariance between structural and reduced-form errors or the first-stage R^2. An unbiased estimate of the attenuation bias of SSIV is given by the coefficient from a regression of the endogenous regressor on its predicted value (using data from one half of the sample but first-stage parameters from the other). The estimator formed from the product of SSIV and the estimated inverse attenuation bias, called USSIV, is consistent as the number of instruments grows, holding the number of observations per instrument constant. Bekker (1994) showed that this sort of asymptotic argument gives a good account of the finite-sample properties of simultaneous-equations estimators, improving considerably on the conventional asymptotic approximation.

In Section 1, we review the literature on finite-sample bias in IV estimators. In Section 2, we develop the basic SSIV and USSIV approach. In Section 3, we discuss "group asymptotics" applied to SSIV and USSIV, as well as to 2SLS. In Section 4, we present a replication study of our two articles using 2SLS to estimate the returns to schooling. SSIV and USSIV coefficient estimates generated from the AK-91 data are similar to conventional IV estimates. A reexamination of the results of AK-92, however, strongly suggests that the 2SLS results reported in that article are almost solely attributable to the finite-sample bias of IV toward OLS.

1. SMALL–SAMPLE BIAS IN 2SLS ESTIMATES

Consider the following two-equation model with one endogenous regressor:

$$y_i = \beta_0' w_{0i} + \beta_1 s_i + \epsilon_i \equiv \beta' x_i + \epsilon_i \qquad (1)$$

and

$$x_i = \pi_0' w_{0i} + \pi_1' w_{1i} + \eta_i \equiv \pi' z_i + \eta_i, \qquad (2)$$

for $i = 1, \ldots, n$ observations, where y_i is the dependent variable (e.g., log wages) and s_i is the endogenous regressor (e.g., years of schooling). z_i is a $(k+p) \times 1$ vector of instrumental variables that includes the p exogenous variables appearing in Equation (1), w_{0i}, plus k additional variables, w_{1i} (e.g., quarter-of-birth dummies). Thus there are k excluded instruments and $k-1$ overidentifying restrictions. x_i is a $(p+1) \times 1$ vector that includes the exogenous regressors along with the endogenous regressor.

The data are more compactly denoted by an $n \times 1$ vector Y, an $n \times (p+1)$ matrix X, and an $n \times (k+p)$ matrix Z. From (1) and (2), we have $Y = X\beta + \epsilon$ and $X = Z\pi + \eta$. The coefficient β_1 is the scalar parameter of interest, assumed to be the last element in the $(p+1) \times 1$ vector β, and π is the $(k+p) \times (p+1)$ matrix of reduced-form parameters. We assume that observations in the sample are iid and that the disturbances satisfy $E(\epsilon_i \mid z_i) = E(\eta_i \mid z_i) = 0$. The residual variance of ϵ_i is denoted σ_ϵ^2. The vector of residual variances in Equation (2), η_i, consists of p zeros for the exogenous convariates, plus the last element corresponding to s_i. The variance of this element is σ_η^2, and its covariance with ϵ_i is $\sigma_{\epsilon\eta}$.

BJB's adaptation of the Buse (1992) approximate bias formula for a simple case with no exogenous regressors (where all variables have mean 0) is

$$\frac{\sigma_{\epsilon\eta}}{\sigma_\eta^2} \times \frac{\sigma_\eta^2}{\pi' Z' Z \pi}(k-2). \qquad (3)$$

BJB pointed out that $\pi' Z' Z\pi/\sigma_\eta^2 k$ is the inverse of the population analog of the F statistic for a test of $\pi = 0$ in the first-stage equation (i.e., substituting π and σ_η^2 for OLS estimates in the usual F statistic formula) and that the approximate bias of IV estimates is proportional to the OLS bias, $\sigma_{\epsilon\eta}/\sigma_\eta^2$. It is also clear that a lower first-stage R^2, keeping constant the number of instruments, leads to more bias unless there is no need to instrument (i.e., $\sigma_{\epsilon\eta} = 0$.)

A simple explanation for this sort of bias in IV estimates is that the *estimated* coefficients used to construct the first-stage fitted values are correlated with the structural-equation error. Let $P_z = Z(Z'Z)^{-1}Z'$. The first-stage fitted values can then be written $P_z X = Z\pi + P_z\epsilon$. The average covariance between $P_z\epsilon$ and the last column of η is asymptotically negligible but has an expectation equal to $\sigma_{\epsilon\eta}[k+p]/n$ in any finite sample.

To see how serious this sort of bias could be, we experimented with the specification reported by AK-91 using 3 quarter-of-birth dummies × 10 year-of-birth dummies plus 3 quarter-of-birth dummies × 50 state-of-birth dummies to form a set of 180 excluded instruments. Conventional IV estimates from this specification (using a sample of over 329,000 observations) generate a schooling coefficient of .093 with standard error of .009. OLS estimates of the same specification generate a schooling coefficient of .067 with a standard error of .0003. Replacing actual quarter of birth with a random draw from a four-point discrete uniform distribution and repeating the IV estimation generated a coefficient of .057 with a reported standard error of .014. Most researchers would probably believe (not knowing that the instruments were fictitious) that they had learned something about the returns to schooling from this estimate.

2. SPLIT-SAMPLE INSTRUMENTAL VARIABLES (SSIV)

SSIV solves the problem of spurious inferences in IV estimation by breaking the link between ϵ and η in Equations (1) and (2). The SSIV estimate is constructed by randomly dividing a single sample into two half samples, denoted 1 and 2. Each sample consists of data matrices $\{Y_j, X_j, Z_j\}$ for $j = 1, 2$. Sample 2 is used to estimate the first-stage equation. The first-stage parameters are then combined with observations on Z_1 to form fitted values for X_1 in sample 1. Finally, Y_1 is regressed on these fitted values and the exogenous regressors in sample 1. The estimator is

$$\begin{aligned}\widehat{\beta}_s &= \left(\widehat{X}_{21}'\widehat{X}_{21}\right)^{-1}\widehat{X}_{21}'Y_1 \\ &= \left[X_2'Z_2(Z_2'Z_2)^{-1}Z_1'Z_1(Z_2'Z_2)^{-1}Z_2'X_2\right]^{-1} \\ &\quad \times \left[X_2'Z_2(Z_2'Z_2)^{-1}Z_1'Y_1\right], \end{aligned} \qquad (4)$$

where $\widehat{X}_{21} = Z_1(Z_2'Z_2)^{-1}Z_2'X_2$ is the cross-sample fitted value. Note that the cross-sample fitted value for exogenous regressors is the value of the exogenous regressors in sample 1.

To develop the properties of SSIV, we begin with the observation that by virtue of independent sampling we have

Assumption 1. The data matrices $\{Y_1, X_1, Z_1\}$ and $\{Y_2, X_2, Z_2\}$ are jointly independent.

Assumption 1 implies that $\{Y_1, X_1\}$ is jointly independent of $\{Y_2, X_2, Z_2\}$ given Z_1. This is used to prove the following proposition.

Proposition 1. (a) Provided that the expectation exists, $E(\widehat{\beta}_s) = E(\widehat{\theta})\beta = \theta\beta$, where $\widehat{\theta}$ is a $(p+1)\times(p+1)$ matrix,

$$\widehat{\theta} = \left[X_2'Z_2(Z_2'Z_2)^{-1}Z_1'Z_1(Z_2'Z_2)^{-1}Z_2'X_2\right]^{-1} \times\left[X_2'Z_2(Z_2'Z_2)^{-1}Z_1'X_1\right]. \quad (5)$$

(b) Let $\widehat{s}_{21}$ represent the cross-sample fitted value of the vector of s_i, and let s_1 represent the endogenous regressor in sample 1. The lower right element of θ is equivalent to the coefficient on $\widehat{s}_{21}$ from a regression of s_1 on $\widehat{s}_{21}$ and all the exogenous regressors. This regression coefficient provides an unbiased estimate of the proportional bias in SSIV estimates of β_1.

