TECHNICAL WORKING PAPER SERIES

SPLIT SAMPLE INSTRUMENTAL VARIABLES

Joshua D. Angrist
Alan B. Krueger

Technical Working Paper No. 150

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 1994

We are grateful to David Card, Guido Imbens, and Jim Powell for helpful discussions and comments. The authors will make a SAS computer program available to anyone interested in using the estimators presented in this paper. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #150
January 1994

SPLIT SAMPLE INSTRUMENTAL VARIABLES

ABSTRACT

Instrumental Variables (IV) estimates tend to be biased in the same direction as Ordinary Least Squares (OLS) in finite samples if the instruments are weak. To address this problem we propose a new IV estimator which we call Split Sample Instrumental Variables (SSIV). SSIV works as follows: we randomly split the sample in half, and use one half of the sample to estimate parameters of the first-stage equation. We then use these estimated first-stage parameters to construct fitted values and second-stage parameter estimates using data from the other half sample. SSIV is biased toward zero, rather than toward the plim of the OLS estimate. However, an unbiased estimate of the attenuation bias of SSIV can be calculated. We use this estimate of the attenuation bias to derive an estimator that is asymptotically unbiased as the number of instruments tends to infinity, holding the number of observations per instrument fixed. We label this new estimator Unbiased Split Sample Instrumental Variables (USSIV). We apply SSIV and USSIV to the data used by Angrist and Krueger (1991) to estimate the payoff to education.

Joshua D. Angrist
Department of Economics
Hebrew University
Mount Scopus Campus
Jerusalem, ISRAEL
and NBER

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

There has been longstanding interest in the finite sample properties of Instrumental Variables (IV) estimators.[1] In an influential early paper, Nagar (1959) used an approximation argument to show that IV and other k-class estimators are biased toward the probability limit of Ordinary Least Squares (OLS) estimates in small samples with normal disturbances. Buse (1992) generalized this result to cases with non-normal disturbances. The Nagar approximation result shows that, other things equal, the bias of IV is greater if the excluded instruments explain a smaller share of the variation in the endogenous variable. Nelson and Startz (1990) demonstrate that, in samples of the size typically used in time series analyses, IV estimates and their t-ratios have non-normal distributions if the first-stage R-square is low. These results suggest that extreme caution should be used when interpreting IV estimates based on small samples.

Recently, Bound, Jaegar and Baker (1993) (henceforth BJB) have argued that finite sample bias may be a problem in cross-sectional studies that use large samples and employ many excluded instruments. In particular, BJB single out some of the IV specifications in our 1991 QJE article as evidence that finite sample bias could be severe even in large samples. In our 1991 paper, we used quarter of birth as an instrument for education in wage equations. We argued that quarter of birth is weakly correlated with years of schooling because students generally enter first grade in the fall of the year in which they turn six, but are permitted to drop out of school on their 16th birthday. Thus, students born earlier in the calendar year are allowed to drop out of school after having completed fewer years of schooling than students born later in the year. Because compulsory schooling laws

---

[1]Throughout the paper we use the terms Instrumental Variables and Two Stage Least Squares (2SLS) interchangeably because 2SLS can be thought of as an IV estimator in which the predicted endogenous regressor is used as the instrument.

and school entry laws vary across states, in some specifications the instrument list included interactions of quarter of birth with state-of-birth dummies.

Our analysis was primarily based on a sample of over 300,000 men from the 1980 Census. Most of the IV estimates of the return to education in our paper are slightly higher than the OLS estimates. BJB argue that these results are due to having weak instruments, and to "over fitting" the first-stage equation by including too many interaction terms in the instrument list. One important finding that BJB note is that seemingly plausible IV coefficient estimates (e.g., close to OLS) and standard errors can be obtained by using a large number of instruments whose values are drawn from a random number generator in a large sample. The possibility of such misleading inferences suggests the importance of developing IV estimators that are not biased toward OLS.

In this paper, we propose a new instrumental variables estimator which we call Split Sample Instrumental Variables (SSIV). SSIV works as follows: we randomly split the sample in half, and use one half of the sample to estimate parameters of the first-stage equation. We then use these estimated first-stage parameters to construct fitted values and second-stage parameter estimates using data from the other half sample. This estimator is a variation of the Two Sample Instrumental Variables (TSIV) estimator in Angrist and Krueger (1992), in which one sample is randomly divided in half to provide two independent samples.[2]

SSIV has several desirable properties vis-a-vis conventional IV and other k-class

[2]Altonji and Segal (1993) have recently discussed the use of sample-splitting to reduce bias in certain GMM estimators.

estimators. First, unlike conventional IV estimates, SSIV estimates are not biased towards OLS estimates in finite samples. Instead, SSIV estimates are biased toward zero regardless of the degree of covariance between structural and reduced form errors. Intuitively, the bias that arises in SSIV estimation is similar to attenuation bias that arises from measurement error in an independent variable in an OLS regression. Because the first-stage parameters in SSIV are subject to sampling error, the predicted endogenous regressor is measured with error. However, SSIV is a consistent estimator.

Second, we can obtain an unbiased estimate of the attenuation bias of SSIV. Namely, the coefficient from a regression of the endogenous regressor on its predicted value (using data from one half sample but first-stage parameters from the other) provides an unbiased estimate of the attenuation bias. Moreover, the product of the SSIV estimate and the inverse attenuation bias is asymptotically unbiased as the number of instruments grows, holding the number of observations per instrument constant. The asymptotic approach we use to establish the properties of USSIV is similar to that used by Deaton (1985) to study the behavior of estimators computed in repeated cross-sections of fixed size. A similar approach has also been used recently by Bekker (1992) to study conventional simultaneous equations estimators. It is worth noting that conventional IV and SSIV are still biased under this form of asymptotics. We call the asymptotically unbiased split sample estimator Unbiased Split Sample Instrumental Variables (USSIV).

In the remainder of the paper, we describe the properties of SSIV and USSIV. We then use these estimators to estimate variations on the specifications reported in Angrist and Krueger (1991). The SSIV and USSIV coefficient estimates have relatively small standard errors, and the estimates are similar to the conventional IV estimates in Angrist and Krueger (1991). This suggests that small sample bias is not responsible for the conclusions drawn from conventional IV in that article.

As a final check on the behavior of SSIV and USSIV estimators with extremely weak instruments, we present IV estimates computed using artificial, uniformly distributed instruments drawn from a random number generator. Using randomly generated instruments, conventional IV estimates are similar to OLS and have relatively small standard errors. But SSIV and USSIV estimates have very large standard errors and are not statistically different from zero in this case. Thus, unlike conventional instrumental variables estimators, SSIV and USSIV would correctly lead a researcher to conclude that the random instruments are of no value.

1. Small Sample Bias in IV Estimates

Consider the following two-equation model, where to simplify notation we assume there is one endogenous regressor:

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

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

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 over-identifying restrictions. x_i is a (p + 1) x 1 vector that includes the exogenous regressors and the endogenous regressor.

The data are more compactly denoted by an n x 1 vector Y, an n x (p+1) matrix X, and an n x (k+p) matrix Z. From (1) and (2), we have

$$Y = X\beta + \varepsilon$$

$$X = Z\pi + \eta.$$

The coefficient β_1 is the scalar parameter of interest, assumed to be the last element in the (p+1) x 1 vector β, and π is the (k + p) x (p + 1) matrix of reduced form parameters. We assume that observations in the sample are independent and identically distributed, and that the disturbances satisfy $E(\varepsilon_i \mid z_i) = E(\eta_i \mid z_i) = 0$. The residual variance of ε_i is denoted σ_ε^2. The vector of residual variances in equation (2), has one non-zero element in the last row corresponding to s_i. For this element, $E[\eta_i^2] = \sigma_\eta^2$ and $E(\varepsilon_i \eta_i) = \sigma_{\varepsilon\eta}$.

The core of the BJB critique is that the IV and 2SLS estimates in our 1991 paper are contaminated by small sample bias, thereby causing the 2SLS estimates to resemble the OLS estimates.[3] The bias toward OLS in a finite sample is easy to grasp intuitively. To see this, consider an extreme example: suppose the values of the instruments (Z) are randomly assigned. In any given sample, by chance some of the variation in X will be explained by Z, the random instruments. Thus, the first-stage R-square will be greater than

[3]BJB also argue that estimates in Angrist and Krueger (1991) are biased because quarter of birth is correlated with earnings for reasons other than compulsory attendance laws. In our 1991 paper, and in Angrist and Krueger (1992), we provide a detailed investigation of this issue. Interested readers are referred to these articles for more on this point.

0 in a finite sample. In this situation, the second stage parameter estimate will be biased toward the probability limit of the OLS coefficient because, with randomly generated instruments, the variation in X that is explained by the instruments will be like the "typical" variation in X.

One problem with the BJB argument for small-sample bias in our 1991 paper is that the bias should differ as the number of instruments changes. This is apparent from BJB's adaptation of the Buse (1992) and Nagar (1959) approximate bias formula. Using the above notation for an example where the only exogenous regressor is a constant, the approximate small sample bias of the 2SLS slope estimate is:

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

BJB point 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 the instruments in the first-stage equation (i.e., substituting π for OLS estimates in the usual F-statistic formula). From this formula, it is clear that if adding additional instruments does not improve the fit of the first-stage equation, then (other things equal) the approximate bias grows with the number of instruments.[4]

Yet for the middle-aged cohorts on whom we focused our analysis, the simple Wald estimates (which use one excluded instrument) are statistically indistinguishable from 2SLS estimates that use 3 quarter dummies interacted with 10 years of birth dummies as excluded instruments (30 instruments), and from 2SLS estimates that add 50 state of birth dummies

[4] Staiger and Stock (1993) make a similar point.

interacted with 3 quarter of birth dummies (a total of 180 instruments). These alternate estimates are presented in Table 1. The remarkable consistency across these model specifications could be compatible with the approximate bias formula in (3) if small-sample bias is not much of a problem, or if the OLS estimates are themselves unbiased. A final possibility is that the bias formula in (3) does not provide a very good approximation to the first moments of 2SLS in this example.[5]

Nevertheless, one of the specifications that we reported -- that using a full set of state-of-birth dummies interacted with quarter-of-birth dummies and quarter-of-birth dummies interacted with year-of-birth dummies (180 instruments) -- may still be biased towards the OLS results. We discovered this possibility by experimenting with 2SLS estimates in a "worst-case" scenario in which the instruments are randomly generated. In particular, we assigned quarter of birth dummies randomly from a random number generator, and then used these random instruments to re-estimate models from our 1991 QJE paper. For most of these models, the 2SLS estimate of the schooling coefficient is close to the OLS estimate of .06. The standard error for the 180-instrument specification was surprisingly low with randomly assigned instruments, typically around .017. The standard error for the 30-instrument model was higher, around .04 with random instruments.