Proof. Substitute $X_1\beta + \epsilon_1$ for Y_1 in (4). Then we can write

$$\widehat{\beta}_s = \widehat{\theta}\beta + \left[X_2'Z_2(Z_2'Z_2)^{-1}Z_1'Z_1(Z_2'Z_2)^{-1}Z_2'X_2\right]^{-1} \times [X_2'Z_2(Z_2'Z_2)^{-1}Z_1'\epsilon_1].$$

Iterating expectations over Z_1 and using assumption 1, we have

$$\begin{aligned} &E\left[X_2'Z_2(Z_2'Z_2)^{-1}Z_1'Z_1(Z_2'Z_2)^{-1}Z_2'X_2\right]^{-1} \times \left[X_2'Z_2(Z_2'Z_2)^{-1}Z_1'\epsilon_1\right] \\ &\quad = E\{E[X_2'Z_2(Z_2'Z_2)^{-1}Z_1'Z_1(Z_2'Z_2)^{-1}Z_2'X_2]^{-1} \\ &\qquad \times \left[X_2'Z_2(Z_2'Z_2)^{-1} \mid Z_1\right] \times Z_1' \times E[\epsilon_1 \mid Z_1]\}, \end{aligned}$$

which is 0 because $E[\epsilon_1 \mid Z_1] = 0$.

Note that $\widehat{\theta}$ is the matrix of coefficients from a regression of the columns of X_1 on $Z_1(Z_2'Z_2)^{-1}Z_2'X_2$. We can write $X_1 = [W_{01} \quad s_1]$, and $Z_1(Z_2'Z_2)^{-1}Z_2'X_2 = [W_{01} \quad \widehat{s}_{21}]$. Regressing $[W_{01} \quad s_1]$ on $[W_{01} \quad \widehat{s}_{21}]$ gives the matrix

$$\widehat{\theta} = \begin{bmatrix} I_p & \widehat{\theta}_p \\ 0 & \widehat{\theta}_{p+1} \end{bmatrix},$$

where I_p is a $p \times p$ identity matrix, $\widehat{\theta}_p$ is $p \times 1$, and $\widehat{\theta}_{p+1}$ is a scalar equal to the coefficient on $\widehat{s}_{21}$ in a regression of s_1 on $\widehat{s}_{21}$ and W_{01}.

Proposition 1 starts with the assumption that the expectation of $\widehat{\theta}$ exists. We have not been able to provide general conditions for the existence of this expectation. Instead, we first consider a special case in which the expectation clearly exists. Then in Section 4 we use an improved asymptotic argument to make the same point in a more general way. The special case we consider is one in which $E[X_1 \mid \widehat{X}_{21}]$ is linear. [This will be approximately true if X and Z are normally distributed and if the sampling variance of $(Z_2'Z_2)^{-1}Z_2'X_2$ is negligible.] We have the following corollary.

Corollary 1.1. Suppose that $E[X_1 \mid \widehat{X}_{21}]$ is linear. Then

$$\theta \equiv E[\widehat{\theta}] = (E[\widehat{X}_{21}'\widehat{X}_{21}])^{-1}E[\widehat{X}_{21}'X_1] \quad (5a)$$

$$= \left\{\pi'E(z_iz_i')\pi + c\sigma_\eta^2 L_1\right\}^{-1}\{\pi'E(z_iz_i')\pi\}, \quad (5b)$$

where $c \equiv \mathrm{tr}\{E[(Z_2'Z_2)^{-1}(Z_1'Z_1)]/n_1\}$ and L_1 is a $(k+p)$ square matrix consisting of all zeros except for a 1 in the lower right corner. If $(Z'Z)$ is the same in the two samples, then $c = (k+p)/n_1$. L_1 reflects the fact that X_i includes only one endogenous regressor, s_i.

Proof. If $E[X_1 \mid \widehat{X}_{21}]$ is linear, then $E[X_1 \mid \widehat{X}_{21}]$ is $\widehat{X}_{21}\{E[\widehat{X}_{21}'\widehat{X}_{21}]^{-1}E[\widehat{X}_{21}'X_1]\}$. Since $\widehat{\theta} = [\widehat{X}_{21}'\widehat{X}_{21}]^{-1}[\widehat{X}_{21}'X_1]$, we can substitute for X_1 to show that $E[\widehat{\theta}] = E[\widehat{X}_{21}'\widehat{X}_{21}/n_1]^{-1} * E[\widehat{X}_{21}'X_1/n_1]$. In the Appendix, the moments in the numerator and denominator are simplified to give (5b).

The corollary shows that $\widehat{\theta}$ represents a kind of attenuation bias arising from the use of reduced-form coefficients from a separate sample. If there are no exogenous variables, then the proportional bias of SSIV is between 0 and 1—that is, the SSIV coefficient will be biased toward 0 in absolute value. More generally, (5b) implies a matrix attenuation bias. As in multivariate measurement-error models with a single mismeasured regressor (e.g., Fuller 1975), matrix attenuation in this case implies attenuation of the coefficient on the single endogenous regressor, β_1. The Appendix shows that

$$\theta\beta = \begin{bmatrix} \beta_0 + \left[c\sigma_\eta^2/(\phi + c\sigma_\eta^2)\right]P^{-1}R\beta_1 \\ \left[\phi/(\phi + c\sigma_\eta^2)\right]\beta_1 \end{bmatrix}, \quad (6)$$

where P and R are submatrices in a partitioned matrix and ϕ is a positive scalar so that $\phi/(\phi + c\sigma_\eta^2)$ is necessarily between 0 and 1 (because c is also positive.)

A consequence of Equation (6) is that, under the conditions of Corollary 1.1, the SSIV estimate is asymptotically unbiased as n gets large with the number of instruments fixed (because c then goes to 0.) Another case of interest is when the vector of reduced-form coefficients, π, is near 0. In this case, it is apparent from Equation (5) that the SSIV estimate of β_1 has expectation near 0. Moreover, increasing the number of instruments with the explained sum of squares fixed also tends to pull SSIV estimates toward 0 (because c then increases.) This property contrasts sharply with the tendency of conventional IV estimates to be biased toward OLS.

2.1 Conventional Asymptotic Results for SSIV

SSIV is consistent because plim $\widehat{\theta}$ equals the identity matrix. The following proposition characterizes the asymptotic distribution of $\widehat{\beta}_s$.

Proposition 2. Define $g_n(\beta) \equiv [Z_1'Y_1/n_1 - (Z_2'X_2/n_2)\beta]$, where $n_1 = \alpha n_2$ for some positive number α. Under standard conditions, $n_1^{1/2}g_n(\beta) \overset{a}{\sim} N(0, \Omega)$, where Ω is a $(p+k) \times (p+k)$ asymptotic covariance matrix. Then $n_1^{1/2}(\widehat{\beta}_s - \beta) \overset{a}{\sim} N(0, \psi)$ where $\psi = (\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{xz})^{-1}\ \Sigma_{xz}\Sigma_{zz}^{-1}\Omega\Sigma_{zz}^{-1}\Sigma_{xz}\ (\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{xz})^{-1}$ and Σ_{xz} and Σ_{zz} denote population average cross-product matrices.

Proof. This proposition is established by showing that SSIV is asymptotically equivalent to the general TSIV estimator discussed by Angrist and Krueger (1992b) and then using the asymptotic covariance matrix given by them. The general TSIV estimator is $[(X_2'Z_2/n_2)\Phi(Z_2'X_2/n_2)]^{-1}[(X_2'Z_2/n_2)\Phi(Z_1'Y_1/n_1)]$, where Φ is any positive-definite weighting matrix. To see that SSIV is asymptotically equivalent to TSIV, set $\Phi = (Z_2'Z_2)^{-1}(Z_1'Z_1)(Z_2'Z_2)^{-1}$.

In general, setting $\Phi = \Omega^{-1}$ gives the optimal TSIV estimator. If $\Omega = \Sigma_{zz}\sigma_\epsilon^2$ (say because $\beta = 0$), then SSIV is the asymptotically efficient TSIV estimator constructed from $Z_1'Y_1/n_1$ and $Z_2'X_2/n_2$. In this case, the asymptotic covariance matrix of SSIV simplifies to $(\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{xz})^{-1}\sigma_\epsilon^2$ under an $(n_1)^{1/2}$ normalization, which is the usual form of the 2SLS asymptotic covariance matrix. Since $n_1 = n/2$ in the SSIV case, however, the asymptotic covariance matrix of SSIV is at least twice the asymptotic covariance matrix of 2SLS. This is not surprising because SSIV uses half as much data as 2SLS to compute the second-moment matrices.

Finally, note that SSIV has a practical advantage over other TSIV estimators in that it is easy to compute using standard OLS regression software. Moreover, like TSIV, the SSIV estimate can be calculated using one sample with information on Z and Y but no information on X and a second sample with information on Z and X but no information on Y. This property sometimes motivates the use of TSIV instead of 2SLS (e.g. Angrist 1990; Angrist and Krueger 1992b.).

2.2 Unbiased Split-Sample Estimation

It seems reasonable to try to improve on SSIV by inflating $\widehat{\beta}_s$ by the inverse of the estimated proportional bias, $\widehat{\theta}$. The resulting estimator, $\widehat{\theta}^{-1}\widehat{\beta}_s$, is not unbiased, however, because it involves a nonlinear function of the (correlated) random variables $\widehat{\theta}$ and $\widehat{\beta}_s$. Nevertheless, the inflated estimator is unbiased under the group-asymptotic argument outlined later. We therefore label the inflated estimator unbiased split-sample instrumental variables (USSIV).

Recall that $\widehat{\theta} = [\widehat{X}_{21}'\widehat{X}_{21}]^{-1}[\widehat{X}_{21}'X_1]$. Then the USSIV estimator is $\widehat{\beta}_u \equiv \widehat{\theta}^{-1}\widehat{\beta}_s = [\widehat{X}_{21}'X_1]^{-1}[\widehat{X}_{21}'Y_1]$. Note that $\widehat{\beta}_u$ can be constructed by using $\widehat{X}_{21}$ as an instrument for X_1 in the regression $Y_1 = X_1\beta + \epsilon_1$. Using $\widehat{X}_{21}$ as an instrument for X_1 instead of including it directly as a regressor eliminates the attenuation bias that arises from estimation of the first-stage reduced form. An important difference between USSIV and SSIV is that USSIV requires data on X_1 but SSIV does not. This means that USSIV *cannot* be used in applications such as that of Angrist and Krueger (1992b), where one sample includes only observations on (Z_1, Y_1) and the other includes only observations on (Z_2, X_2).

3. GROUP ASYMPTOTICS

In this section, we develop an asymptotic argument that appears to capture important features of the finite-sample behavior of $\widehat{\beta}_s$ and $\widehat{\beta}_u$. The group-asymptotics approach derives the limiting characteristics of $\widehat{\beta}_s$ as the number of instruments grows, but the number of observations per instrument is held fixed. In the context of AK-91, this can be thought of as obtaining additional instruments by adding new cross-sections for new years of data, or by adding additional cross-sections from new states, regions, or cohorts. This is the same type of argument used by Deaton (1985) in his study of panel data created from an asymptotically lengthening time series of cross-sections. The group-asymptotics approach is also similar to the parameter sequence used in Bekker's (1994) study of simultaneous-equations estimators. As noted by Bekker, the rationale for this approach is not really important; what matters is that it gives a good account of finite-sample properties.

Under group asymptotics, each cross-section replication provides m additional observations. In particular, the tth cross-section replication is assumed to contain iid data matrices of length m with observations on $\{Y_t, X_t, Z_t\}$ for $t = 1, \ldots, T$. We split these observations into data matrices for half-samples of size m_1 and m_2, denoted by $\{Y_{jt}, X_{jt}, Z_{jt}\}$, for $j = 1, 2$. An important feature of this replication sequence is that there is assumed to be a different matrix of reduced-form coefficients associated with each replication. In particular, we imagine that at each replication a reduced-form coefficient matrix, π_t, is also drawn. The π_t are themselves iid random matrices satisfying $E[X_{jt} - Z_{jt}\pi_t \mid Z_{jt}] = 0$, with $E[(X_{jt} - Z_{jt}\pi_t)(X_{jt} - Z_{jt}\pi_t)' \mid Z_{jt}]$ having one nonzero element equal to σ_η^2. Each π_t is independent of the data in each half sample.

The fact that π_t varies with t motivates the use of interaction terms in the instrument list. The matrix of fitted values is therefore

$$\widehat{X}_{21} = \left[\widehat{X}_{21,1}' \cdots \widehat{X}_{21,t}' \cdots \widehat{X}_{21,T}'\right]',$$

where $\widehat{X}_{21,t} = Z_{1t}(Z_{2t}'Z_{2t})^{-1}Z_{2t}'X_{2t}$. Similarly, the data matrices from each replication are stacked:

$$Y_j = [Y_{j1}', \ldots, Y_{jt}', \ldots, Y_{jT}']'$$
$$X_j = [X_{j1}', \ldots, X_{jt}', \ldots, X_{jT}']'$$
$$Z_j = [Z_{j1}', \ldots, Z_{jt}', \ldots, Z_{jT}']'$$

for $j = 1, 2$.

Consider the SSIV estimator constructed by pooling all replications and allowing a separate reduced form for each replication. We define the *group-asymptotic probability limit* of $\widehat{\beta}_s$ as the probability limit of this estimator when the number of groups (T) becomes infinite while the group size (m) is fixed. This probability limit turns out to be similar to the expectation derived in Proposition 1, in which a linear conditional expectation, $E[X_1 \mid \widehat{X}_{21}]$, was assumed. Proposition 3 uses group asymptotics to compare the bias of SSIV and conventional 2SLS.

Proposition 3. (a) For SSIV, it is useful to write

$$\widehat{X}_{21,t} = Z_{1t}\pi_t + Z_{1t}(Z_{2t}'Z_{2t})^{-1}Z_{2t}'\eta_{2t}$$
$$Y_{1t} = (Z_{1t}\pi_t + \eta_{1t})\beta + \epsilon_{1t}.$$

The group-asymptotic probability limit of SSIV is

$$\operatorname*{plim}_{T\to\infty}[(1/T)\Sigma_t\widehat{X}_{21,t}'\widehat{X}_{21,t}/m_1]^{-1}[(1/T)\Sigma_t\widehat{X}_{21,t}'Y_{1t}/m_1]$$
$$= \{E[\pi_t'(Z_{1t}'Z_{1t}/m_1)\pi_t] + c^*\sigma_\eta^2 L_1\}^{-1} \times E[\pi_t'(Z_{1t}'Z_{1t}/m_1)\pi_t]\beta,$$

where $c^* \equiv \text{tr}\{E[(Z_{2t}'Z_{2t})^{-1}(Z_{1t}'Z_{1t})]/m_1\}$. (b) For 2SLS, it is useful to write

$$\widehat{X}_t = Z_t(Z_t'Z_t)^{-1}Z_t'X_t = Z_t\pi_t + Z_t(Z_t'Z_t)^{-1}Z_t'\eta_t$$
$$Y_t = (Z_t\pi_t + \eta_t)\beta + \epsilon_t. \qquad (7)$$

The group-asymptotic probability limit of 2SLS is

$$\begin{aligned}\text{plim}[(1/T)\Sigma_t[\widehat{X}_t'\widehat{X}_t/m]^{-1}[(1/T)\Sigma_t[\widehat{X}_t'Y_t/m] \\ = \beta + [(k+p)/m]\{E[\pi_t'(Z_t'Z_t/m)\pi_t] \\ + [(k+p)/m]\sigma_\eta^2 L_1\}^{-1}\ell_1\sigma_{\epsilon\eta},\end{aligned}$$

where ℓ_1 is a $(p+1)$ column vector consisting of zeros in the first p rows and a 1 in the last row.