BJB subsequently reported results from a similar set of simulation exercises in their paper, focusing on the 180-instrument models. They conclude that "the similarity of the average point estimates to the OLS results and of the average estimated standard errors to

---

[5]See Sargan (1974) for a discussion of the validity of the Nagar expansion. Note that the Nagar/Buse approximate bias becomes infinite if the true π is zero.

those using the actual quarter of birth data is striking."

Note, however, that the estimated standard errors from this simulation exercise are greater than the comparable 2SLS standard errors estimated using the actual data as instruments (.009 and .016 for the 180-instrument and 30-instrument models). For a model estimated using 180 quarter-of-birth interactions as instruments, the ratio of the estimated variance using random instruments to the estimated variance with real instruments is 3.6 = $(.017/.009)^2$. This suggests that something is being accomplished by the real instruments. Nevertheless, it is disturbing that IV estimates with randomly generated instruments produce a possibly misleading inference. In the remainder of this paper, we show that misleading inferences of this nature are unlikely to occur using split-sample estimators.

2. Split Sample Instrumental Variables (SSIV)

2.1 Basic Results

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. These 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. Algebraically, the estimator is:

$$\hat{\beta}_s = (\hat{X}_{21}'\hat{X}_{21})^{-1}\hat{X}_{21}'Y_1$$

$$= [X_2'Z_2(Z_2'Z_2)^{-1} Z_1'Z_1 (Z_2'Z_2)^{-1} Z_2'X_2]^{-1}[X_2'Z_2(Z_2'Z_2)^{-1} Z_1'Y_1] \quad (4)$$

where

$$\hat{X}_{21} = Z_1(Z_2'Z_2)^{-1} Z_2'X_2$$

is the cross-sample fitted value. Notice that for columns of X that are included in Z, the estimated reduced form coefficient from sample 2 will be exactly equal to 1. The cross-sample fitted value for exogenous regressors is therefore the actual value of the exogenous regressors in sample 1.

An important feature of SSIV is the fact that, as long as observations from the two half-samples are independent, the bias of the SSIV estimator can be expressed in a simple form without approximation. The formal independence assumption we impose is

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

Assumption 1 is somewhat stronger than necessary, but seems a natural assumption to make in our cross-section framework where the observations are i.i.d. and the samples have been randomly divided. Assumption 1 implies that $\{Y_1, X_1\}$ is jointly independent of $\{Y_2, X_2, Z_2\}$ given Z_1. This implication is used to prove the following:

Proposition 1.

$E(\beta_s) = E(\theta)\beta = \theta\beta$ where θ is a (p+1) x (p+1) matrix,

$$\theta = [X_2'Z_2(Z_2'Z_2)^{-1} Z_1'Z_1 (Z_2'Z_2)^{-1} Z_2'X_2]^{-1}[X_2'Z_2(Z_2'Z_2)^{-1} Z_1'X_1],$$

Let $\hat{S}_{21}$ represent the cross-sample fitted value of S, and let S_1 represent the endogenous regressor in sample 1. The lower-right-hand-corner element of $\hat{\theta}$ is equivalent to the coefficient on $\hat{S}_{21}$ from a regression of S_1 on $\hat{S}_{21}$ and all the exogenous regressors. As shown below, this provides an estimate of the proportional bias in SSIV of β_1.

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

$$\hat{\beta}_s = \hat{\theta}\beta + [X_2'Z_2(Z_2'Z_2)^{-1} Z_1'Z_1 (Z_2'Z_2)^{-1} Z_2'X_2]^{-1}X_2'Z_2(Z_2'Z_2)^{-1} Z_1'\epsilon_1]$$

Iterating expectations over Z_1, we have

$$E[X_2'Z_2(Z_2'Z_2)^{-1} Z_1'Z_1 (Z_2'Z_2)^{-1} Z_2'X_2]^{-1}[X_2'Z_2(Z_2'Z_2)^{-1} Z_1'\epsilon_1]$$

$$= E\{ E[X_2'Z_2(Z_2'Z_2)^{-1} Z_1'Z_1 (Z_2'Z_2)^{-1} Z_2'X_2]^{-1}[X_2'Z_2(Z_2'Z_2)^{-1} Z_1'\epsilon_1 \mid Z_1] \}.$$

Using Assumption 1, this is

$$E\{ E[X_2'Z_2(Z_2'Z_2)^{-1} Z_1'Z_1 (Z_2'Z_2)^{-1} Z_2'X_2]^{-1}[X_2'Z_2(Z_2'Z_2)^{-1}] \cdot Z_1' \cdot$$

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

Now, note that $\hat{\theta}$ 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 $X_1 = [W_{01}\ S_1]$, and $Z_1(Z_2'Z_2)^{-1}Z_2'X_2 = [W_{01}\ \hat{S}_{21}]$. Regressing $[W_{01}\ S_1]$ on $[W_{01}\ \hat{S}_{21}]$ gives the matrix of coefficients:

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

where I_p is a pxp identity matrix, $\hat{\theta}_p$ is px1, $\hat{\theta}_{p+1}$ is a scalar equal to the coefficient on $\hat{S}_{21}$ in a regression of S_1 on $\hat{S}_{21}$ and W_{01}.

Note that for conventional IV estimators, $\hat{\theta}$ is always one. In SSIV estimation, $\hat{\theta}$ is a matrix of coefficients from a regression of X_1 on $\hat{X}_{21}$, and reflects a kind of attenuation bias arising from the use of reduced form coefficients from a separate sample. We illustrate this point more formally for a special case where $E[X_1 \mid \hat{X}_{21}]$ is linear. Although this example is not particularly realistic, we show in the next section that it illustrates some general features of $\hat{\theta}$.[6]

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

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

(5b) $$= \{\pi' E(Z_i Z_i')\pi + c\sigma_\eta^2 L_1\}^{-1}\{\pi' E(Z_i Z_i')\pi\}$$

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

Proof. If $E[X_1 \mid \hat{X}_{21}]$ is linear, then $E[X_1 \mid \hat{X}_{21}]$ can be written $\hat{X}_{21}\{E[\hat{X}_{21}'\hat{X}_{21}]^{-1} E[\hat{X}_{21}'X_1]\}$. Since $\hat{\theta} = [\hat{X}_{21}'\hat{X}_{21}]^{-1} [\hat{X}_{21}'X_1]$, we can substitute for X_1 to show that $E[\hat{\theta}] = E[\hat{X}_{21}'\hat{X}_{21}/n]^{-1} * E[\hat{X}_{21}'X_1/n]$. In the appendix, the moments in the numerator and denominator are simplified to complete the proof. The intuition behind equation (5b) is that the denominator reflects the sampling variance in the estimated first stage parameters. Note that if the cross-product matrices (Z'Z) in the two samples are the same, then $c = (k+p)/n$. The matrix L_1 reflects the fact that X_i includes only 1 endogenous regressor, s_i.

[6]One situation where $E[X_1 \mid \hat{X}_{21}]$ is linear is when $\hat{X}_{21}$ and X_1 are jointly normally distributed. X_1 is normally distributed if Z_1 and η are normally distributed. $\hat{X}_{21}$ will also be approximately normally distributed in this case if n is large enough to make the sampling variance of $(Z_2'Z_2)^{-1}Z_2'X_2$ negligible.

From (5b), it is apparent that if there are no exogenous variables, then the proportional bias of SSIV is between zero and one -- that is, the SSIV coefficient will be biased toward 0 in absolute value. More generally, (5b) implies a matrix attenuation bias, familiar from multivariate measurement error models (e.g., Fuller 1975). Matrix attenuation does not necessarily imply attenuation of the individual coefficients in the vector β. In this case, however, it does imply attenuation of the coefficient on the single endogenous regressor, β_1. To see this write the (p+1) x (p+1) matrix, $\pi' E(Z_i Z_i')\pi$, as a partitioned matrix,

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

where P is p x p, Q is a scalar, and R is a p x 1 vector. Given (5b), it is possible to show that

$$\text{(6)} \qquad \theta\beta = \begin{matrix} \beta_0 + [c\sigma_\eta^2/(\phi + c\sigma_\eta^2)]P^{-1}R\beta_1 \\ [\phi/(\phi + c\sigma_\eta^2)]\beta_1 \end{matrix}$$

where $\phi \equiv [Q - R'P^{-1}R]$ is a positive scalar, so that $\phi/(\phi + c\sigma_\eta^2)$ is necessarily between zero and one. Equation (6) is derived in the appendix.

It is of interest to consider the bias formulas, (5b) and (6), in a number of special cases. First, if the reduced form residual variance in Sample 2 (σ_η^2) equals zero, SSIV estimates are unbiased. Another case of interest is when the true reduced form coefficients, π, are equal to zero (or are very small). In this case, it is apparent from equation (5) that the SSIV estimate of β_1 has expectation zero (or near zero). Moreover, increasing the number of instruments also tends to pull SSIV estimates towards zero. These results contrast sharply with the tendency of conventional IV estimates to be biased towards OLS, or to diverge to infinity if the true reduced form coefficients are zero.

2.2 Conventional Asymptotic Results for SSIV

We complete this section by summarizing important asymptotic properties of $\hat{\beta}$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 fraction, α. Under standard conditions, $n^{1/2}E[g_n(\beta)] \sim N(0, \Omega)$, where Ω is a (p+k)x(p+k) covariance matrix. If this holds, we have:

(i) $n^{1/2}(\hat{\beta}_s - \beta) \sim N(0, \psi)$ 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 cross-product matrices.

(ii) If $\beta = 0$ and $E[\epsilon^2 \mid Z] = \sigma_\epsilon^2$, then SSIV is the most efficient TSIV estimator for a given sample split.