Proof. For Part (a), the first step is to note that

$$\begin{aligned}\underset{T\to\infty}{\text{plim}}[(1/T)\Sigma\widehat{X}_{21,t}'\widehat{X}_{21,t}'/m_1]^{-1}[(1/T)\Sigma\widehat{X}_{21,t}'Y_{1t}/m_1] \\ = E[\widehat{X}_{21,t}'\widehat{X}_{21,t}/m_1]^{-1} * E[\widehat{X}_{21,t}'Y_{1t}/m_1].\end{aligned}$$

The proof is completed by using the definitions of $\widehat{X}_{21,t}$ and Y_{1t} and the independence assumption to evaluate population moments, as in the proof of Corollary 1.1 (given in the Appendix). As in Corollary 1.1, the matrix attenuation in Proposition 3 implies scalar attenuation of the coefficient β_1. Similarly, proof of Part (b) begins with the observation that the desired probability limit is $E[\widehat{X}_t'\widehat{X}_t/m]^{-1} * E[\widehat{X}_t'Y_t/m]$. In this case, however, ϵ_t and η_t are in the same sample and have covariance $\sigma_{\epsilon\eta}$ for the element of η_i corresponding to s_i.

This proposition shows that neither SSIV or 2SLS is consistent under group asymptotics, although the two estimators are biased in different ways. The group-asymptotic probability limit of SSIV is the same as that presented in Corollary 1.1 for a special case and reflects a bias toward 0. The group-asymptotic probability limit of 2SLS is the same as Bekker's (1994) formula for the bias of 2SLS. Like the Nagar (1959) and Buse (1992) approximation results, this formula reflects a bias toward OLS. (Because SSIV and 2SLS are not consistent under group asymptotics, the standard errors for the SSIV and 2SLS estimates reported in Section 4 are based on the usual asymptotic approximation.)

The group-asymptotic properties of USSIV are summarized in the following proposition.

Proposition 4.

(a) $\underset{T\to\infty}{\text{plim}}[(1/T)\Sigma_t\widehat{X}_{21,t}'X_{1t}/m_1]^{-1}[(1/T)\Sigma_t\widehat{X}_{21,t}'Y_{1t}/m_1] = \beta$.

(b) $T^{1/2}(\widehat{\beta}_u - \beta) \overset{a}{\sim} N(0, \Lambda)$

where

$$\begin{aligned}\Lambda = (1/m_1)E[X_{1t}'\widehat{X}_{21,t}'/m_1]^{-1} \\ \times E[\widehat{X}_{21,t}'\widehat{X}_{21,t}/m_1]E[\widehat{X}_{21,t}'X_{1t}/m_1]^{-1}\sigma_\epsilon^2.\end{aligned}$$

Proof. To prove (a), we only need to show that

$$\underset{T\to\infty}{\text{plim}}[(1/T)\Sigma_t\widehat{X}_{21,t}'X_{1t}/m_1] = E[\pi_t'Z_{1t}'Z_{1t}\pi_t/m_1].$$

Writing $X_{1t} = Z_{1t}\pi_t + \eta_{1t}$ and using the definition of $\widehat{X}_{21,t}$ gives $E[\widehat{X}_{21,t}'X_{1t}/m_1] = E[\pi_t'Z_{1t}'Z_{1t}\pi_t/m_1]$, as in the Appendix proof of Corollary 1.1. To derive the variance formula in (b), substitute for Y_1 in $\widehat{\beta}_u$:

$$\widehat{\beta}_u = [X_{21}'X_1]^{-1}[\widehat{X}_{21}'Y_1] = \beta + [\widehat{X}_{21}'X_1]^{-1}[\widehat{X}_{21}'\epsilon_1]$$

so that

$$\begin{aligned}T^{1/2}(\widehat{\beta}_u - \beta) = [(1/T)\Sigma_t\widehat{X}_{21,t}'\widehat{X}_{1t}/m_1]^{-1} \\ * T^{1/2}[(1/T)\Sigma_t\widehat{X}_{21,t}'\epsilon_{1t}/m_1] \\ \overset{a}{\sim} E[\widehat{X}_{21,t}'X_{1t}/m_1]^{-1} \\ * T^{1/2}[(1/T)\Sigma_t\widehat{X}_{21,t}'\epsilon_{1t}/m_1].\end{aligned}$$

Using the fact that ϵ_{1t} is mean independent of $\widehat{X}_{21,t}$ with a scalar covariance matrix completes the proof.

Thus USSIV is consistent under group asymptotics. The USSIV coefficient estimates and group-asymptotic standard errors are also easy to compute. Note that $\widehat{\beta}_u$ is a just-identified 2SLS estimator in sample 1, so it can be written

$$\widehat{\beta}_u = [X_1'\widehat{X}_{21}(\widehat{X}_{21}'\widehat{X}_{21})^{-1}\widehat{X}_{21}'X_1]^{-1}X_1'\widehat{X}_{21}'(\widehat{X}_{21}'\widehat{X}_{21})^{-1}\widehat{X}_{21}'Y_1.$$

The conventional 2SLS covariance matrix estimator for an estimate of this form is $[X_1'\widehat{X}_{21}'(\widehat{X}_{21}\widehat{X}_{21})^{-1}\widehat{X}_{21}'X_1]^{-1}\sigma_\epsilon^2$. Software-reported 2SLS standard errors therefore provide a consistent estimate of the sampling variance of $\widehat{\beta}_u$ under group asymptotics.

We also have the following corollary, which gives conventional asymptotic results for USSIV for the case in which $m_1 = m/2$.

Corollary 4.1. $([m/2]T)^{1/2}(\widehat{\beta}_u - \beta)$ has the same conventional asymptotic covariance matrix (i.e., letting m get large with fixed T) as 2SLS.

This can be proved directly or by taking the limit of Λ as m becomes infinite. An implication of the corollary is that the average of USSIV and its complement (reversing the roles of samples 1 and 2) has the same limiting distribution under conventional asymptotics as does 2SLS. Analysis of the group-asymptotic distribution of the combined estimator is more complicated, however, because the group-asymptotic covariance of the two possible USSIV estimators is not 0. We therefore leave an investigation of the combined USSIV estimator for future work.

4. IV ESTIMATES OF THE RETURN TO SCHOOLING

4.1 Angrist and Krueger (1991)

AK-91 argued that quarter of birth provides a legitimate instrumental variable for years of schooling because children born earlier in the year enter school at an older age and are therefore allowed to drop out (on their 16th or 17th birthday) after having completed less schooling than children born later in the year. In particular, men born in earlier quarters get less schooling, are less likely to graduate from high school, and earn less than men born in later quarters. These relationships are statistically significant in data on single-year birth cohorts from 1920–1959 and in both the 1970 and 1980 Census.

2SLS estimates of the return to education based on quarter-of-birth instruments are close to OLS estimates, suggesting that omitted variables do not bias the OLS estimates. Here we focus on estimates for men in their 40s (i.e., men born 1930–1939 in the 1980 Census and men born 1920–1929 in the 1970 Census) because the age–earnings profile is fairly flat for this age group. This minimizes potential problems due to correlation between age and quarter of birth.

The first two columns of Table 1 report OLS and 2SLS estimates of the education coefficient from log-wage equations estimated using the Census samples. The sample sizes range from close to 250,000 in the 1970 Census to close to 330,000 in the 1980 Census, and the specifications are the same as those reported by AK-91 (tables IV and V). The 2SLS model uses 30 quarter-of-birth × year-of-birth interaction terms as excluded instruments, including year-of-birth main effects as exogenous covariates.

As noted previously, OLS and 2SLS estimates are remarkably similar in both data sets. This naturally raises the question of whether such similarity is a real finding or a spurious result attributable to the finite-sample bias of IV. For comparison, SSIV and USSIV results are presented in columns 3 and 4. Each of the SSIV estimates is somewhat smaller than the corresponding IV estimate, as one would expect because SSIV is biased toward 0. The SSIV estimate is above the OLS estimate for the 1980 sample, whereas it is below it for the 1970 sample. But in each case the SSIV and OLS estimates are not statistically different. The SSIV estimates are significantly different from 0, with standard errors 50% larger than the standard errors of 2SLS estimates.

The proportional attenuation bias (θ) of SSIV is estimated to be 78% in the 1980 sample and 93% in the 1970 sample, with a standard error of about 12% in each case. Column (4) reports USSIV estimates, which inflate the SSIV estimates by the inverse of $\widehat{\theta}$. The USSIV estimates tend to be above the OLS estimates and are also remarkably close to 2SLS estimates.