Proof. This proposition is a straightforward application of results in our earlier paper on TSIV (Angrist and Krueger 1992, Lemma 1). The general TSIV estimator is $[X_2'Z_2\Phi^{-1}X_2'X_2]^{-1}[X_2'Z_2\Phi^{-1}Z_1'Y_1]$, where Φ is any positive definite weighting matrix. To see that SSIV is a special case of TSIV, note that without changing asymptotic results, we can normalize each Z_j by multiplying times $(Z_j'Z_j)^{-1}$. Using the normalized instruments and setting $\Phi = (Z_1'Z_1)^{-1}$ in the TSIV formula gives SSIV. In general, setting $\Phi = \Omega^{-1}$ gives the optimal TSIV estimator. If we assume $E[\epsilon^2 \mid Z] = \sigma_\epsilon^2$ and $\beta=0$, then $\Omega = \Sigma_{zz}^{-1}\sigma_\epsilon^2$. In that case, choosing $\Phi = (Z_1'Z_1)^{-1}$ gives the optimal TSIV estimator.

SSIV has a number of practical advantages over other TSIV estimators. First, SSIV is easy to compute using regression software. Second, if $\Omega = \Sigma_{zz}^{-1}\sigma_\epsilon^2$, then the asymptotic covariance matrix of SSIV simplifies to $(\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{xz})^{-1}\sigma_\epsilon^2$. This is the usual form of the 2SLS asymptotic covariance matrix. It can be computed from regression software by calculating the true 2SLS residuals (using regressors and not fitted values) and then adjusting the software-reported residual variance estimate.

3. Group Asymptotics

Proposition 1 characterizes the expectation of $\hat{\beta}_s$ without resorting to complicated approximation arguments. A simple unbiased estimator of the proportional bias of $\hat{\beta}_s$ has also been derived. To explore the nature of this bias further, without relying on normality or linearity assumptions for simplifications, we develop an asymptotic argument that we think captures important features of the finite sample behavior of $\hat{\beta}_s$.

The approach we use is based on a parameter and data sequence that involves replicating cross-sections of equal size to estimate the same parameter vector, β. In the context of Angrist and Krueger (1991), this replication process 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. The group-asymptotics approach derives the limiting characteristics of $\hat{\beta}_s$ as the number of instruments grows, but the number of observations per instrument is held fixed. This involves the same type of argument used by Deaton (1985) in his study of panel data created from a time series of cross-sections. The asymptotic results in Deaton's paper are based on the time dimension of the panel, rather than more conventional cross-section asymptotics. Our group-asymptotics approach is also similar to the asymptotic argument used in Bekker's (1992) study of simultaneous equations estimators.

We think of each cross-section as representing a group of m observations (e.g., state of birth), and each group as providing an additional excluded instrument. The t^{th} cross-section replication is assumed to contain 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 $\{Y_{jt}, X_{jt}, Z_{jt}\}$, for $j=1,2$. For simplicity, we assume that the replications are i.i.d. An important feature of this sequence of replications is that there is assumed to be a different matrix of reduced form coefficients associated with each replication. In particular, we assume that at each replication two half-samples and a reduced form coefficient matrix, π_t, are drawn. The π_t, are themselves i.i.d. random matrices satisfying $E[X_{jt} - Z_{jt}\pi_t \mid Z_{jt}] = 0$ and $E[(X_{jt} - Z_{jt}\pi_t)^2 \mid Z_{jt}] = \sigma_\eta^2$ for $j=1,2$ and the residual component corresponding to s_i. Each π_t is also assumed to be independent of the data in each half sample, and the two samples for each replication are assumed to be independent of each other (as in Assumption 1).

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,

$$\hat{X}_{21} = [\hat{X}_{21,1}' \ldots \hat{X}_{21,t}' \ldots \hat{X}_{21,T}']'$$

where $\hat{X}_{21,t} = Z_{1t}(Z_{2t}'Z_{2t})^{-1} Z_{2t}'X_{2t}$. 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. The resulting estimator can be written:

$$\hat{\beta}_s = (\hat{X}_{21}'\hat{X}_{21})^{-1}\hat{X}_{21}'Y_1$$

$$= [(1/T)\Sigma_t \hat{X}_{21,t}'\hat{X}_{21,t}/m]^{-1} [(1/T)\Sigma_t \hat{X}_{21,t}'Y_{1t}/m].$$

We define the *group-asymptotic probability limit* of $\hat{\beta}_s$ as the probability limit when the number of groups becomes infinite while the group size is fixed. This probability limit turns out to be similar to the expectation derived in Proposition 1, where a linear conditional expectation was assumed. The corresponding group-asymptotic result is presented below:

<u>Proposition 3</u>. $\underset{T\to\infty}{\text{plim}}\ [(1/T)\Sigma_t \hat{X}_{21,t}'\hat{X}_{21,t}/m]^{-1} [(1/T)\Sigma_t \hat{X}_{21,t}'Y_{1t}/m]$

$$= \{E[\pi_t'(Z_{1t}'Z_{1t}/m)\pi_t] + c^*\sigma_\eta^2L_1\}^{-1}E[\pi_t('Z_{1t}'Z_{1t}/m)\pi_t]\beta$$

where $c^* \equiv tr\{E[(Z_{2t}'Z_{2t})^{-1}(Z_{1t}'Z_{1t})]/m\}$.

Proof. The first step is to note that

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

We also have,

$$\hat{X}_{21,t} = Z_{1t}\pi_t + Z_{1t}(Z_{2t}'Z_{2t})^{-1}Z_{2t}'\eta_{2t}$$

$$Y_{1t} = (Z_{1t}\pi_t + \eta_{1t})\beta + \epsilon_{1t}$$

The proof is completed by using these expressions 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, $\{E[\pi_t'(Z_{1t}'Z_{1t}/m)\pi_t] + c^*\sigma_\eta^2L_1\}^{-1}E[\pi_t('Z_{1t}'Z_{1t}/m)\pi_t]$, implies scalar attenuation of the coefficient β_1.

Although Proposition 3 is a special kind of asymptotic result, the group-asymptotics characterization of $\hat{\beta}_s$ should provide at least as good an approximation to the first moments of the SSIV estimator as Nagar-type approximation results. To see this, consider the group-asymptotics probability limit of conventional 2SLS in T repeated cross-sections as T grows to infinity. As in the proof of Proposition 3, the 2SLS estimator has group-asymptotics probability limit

$$\text{plim}\ [(1/T)\textstyle\sum_t[\hat{X}_t'\hat{X}_t/m]^{-1}\ [(1/T)\sum_t[\hat{X}_t'Y_t/m] = E[\hat{X}_t'\hat{X}_t/m]^{-1} * E[\hat{X}_t'Y_t/m].$$

where,

(7) $$\hat{X}_t = Z_t(Z_t'Z_t)^{-1}Z_t'X_t = Z_t\pi_t + Z_t(Z_t'Z_t)^{-1}Z_t'\eta_t$$

$$Y_t = (Z_t\pi_t + \eta_t)\beta + \epsilon_t\ .$$

In this case, 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. Using (7) it is therefore possible to show that

(8) $$E[\hat{X}_t'\hat{X}_t/m]^{-1} * E[\hat{X}_t'Y_t/m]$$

$$= \beta + [(k+p)/m]\{E[\pi_t'(Z_{1t}'Z_{1t}/m)\pi_t] + [(k+p)/m]\sigma_\eta^2L_1\}^{-1}\ell_1\sigma_{\epsilon\eta},$$

where ℓ_1 is a (p + 1) vector consisting of zeros in the first p rows and a 1 in the last row.[7]

From (8) it is apparent that 2SLS is not consistent under group asymptotics.[8] Note that (8) is similar to the Nagar bias result used by Buse (1992) and BJB (see equation (3), above). The principal difference between (8) and the Nagar approximation formula is that, through the term σ_η^2 in the denominator, (8) reflects sampling variability in estimates of the reduced form. We conjecture that because of this, formula (8) will usually do a better job characterizing the small-sample properties of conventional 2SLS estimators than will the Nagar formula.

---

[7]As in the proof of Proposition 2, algebra for this result follows steps similar to those outlined in the Appendix for Corollary 1.1 .

[8]Equation (8) is a variant of the result derived by Bekker (1992) using a parameter sequence that fixes the explained sum of squares in the reduced form sum of squares, $\pi'Z'Z\pi$, while the number of instruments grows.

4. Unbiased Split-Sample Estimation

It seems reasonable to try to improve on SSIV by inflating the estimates, $\hat{\beta}_s$, by the inverse of the estimated proportional bias, $\hat{\theta}$. The resulting estimator, $\hat{\theta}^{-1}\hat{\beta}_s$, is not unbiased, however, because it involves a nonlinear function of the (correlated) random variables $\hat{\theta}$ and $\hat{\beta}_s$. Nevertheless, the inflated estimator is unbiased under the group-asymptotic argument outlined above. We therefore label this inflated estimator as an Unbiased Split-Sample Instrumental Variables (USSIV) estimator.

Recall that $\hat{\theta} = [\hat{X}_{21}'\hat{X}_{21}]^{-1} [\hat{X}_{21}'X_1]$. Then the USSIV estimator can be written

$$\hat{\beta}_u \equiv \hat{\theta}^{-1}\hat{\beta}_s$$

$$= [\hat{X}_{21}'X_1]^{-1}[\hat{X}_{21}'Y_1] = [(1/T)\Sigma_t \hat{X}_{21,t}'X_{1t}/m\ ]^{-1} [(1/T)\Sigma_t \hat{X}_{21,t}'Y_{1t}/m\ ]$$

Note that $\hat{\beta}_u$ can be constructed by using $\hat{X}_{21}$ as an instrument foı X_1 in the regression

$$Y_1 = X_1\beta + \epsilon_1.$$

Using $\hat{X}_{21}$ 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.

The properties of $\hat{\beta}_u$ are summarized in the next proposition:

Proposition 4. (i) $\operatorname*{plim}_{T\to\infty} [(1/T)\Sigma_t \hat{X}_{21,t}'X_{1t}/m\ ]^{-1} [(1/T)\Sigma_t \hat{X}_{21,t}'Y_{1t}/m\ ]$

$= \beta.$

(ii) $T^{1/2}(\hat{\beta}_u - \beta) \sim N(0, \Lambda)$ where

$$\Lambda = E[X_{1t}'\hat{X}_{21,t}/m]^{-1}E[\hat{X}_{21,t}'\hat{X}_{21,t}/m]E[\hat{X}_{21,t}'X_{1t}/m]^{-1}\sigma_\epsilon^2$$