Table 1. Quarter-of-Birth Estimates With 30 Instruments

Parameter	OLS (1)	2SLS (2)	SSIV (3)	USSIV (4)
	A. 1980 Census, men born 1930–1939			
β	.063	.081	.070	.089
	(.0003)	(.016)	(.023)	(.030)
θ	—	—	.780	—
			(.118)	
First-stage F (df = 30)	—	4.75	2.41	2.41
	B. 1970 Census, men born 1920–1929			
β	.070	.069	.059	.063
	(.0004)	(.015)	(.023)	(.024)
θ	—	—	.934	—
			(.127)	
First-stage F (df = 30)	—	4.54	2.03	2.03

NOTE: Models include 9 year-of-birth dummies, marital status, region dummies, SMSA dummy, and a race dummy as exogenous regressors. Sample size for 1980 sample for OLS and 2SLS is 329,509; for SSIV and USSIV, the first-stage equation was estimated with 164,474 observations and the second-stage with 165,035 observations. Sample size for the 1970 sample for OLS and 2SLS is 244,099; for SSIV and USSIV the first-stage equation was estimated with 121,956 observations and the second-stage with 122,143 observations.

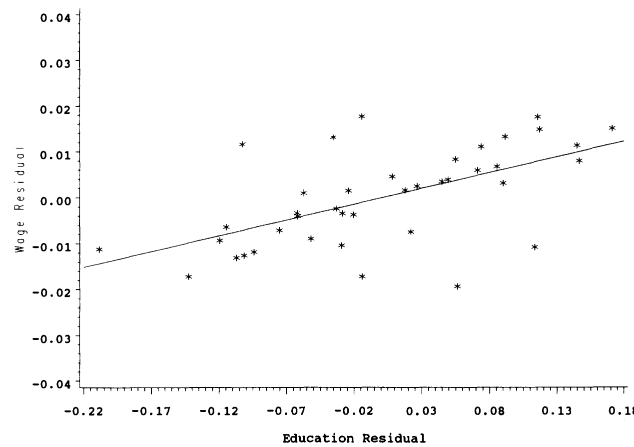

Figure 1. Split-sample Graph of Average Wages by Quarter of Birth Against Average Schooling by Quarter of Birth in the 1970 Census. The scatter shows averages computed from two half samples, one for earnings and one for schooling, both drawn from the 1970 Census for men born 1920–1929. Points plotted in the figure are residuals from a regression on year-of-birth effects. The OLS regression line through the average is also shown.

The split-sample approach can also be used to produce a graphical impression of the SSIV slope estimate. To do this, we randomly split the sample in half and then graphed average earnings by quarter of birth in one sample against average education by quarter of birth in the other sample (after removing year effects). Figures 1 and 2 show these graphs for the 1970 and 1980 samples. The plots clearly show upward-sloping relationships. The slope of the regression line drawn in the figures can be shown to be an SSIV estimator for this example because $Z_1'Z_1$ is roughly proportional to I_{k+p} in this case. For both the 1970 and 1980 data, the slope is roughly .069.

Table 2 reports a set of OLS, 2SLS, SSIV, and USSIV results for models estimated using 150 quarter-of-birth × state-of-birth interactions plus 30 quarter-of-birth × year-of-birth interactions as the excluded instruments, with data from the

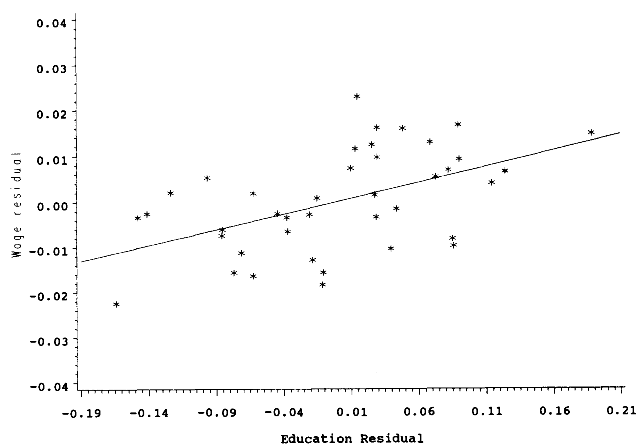

Figure 2. Split-sample Graph of Average Wages by Quarter of Birth Against Average Schooling by Quarter of Birth in the 1980 Census. The scatter shows averages computed from two half samples, one for earnings and one for schooling, both drawn from the 1980 Census for men born 1930–1939. Points plotted in the figure are residuals from a regression on year-of-birth effects. The OLS regression line through the averages is also shown.

Table 2. Quarter-of-Birth Estimates With 180 Instruments

Parameter	OLS (1)	2SLS (2)	SSIV (3)	USSIV (4)
	Type of estimate			
	1980 Census, men born 1930–1939			
β	.063	.083	.031	.076
	(.0003)	(.009)	(.011)	(.028)
θ	—	—	.408	—
			(.057)	
First-stage F (df = 180)	—	2.43	1.70	1.70

NOTE: Models include 9 year-of-birth dummies, 48 state-of-birth dummies, marital status, region dummies, SMSA dummy, and a race dummy as exogenous regressors. Sample size for 1980 sample for OLS and 2SLS is 329,509; for SSIV and USSIV the first-stage equation was estimated with 164,474 observations and the second-stage with 165,035 observations.

Table 3. Quarter-of-Birth Estimates—Results of 31 Monte Carlo Replications of Split (180-instrument specification)

	Actual instruments		Random instruments	
Statistic	SSIV (1)	USSIV (2)	SSIV (3)	USSIV (4)
	Summary statistics for education coefficients			
Mean	.048	.112	.002	.021
Median	.050	.114	.004	.034
Standard deviation of coefficients	.010	.024	.014	.187
25th percentile	.042	.099	−.006	−.080
75th percentile	.055	.129	.014	.133

NOTE: Models include 9 year-of-birth dummies and 50 state-of-birth dummies as exogenous regressors. The conventional 2SLS estimate and standard error using random instruments is .057 (.014).

1980 Census sample. This model has a first-stage F statistic for the excluded instruments of 2.4 (compared to 4.8 in the 30-instrument model) and corresponds to the models reported in table VII of AK-91. BJB and Staiger and Stock (1994) argued that the low first-stage F statistic means that IV estimates of these models are likely to be seriously biased. (On the other hand, Hall, Rudebusch, and Wilcox [1994] noted that pretesting for instrument relevance using first-stage F tests or other criteria can exacerbate the poor finite-sample properties of 2SLS.) For the 180-excluded-instruments specification, the SSIV education coefficient is .031, about 40% as large as the IV estimate, though still significantly different from 0. The estimated proportional attenuation bias of SSIV, however, is also on the order of 40%. Consequently, the USSIV estimate is .076, only slightly less than the 2SLS estimate and above the OLS estimate.

Would SSIV and USSIV provide misleading results in the extreme case of fictitious, randomly assigned instruments? To investigate this, as well as the sensitivity of SSIV and USSIV estimates to alternative splits, we conducted a small-scale Monte Carlo exercise in which we randomly divided the sample and calculated SSIV and USSIV estimates 31 times. For each replication, we divided the data using a different (randomly generated) seed number. The specifications estimated here use the 180 quarter-of-birth interactions as excluded instruments, as in Table 2.

The Monte Carlo results for the actual instruments are reported in columns (1) and (2) of Table 3. The average SSIV estimate is .048, with a Monte Carlo standard deviation of .010 in 31 replications. This is somewhat higher than the SSIV estimate reported in Table 2 for a similar specification. (We omitted region dummies, marital status, the SMSA dummy, and the race dummy from the first and second stages of the models used for the Monte Carlo replications. The estimates in Table 3 and Table 2 are therefore not strictly comparable.) The median SSIV estimate is .05, with upper and lower quartiles of .055 and .042. The average estimate of the proportional attenuation bias in SSIV in these 31 replications (not shown in the table) is .433 with a Monte Carlo standard deviation of .05. The average USSIV estimate is .112 with a standard deviation of .024. Lower and upper quartiles for USSIV estimates are .099 and .129, giving an interquartile range of .03. This suggests that the SSIV estimates are less sensitive than USSIV estimates to the sample split.