Proof. To prove (i) we need to show that

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

$$\hat{X}_{21,t} = Z_{1t}(Z_{2t}'Z_{2t})^{-1}Z_{2t}'X_{2t} = Z_{1t}\pi_t + Z_{1t}(Z_{2t}'Z_{2t})^{-1}Z_{2t}'\eta_{2t}$$

$$X_{1t} = Z_{1t}\pi_t + \eta_{1t}$$

gives $E[\hat{X}_{21,t}'X_{1t}/m] = E[\pi_t'Z_{1t}'Z_{1t}\pi_t/m]$, exactly as in line (A5) of the proof to Corollary 1.1 . To derive the variance formula in (ii), substitute for Y_1 in $\hat{\beta}_u$:

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

so that

$$T^{1/2}(\hat{\beta}_u - \beta) = [(1/T)\Sigma_t \hat{X}_{21,t}'X_{1t}/m]^{-1} * T^{1/2}[(1/T)\Sigma_t \hat{X}_{21,t}'\epsilon_{1t}/m].$$

$$\sim E[\hat{X}_{21,t}'X_{1t}/m]^{-1} * T^{1/2}[(1/T)\Sigma_t \hat{X}_{21,t}'\epsilon_{1t}/m].$$

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

Proposition 4 provides a basis for inference using USSIV under group-asymptotics. The USSIV coefficient estimates and estimated asymptotic standard errors are also easy to compute. Note that $\hat{\beta}_u$ is a just-identified 2SLS estimator in sample 1, so that it can be written

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

The usual 2SLS covariance matrix estimator for a coefficient estimator of this form is $[X_1'\hat{X}_{21}(\hat{X}_{21}'\hat{X}_{21})^{-1}\hat{X}_{21}'X_1]^{-1}\sigma_\epsilon^2$. Multiplying this by mT provides a consistent estimate of Λ. Conventional 2SLS standard errors therefore provide a consistent estimate of the sampling variance of $\hat{\beta}_u$ under group-asymptotics.

5. Application: Angrist and Krueger (1991)

We apply SSIV and USSIV to the data set used by Angrist and Krueger (1991) to estimate the monetary return to education. In that paper we argued that quarter of birth provides a legitimate instrumental variable for years of schooling because children's age at school entry is related to their date of birth, and because compulsory schooling laws require children to attend school until their 16th or 17th birthday. We found that quarter of birth is weakly correlated with education and earnings for men born between 1920-49. Conventional IV 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 because the age-earnings profile is fairly flat for this age group. This avoids potential problems due to correlation between age and quarter of birth.

The first two columns of Table 2 report OLS and IV estimates of the education coefficient from log-wage equations. The analysis uses large samples from the 1980 and 1970 Censuses, and the specifications are the same as in Angrist and Krueger (1991). The IV model in Table 2 uses 30 quarter of birth x year of birth interactions as excluded instruments. SSIV and USSIV results are in columns 3 and 4. The standard error for the

SSIV estimates is about 50 percent greater than the conventional IV standard errors. Moreover, each of the SSIV estimates is somewhat less than the corresponding IV estimates, as one would expect since SSIV is biased toward zero. 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 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 the USSIV estimates.[9] These estimates inflate the SSIV estimates by the inverse of θ. The USSIV estimates tend to be above the OLS estimates, and are remarkably close to the conventional IV estimates.

The split sample approach can also be used to produce a visual impression of the data. Specifically, we randomly split the sample in half and then removed year-of-birth effects from earnings using one half of the sample, and removed year-of-birth effects from education in the other half. Finally, we graphed the average residual earnings by quarter of birth against the average residual education by quarter of birth. Figures 1a and 1b show the graphs for the 1970 and 1980 samples. The plots clearly show upward sloping relationships. The slope of this line can be shown to be an SSIV estimator in which each quarter of birth is given equal weight. For both the 1970 and 1980 data, the slope is .069.

Table 3 contains a set of OLS, IV, SSIV and USSIV results for models estimated using 150 quarter of birth x state of birth interactions as well 30 quarter of birth x year of

---

[9]The estimate of θ for the schooling coefficient is estimated from a regression of actual schooling in sample 1 on the cross-sample fitted schooling variable and all the exogenous variables.

birth interactions as the excluded instruments, with data from the 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 BJB and Staiger and Stock (1993) argue that IV estimates of these models are likely to be severely biased as a result. For this 180-excluded-instruments specification, the SSIV education coefficient is .031, about 40 percent as large as the IV estimate. However, the estimated proportional attenuation bias of SSIV in this case is also on the order of 40%. Thus, the USSIV estimate increases to .076, only slightly less than the IV estimate and above the OLS estimate. This result suggests that finite sample bias is not a not a serious problem with conventional IV here.

The SSIV and USSIV estimates reported in Tables 2 and 3 are based on a single random split of the data. To investigate the sensitivity of our results 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, each time using a different (randomly generated) seed number to split the data. The specifications estimated here use the 180 quarter of birth interactions as excluded instruments, as in Table 3. These results are reported in the first two columns of Table 4. 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 3 for a similar specification.[10] 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

---

[10]To save time and money, 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 4 and Table 3 are therefore not directly comparable.