Results of the same experiment using randomly assigned fictitious instruments are reported in columns (3) and (4) of Table 3. The SSIV coefficient estimates are centered on 0, with a Monte Carlo standard deviation of .014. The average estimate of θ in this experiment is .086, suggesting substantial bias downward, and this estimate is not significantly different from 0. An insignificant estimate of θ means that the researcher cannot reject the hypothesis that the true vector of reduced-form coefficients is 0. In that case, the SSIV estimate is 0 regardless of the correlation between ϵ and η in Equations (1) and (2) and the (group-asymptotic) moments of the USSIV coefficient do not exist.

The USSIV coefficient estimates with randomly generated instruments are highly variable, with a Monte Carlo standard deviation of .187. Their individual standard errors are also high—on the order of .13—which is about double the size of the OLS coefficient estimate. Although the USSIV coefficients are centered near 0, the key result here is that they have very large sampling variance and would be unlikely to lead to an apparently credible inference. In contrast, using the same randomly generated instruments in 2SLS estimation yields a coefficient estimate of .057 with a reported standard error of .014. Thus, unlike SSIV and USSIV, 2SLS results with fictitious instruments look remarkably like the OLS estimates.

4.2 Angrist and Krueger (1992a)

In AK-92, we used the 1970–1972 draft lotteries to construct instruments for the education of men at risk of induction during the Vietnam era. The lotteries worked by assigning a random sequence number (RSN) to dates of birth in cohorts at risk of being drafted. The lowest numbers were called first, up to an administratively determined ceiling. Men with numbers above the ceiling were not drafted. In certain years, men could be deferred or exempted from military service by remaining in school and thereby obtaining an educational deferment. Thus draft-lottery numbers affected both the likelihood of serving in the military and the incentive to seek additional schooling.

Angrist (1990) showed that low lottery numbers are associated with an increased probability of military service and reduced Social Security earnings. If this link represents the casual effect of veteran status, then the impact of military service on earnings must be accounted for if the draft lottery is also to be used to identify the effect of schooling on earnings. We therefore proposed the following model:

$$y_i = \beta_0' w_{0i} + \beta_1 s_i + \gamma v_i + \epsilon_i, \quad (8)$$

$$s_i = \pi_{10}' w_{0i} + \pi_{11}' w_{1i} + \eta_i, \quad (9)$$

and

$$v_i = \pi_{20}' w_{0i} + \pi_{21}' w_{1i} + u_i, \quad (10)$$

where v_i is a dummy for veteran status and w_{1i} is a vector of excluded instruments. The first equation captures the partial effects of the two endogenous regressors s_i and v_i on the outcome y_i. The latter two equations are reduced forms. The excluded instruments, w_{1i}, are dummies that indicate groups of consecutive RSN's interacted with dummies for years of birth from 1944–1953.

The data set used to estimate (8)–(10) consists of a sample of over 25,000 observations from six March Current Population Surveys (CPS's) that were specially prepared for us and includes information on draft-lottery numbers. The CPS extracts contain labor-market information for the years 1979 and 1981–1985. These data show that men born from 1950–1953 with low lottery numbers were indeed significantly and substantially more likely to have served in the military than men with high numbers.

Table 4 reports CPS estimates of Equation (9), which relates education to dummies for lottery numbers. Results from two models are reported, one where v_i is treated as an exogenous covariate and one where v_i is treated as endogenous in (8). The instruments are three dummies for coarse lottery-number groups (RSN 1–75, 76–150, and 151–225), interacted with 10 years of birth. The first-stage estimates do not show a consistent pattern and only a few of the individual coefficient estimates are positive. But the joint test of significance has a marginal significance level under 10% for both sets of estimates.

Table 5 reports OLS, 2SLS, SSIV, and USSIV estimates of schooling coefficients from Equation (8) corresponding to the first-stage estimates in Table 4. The estimates are for models in which veteran status is treated as an exogenous covariate (treating veteran status as endogenous has little impact on the estimated schooling coefficients.) For comparison, the OLS estimate of .059 is reported in column (1). The 2SLS estimate is .021 with a standard error of .029. The SSIV estimate is essentially 0, with a somewhat larger standard error than the 2SLS estimate. The attenuation bias in the SSIV estimate is .176, but the standard error associated with this parameter is .167. Thus the null hypothesis that the true reduced-form coefficients are 0 cannot be rejected. The USSIV estimate, although inflated by the inverse attenuation bias, is also virtually 0.

The main specification reported by AK-92 is replicated in Panel A of Table 6. Column (2) shows the 2SLS estimate generated by using 3 lottery-number dummies (indicating groups of 25 consecutive RSN's) interacted with 10 year-

Table 4. Lottery Number and Educational Attainment

	Veteran status exogenous[a,c]			*Veteran status endogenous*[b,c]		
		RSN			*RSN*	
Year of birth	*1–75* (1)	*76–150* (2)	*151–225* (3)	*1–75* (4)	*76–150* (5)	*151–225* (6)
1944	−.194	−.563	−.471	−.197	−.556	−.485
	(.171)	(.174)	(.177)	(.171)	(.174)	(.177)
1945	.375	.230	.457	.378	.235	.460
	(.175)	(.170)	(.175)	(.175)	(.170)	(.175)
1946	.301	.332	.298	.288	.328	.303
	(.157)	(.156)	(.163)	(.157)	(.156)	(.163)
1947	.003	−.228	.008	−.001	−.232	−.012
	(.151)	(.151)	(.147)	(.151)	(.151)	(.148)
1948	−.003	.068	−.004	−.012	.050	−.018
	(.156)	(.152)	(.153)	(.156)	(.152)	(.153)
1949	.057	.091	.079	.031	mi.076	.082
	(.155)	(.153)	(.150)	(.155)	(.153)	(.150)
1950	−.078	.097	−.131	−.134	.056	−.158
	(.151)	(.152)	(.152)	(.151)	(.152)	(.153)
1951	.358	.160	.139	.312	.122	.133
	(.147)	(.151)	(.146)	(.150)	(.151)	(.146)
1952	−.023	.025	.080	−.084	.011	.081
	(.147)	(.149)	(.149)	(.150)	(.149)	(.149)
1953	−.026	.104	.151	−.035	.102	.149
	(.149)	(.148)	(.150)	(.149)	(.148)	(.150)
P value for joint *F* test (df = 30)		.071			.080	

[a]Dependent variable is years of schooling.
[b]Dependent variable is years of schooling after removing the effect of predicted veteran status and covariates.
[c]Covariates are two race dummies, central city dummy, balance of SMSA dummy, marriage dummy, five year dummies, nine year-of-birth dummies, and eight region dummies. Veteran status is also a covariate when it is treated as exogenous. Sample size is 25,781. Standard errors are shown below the coefficients.

Table 5. Lottery Estimates With 30 Instruments

	Type of estimate			
	OLS	2SLS	SSIV	USSIV
Parameter	(1)	(2)	(3)	(4)
*Actual instruments (3 lottery dummies * 10 years of birth)*				
β	.059	.021	.0002	.0014
	(.001)	(.029)	(.032)	(.184)
θ	—	—	.176	—
			(.167)	
First-stage F (df = 30)	—	1.40	1.19	1.19

NOTE: The sample includes 25,781 observations on men born 1944–1953 in the March 1979 and 1981–1985 CPS Special Extracts. The table reports OLS estimates and 2SLS estimates of regressions in which years of schooling is the sole endogenous regressor. Other covariates include veteran status, a dummy for Blacks, a dummy for Hispanic and other races, dummies for residence in central city, other SMSA, and married with spouse present, five year dummies, nine year-of-birth dummies, and eight region dummies. The SSIV and USSIV estimates in both panels are based on a single sample split with 12,967 observations used for the cross-sample fitted values and 12,814 observations used for the second stage. The instruments include 3 lottery-number dummies (indicating RSN 1–75, 76–150 and 151–225) interacted with 10 year-of-birth dummies.

of-birth dummies. Coefficients on the additional covariates are not reported (for these see table 3 of AK-92). Although the results, using relatively few instruments, in Table 5 suggest that lottery-based estimation is not very informative, the 2SLS estimate in Table 6 using 130 instruments is .066 with a standard error of .015, a finding close to the OLS estimate of .059. The conventional asymptotic standard error of this estimate does not provide a warning of weak instruments based on the usual normal approximation.

In contrast with the 2SLS estimates, SSIV estimates in column (3) of Table 6 are .005 with a standard error of .016. Thus, unlike for most of the specifications reported by AK-91, the SSIV estimate in this case does not confirm the conventional 2SLS findings. The implied attenuation of SSIV is .088 with a standard error of .084. This is consistent with the null hypothesis that the lottery instruments are actually worthless for estimating schooling coefficients. Inflating SSIV by the attenuation bias generates a USSIV estimate of .062, but the standard error of this estimate is .177, again suggesting that little is learned from lottery-based instruments about the returns to schooling.

Table 6. Lottery Estimates With 130 Instruments

	Type of estimate			
	OLS	2SLS	SSIV	USSIV
Parameter	(1)	(2)	(3)	(4)
*A. Actual instruments (13 lottery dummies * 10 years of birth)*				
β	.059	.066	.005	.062
	(.001)	(.015)	(.016)	(.177)
θ	—	—	.088	—
			(.084)	
First-stage F df = 130	—	1.11	1.12	1.12
*B. Random instruments (13 multinominal dummies * 10 years of birth)*				
β	.059	.049	.025	1.65
	(.001)	(.018)	(.018)	(10.1)
θ	—	—	.015	—
			(.094)	
First-stage F (df = 130)	—	.82	.84	.84

NOTE: The sample includes 25,781 observations on men born 1944–53 in the March 1979 and 1981–1985 CPS Special Extracts. The table reports OLS estimates and 2SLS estimates of regressions in which years of schooling is the sole endogenous regressor. Other covariates include veteran status, a dummy for Blacks, a dummy for Hispanic and other races, dummies for residence in central city, other SMSA, and married with spouse present, five year dummies, nine year-of-birth dummies, and eight region dummies. The SSIV and USSIV estimates in both panels are based on a single sample split with 12,967 observations used for the cross-sample fitted values and 12,814 observations used for the second-stage. The instruments include 13 lottery-number dummies (indicating group of 25 consecutive RSN's), interacted with 10 year-of-birth dummies. The same sample split was used to compute estimates in both panels and in Tables 5 and 6.

As a final check on these models, estimates from the same specification using 13 fictitious randomly generated lottery-number dummies as instruments are reported in Panel B of Table 6. The 2SLS estimate here is .049 with a standard error of .018. This is smaller than the estimate using the actual instruments but does not lead to a dramatically different inference. As with the actual instruments, however, SSIV and USSIV provide strong evidence that the 2SLS result is spurious. The SSIV estimate is .025 with a standard error of .018 and the USSIV estimate is 1.65 with a standard error of 10.1.

5. CONCLUSIONS

SSIV and USSIV provide valuable complements to conventional 2SLS. SSIV estimates are biased toward 0 rather than toward the OLS probability limit. Thus with SSIV there is little risk of spurious or misleading inferences generated solely as a consequence of finite-sample bias. Moreover, the estimated SSIV attenuation bias can be used to inflate SSIV estimates and provide an asymptotically unbiased (under group asymptotics) USSIV estimate.

Our reinvestigation of Angrist and Krueger (1991) shows that SSIV and USSIV produce relatively precise parameter estimates that are close to the conventional 2SLS and OLS estimates. All of the IV estimators used here—2SLS, SSIV, and USSIV—lead to similar results for the 30 instrument specifications reported in that article. There is evidence of a problem for 2SLS estimates in the 180-instrument specifications, as well as for the SSIV estimates, which are biased toward 0. But the bias-corrected USSIV estimator generates statistically significant estimates close to 2SLS and OLS estimates for the 180-instrument specification.

In contrast, our reexamination of results from Angrist and Krueger (1992a) fails to support the findings reported in that article. 2SLS estimates are close to OLS estimates in a 130-instrument specification, but SSIV estimates are essentially 0, and both SSIV and USSIV estimates are statistically insignificant. These findings therefore suggest that draft-lottery numbers are not useful for estimating schooling coefficients.

A natural extension of the research agenda begun here is to develop more efficient estimators that use sample splitting to reduce bias with a minimal increase in sampling variance. For example, an estimator based on combining the two SSIV and USSIV estimators that could be produced from any single split will have lower sampling variance than the SSIV and USSIV estimators introduced here. With Guido Imbens, we are also working on a jackknifed "leave-one-out" version of USSIV based on a separate first stage and fitted value for each observation. This estimator has the same asymptotic distribution as 2SLS with the desirable bias properties of USSIV.

ACKNOWLEDGMENTS

Thanks go to Kevin McCormick, Ronald Tucker, and Greg Weyland at the Census Bureau for creating a special Current Population Extract for us. Thanks also go to David Card, Gary Chamberlain, Guido Imbens, Jim Powell, George Tauchen, and seminar participants at National Bureau of Economic Research Labor Studies meeting for helpful discussions and comments, and to three anonymous referees for detailed written comments. We will make an SAS computer program available to anyone interested in using the estimators presented in this article. This research was supported by National Science Foundation Grant SES-9012149.

APPENDIX: PROOFS

Proof of Corollary 1.1. We need to show that

$$E[\widehat{X}_{21}'\widehat{X}_{21}/n_1] = \{\pi' E(z_i z_i')\pi + c\sigma_\eta^2 L_1\} \quad (A.1)$$

and

$$E[\widehat{X}_{21}' X_1/n_1] = \{\pi' E(z_i z_i')\pi\} \quad (A.2)$$

where $c \equiv \mathrm{tr}\{E[(Z_2'Z_2)^{-1}(Z_1'Z_1)]/n_1\}$ and L_1 is a $(k+p)$ square matrix consisting of zeros except for a 1 in the lower right corner. Note that

$$\widehat{X}_{21} = Z_1(Z_2'Z_2)^{-1}Z_2'X_2 = Z_1\pi + Z_1(Z_2'Z_2)^{-1}Z_2'\eta_2 \quad (A.3)$$

and

$$X_1 = Z_1\pi + \eta_1. \quad (A.4)$$

Using the independence of the two samples and the fact that $E[\eta_2 \mid Z_2] = 0$, $E[\widehat{X}_{21}'\widehat{X}_{21}/n_1]$ simplifies to

$$E[\pi' Z_1' Z_1 \pi/n_1] + E[\eta_2' Z_2(Z_2'Z_2)^{-1}Z_1'Z_1(Z_2'Z_2)^{-1}Z_2'\eta_2/n_1].$$

We have, $E[\pi' Z' Z\pi/n_1] = \{\pi' E(z_i z_i')\pi\}$ by virtue of iid sampling. To simplify $E[\eta_2' Z_2(Z_2'Z_2)^{-1}Z_1'Z_1(Z_2'Z_2)^{-1}Z_2'\eta_2/n_1]$, let η_2^* be the column of η_2 corresponding to s_i. Then,

$$E[\eta_2' Z_2(Z_2'Z_2)^{-1}Z_1'Z_1(Z_2'Z_2)^{-1}Z_2'\eta_2/n_1] = E[\eta_2^{*\prime} Z_2(Z_2'Z_2)^{-1}Z_1'Z_1(Z_2'Z_2)^{-1}Z_2'\eta_2^*]L_1/n_1.$$

Using properties of the trace operator, we have

$$E[\eta_2^{*\prime} Z_2(Z_2'Z_2)^{-1}Z_1'Z_1(Z_2'Z_2)^{-1}Z_2'\eta_2^*] = E[\mathrm{tr}\{Z_2'\eta_2^*\eta_2^{*\prime}Z_2(Z_2'Z_2)^{-1}Z_1'Z_1(Z_2'Z_2)^{-1}\}].$$

Iterating expectations, passing the expectation through the trace, and using the fact that $E[\eta_2^*\eta_2^{*\prime} \mid Z_2] = \sigma_\eta^2 I_n$, gives $E[\mathrm{tr}\{Z_2'\eta_2^*\eta_2^{*\prime}Z_2(Z_2'Z_2)^{-1}Z_1'Z_1(Z_2'Z_2)^{-1}\}] = E[\mathrm{tr}\{Z_1'Z_1(Z_2'Z_2)^{-1}\}]\sigma_\eta^2$. This establishes (A.1).

To simplify (A.2), use (A.3) and (A.4) to write

$$E[\widehat{X}_{21}'X_1] = E[\pi' Z_1' Z_1 \pi] + E[\pi' Z_1' \eta_1] + E[\eta_2' Z_2(Z_2'Z_2)^{-1}Z_1'Z_1\pi] + E[\eta_2' Z_2 (Z_2'Z_2)^{-1}Z_1'\eta_1]. \quad (A.5)$$

Because the two samples are independent and η_j is mean-independent of Z_j, only the first term on the right side of (A.5) is nonzero.

Derivation of Equation (6). Recall that $\beta = [\beta_0' \ \beta_1]'$. Write the $(p+1)\times(p+1)$ matrix, $\pi' E(z_i z_i')\pi$, as a conformably partitioned matrix:

$$\begin{bmatrix} P & R \\ R' & Q \end{bmatrix},$$

where p is $p \times p$, Q is a scalar, and R is $p \times 1$. Moreover, let $q = c\sigma_\eta^2$. Using the partitioned inversion formula (Theil 1971, p. 18), we have

$$[\pi' E(z_i z_i')\pi]^{-1} = \begin{bmatrix} P & R \\ R' & Q \end{bmatrix}^{-1}, = \begin{bmatrix} P^{-1} + P^{-1}RR'P^{-1}(1/\phi) & -P^{-1}R(1/\phi) \\ R'P^{-1}(1/\phi) & (1/\phi) \end{bmatrix}, \quad (A.6)$$

where $\phi \equiv Q - R'P^{-1}R$ is a scalar. We can use (A.6) to write

$$\{\pi' E(Z_i Z_i')\pi + c\sigma_\eta^2 L_1\}^{-1} = \begin{bmatrix} P & R \\ R' & Q+q \end{bmatrix}^{-1} = [\phi/(\phi+q)]$$

and

$$\left[(1/\phi)\begin{bmatrix} P^{-1}\phi + P^{-1}RR'P^{-1} & -P^{-1}R \\ -P^{-1}R' & 1 \end{bmatrix} + (1/\phi)\begin{bmatrix} P^{-1}q & 0 \\ 0 & 0 \end{bmatrix}\right].$$

The first term in curly brackets equals $[\pi' E(z_i z_i')\pi]^{-1}$. Therefore,

$$\{\pi' E(Z_i Z_i')\pi + c\sigma_\eta^2 L_1\}^{-1}\{\pi' E(Z_i Z_i')\pi\} = [\phi/(\phi+q)]I_{p+1} + [q/(\phi+q)]\begin{bmatrix} I_p & P^{-1}R \\ 0 & 0 \end{bmatrix}.$$

Multiplying this times β gives Equation (6) in the text. Because $\{\pi' E(z_i z_i')\pi\}$ is positive definite, $1/\phi \equiv 1/[Q - R'P^{-1}R]$, which is the lower right diagonal element of $[\pi' E(z_i z_i')\pi]^{-1}$, must be positive. Finally, note that $c > 0$ because $(Z_2'Z_2)^{-1}(Z_1'Z_1)$ is positive definite. Thus the proportional bias in estimates of β_1, $[\phi/(\phi+q)] = [\phi/(\phi + c\sigma_\eta^2)]$, is between 0 and 1.

[Received March 1993. Revised September 1994.]

REFERENCES

Altonji, J., and Segal, L. M. (1994), "Small-Sample Bias in GMM Estimation of Covariance Structures," Technical Working Paper 156, National Bureau of Economic Research, Cambridge, MA.

Angrist, J. D. (1990), "Lifetime Earnings and the Vietnam-Era Draft Lottery: Evidence From Social Security Administrative Records," *American Economic Review,* 80, 313–336.

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

——— (1992a), "Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery," Working Paper 4067, National Bureau of Economic Research, Cambridge, MA.

——— (1992b), "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.

Bekker, P. A. (1994), "Alternative Approximations to the Distributions of Instrumental Variables Estimators," *Econometrica,* 62, 657–682.

Bound, J., Jaeger, D., and Baker, R. (in press), "Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak," *Journal of the American Statistical Association*, 90.

Buse, A. (1992), "The Bias of Instrumental Variables Estimators," *Econometrica*, 60, 173–180.

Deaton, A. (1985), "Panel Data From a Time Series of Cross-Sections," *Journal of Econometrics*, 30, 109–126.

Fuller, W. A. (1975), "Regression Analysis for Sample Surveys," *Sankhyā*, Ser. C, 37, 312–326.

Hall, R. A., Rudebusch, G. D., and Wilcox, D. W. (1994), "Judging Instrument Relevance in Instrumental Variables Estimation," Discussion Paper 94-3, Federal Reserve Board Division of Monetary Affairs, Division of Research and Statistics, Finance and Economics, Washington, DC.

Maddala, G. S., and Jeong, J. (1992), "On the Exact Small Sample Distribution of the Instrumental Variables Estimator," *Econometrica*, 60, 181–183.

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., and Startz, R. (1990), "The Distribution of the Instrumental Variables Estimator and its t-ratio When the Instrument is a Poor One," *Journal of Business*, 63, part 2, S125–S140.

Staiger, D., and Stock, J. H. (1994), "Instrumental Variables Regressions With Weak Instruments," Technical Working Paper 151, National Bureau of Economic Research, Cambridge, MA.

Theil, H., (1971), *Principles of Econometrics*, Chicago: University of Chicago Press.

References

Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records
Joshua D. Angrist
The American Economic Review, Vol. 80, No. 3. (Jun., 1990), pp. 313-336.
Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28199006%2980%3A3%3C313%3ALEATVE%3E2.0.CO%3B2-H

Does Compulsory School Attendance Affect Schooling and Earnings?
Joshua D. Angrist; Alan B. Krueger
The Quarterly Journal of Economics, Vol. 106, No. 4. (Nov., 1991), pp. 979-1014.
Stable URL:
http://links.jstor.org/sici?sici=0033-5533%28199111%29106%3A4%3C979%3ADCSAAS%3E2.0.CO%3B2-P

The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples
Joshua D. Angrist; Alan B. Krueger
Journal of the American Statistical Association, Vol. 87, No. 418. (Jun., 1992), pp. 328-336.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199206%2987%3A418%3C328%3ATEOAAS%3E2.0.CO%3B2-A

LINKED CITATIONS

Alternative Approximations to the Distributions of Instrumental Variable Estimators
Paul A. Bekker
Econometrica, Vol. 62, No. 3. (May, 1994), pp. 657-681.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199405%2962%3A3%3C657%3AAATTDO%3E2.0.CO%3B2-E

Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak
John Bound; David A. Jaeger; Regina M. Baker
Journal of the American Statistical Association, Vol. 90, No. 430. (Jun., 1995), pp. 443-450.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199506%2990%3A430%3C443%3APWIVEW%3E2.0.CO%3B2-U

The Bias of Instrumental Variable Estimators
A. Buse
Econometrica, Vol. 60, No. 1. (Jan., 1992), pp. 173-180.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199201%2960%3A1%3C173%3ATBOIVE%3E2.0.CO%3B2-U

On the Exact Small Sample Distribution of the Instrumental Variable Estimator
G. S. Maddala; Jinook Jeong
Econometrica, Vol. 60, No. 1. (Jan., 1992), pp. 181-183.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199201%2960%3A1%3C181%3AOTESSD%3E2.0.CO%3B2-3

The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations
A. L. Nagar
Econometrica, Vol. 27, No. 4. (Oct., 1959), pp. 575-595.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28195910%2927%3A4%3C575%3ATBAMMO%3E2.0.CO%3B2-T

LINKED CITATIONS

The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument is a Poor One

Charles R. Nelson; Richard Startz

The Journal of Business, Vol. 63, No. 1, Part 2: A Conference in Honor of Merton H. Miller's Contributions to Finance and Economics. (Jan., 1990), pp. S125-S140.

Stable URL:

http://links.jstor.org/sici?sici=0021-9398%28199001%2963%3A1%3CS125%3ATDOTIV%3E2.0.CO%3B2-6