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. The lower and upper quartiles for USSIV estimates are .099 and .129, giving an inter-quartile range of .03. This is considerably larger than the SSIV inter-quartile range of .013. Overall, however, both the SSIV and USSIV estimates with the actual instruments do not appear to be overly sensitive to the sample split.[11]

Finally, we investigated the performance of SSIV and USSIV in circumstances when the instruments have extremely weak explanatory power. In particular, we randomly assigned individuals' quarter of birth based on a uniform distribution with equal probability of being born in each month. We then conducted another Monte Carlo exercise in which we randomly divided the sample 31 times and estimated SSIV and USSIV for the 180-excluded-instrument model each time. These results are reported in columns 3 and 4 of Table 4. In general, SSIV and USSIV perform well in this situation. The SSIV coefficient estimates are centered on zero, with a relatively small Monte Carlo standard deviation. The average estimate of θ in this experiment is .086, suggesting substantial bias downwards. Such a low value of θ would be an indication that the instruments are exceedingly weak.

The USSIV coefficient estimates with randomly generated instruments are highly variable, and their individual standard errors are high -- on the order of .13 -- which is about double the OLS coefficient estimate. Although the USSIV coefficients are centered near

---

[11]In future work, we plan to explore this issue in a full-scale Monte Carlo study of SSIV, USSIV, other split-sample estimators, and conventional IV estimators. The purpose of this new study is to compare the estimators on mean-squared error and other grounds and to develop a framework for inference which accounts for variability due to the sample split in the reported sampling variance.

zero, the key result here is that they have huge sampling variances.

Simulations of both SSIV and USSIV using random instruments suggest these estimators would lead a researcher to conclude that the random instruments are not useful. In contrast, using the same randomly generated instruments in conventional IV estimation yields a coefficient estimate of .057 with a standard error of .014. Thus, unlike SSIV and USSIV, the conventional IV results look plausible and remarkably like OLS, even with randomly assigned instruments.

6. Conclusion

Split sample IV has many advantages over conventional IV. In particular, SSIV is biased toward zero, rather than toward the probability limit of the OLS estimate as is the case in IV. Moreover, this result does not rely on approximations (which may not be valid) as is the case with the Nagar (1959) and Buse (1990) IV bias formula. It is also possible to derive an unbiased estimate of the attenuation bias of SSIV. This estimate can then be used to inflate the SSIV estimate. The resulting estimator (USSIV) is asymptotically unbiased as the number of instruments tends to infinity, holding the number of observations per instrument fixed. In contrast, conventional IV is still biased under this type of asymptotics.

One shortcoming of the split-sample approach is that SSIV and USSIV estimators are asymptotically less efficient than conventional IV. But it is probably possible to improve the efficiency of SSIV and USSIV by combining the estimates described here with parallel estimates that use the opposite halves of the data to estimate the first and second stage equations. This approach will use all of the data, although the two estimates of the coefficient will not be independent. We are working on deriving the sampling variance of the estimator in this case, and on developing more efficient split-sample estimators.

We have also applied SSIV and USSIV to the data used by Angrist and Krueger (1991) to estimate the payoff to education. Our re-investigation shows that SSIV and USSIV produce "reasonable" standard errors, and parameter estimates that are close to the conventional IV and OLS estimates. Overall, we view these results as consistent with the conclusions based on conventional IV in Angrist and Krueger (1991). All the IV estimators used here -- 2SLS, SSIV, and USSIV -- lead to similar results for the 30 instrument specifications, so that small-sample bias does not appear to be a problem for the 30-instrument case. Our experiments with random instruments suggest that small-sample bias is potentially a problem for conventional IV estimates in the 180 instrument specification, as well as for the SSIV estimates, which are biased toward zero. But the USSIV estimator corrects the downward bias in SSIV estimates, and in this case generates estimates close to 2SLS for the 180-instrument specification.

Finally, our experiments with SSIV and USSIV show that these estimators do not give a misleading impression when the instruments are randomly assigned. This is a marked contrast with conventional IV estimators. In simulations where we randomly generate the instrumental variables, SSIV and USSIV tend to yield coefficients that are close to zero, and the estimated standard errors for USSIV are large. These findings suggest that SSIV and USSIV are likely to be a useful tool in other applications where researchers are concerned about the possibility of finite-sample bias in conventional IV estimates.

Appendix

Proof of Corollary 1.1. We need to show that:

(A1) $E[\hat{X}_{21}'\hat{X}_{21}/n] = \{\pi'E(Z_iZ_i')\pi + c\sigma_\eta^2 L_1\}$

(A2) $E[\hat{X}_{21}'X_1/n] = \{\pi'E(Z_iZ_i')\pi\}$

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

(A3) $\hat{X}_{21} = Z_1(Z_2'Z_2)^{-1}Z_2'X_2 = Z_1\pi + Z_1(Z_2'Z_2)^{-1}Z_2'\eta_2$

(A4) $X_1 = Z_1\pi + \eta_1$.

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

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

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

$$E[\eta_2'Z_2(Z_2'Z_2)^{-1}Z_1'Z_1(Z_2'Z_2)^{-1}Z_2'\eta_2/n]$$

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

Using properties of the trace operator, we have

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

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

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

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

$E[tr\{Z_2'\eta_2^*\eta_2^{*\prime}Z_2(Z_2'Z_2)^{-1}Z_1'Z_1(Z_2'Z_2)^{-1}\}] = E[tr\{Z_1'Z_1(Z_2'Z_2)^{-1}\}]\sigma_\eta^2$. This establishes A1.

To simplify A2, use A3 and A4 to write

$$\text{(A5)} \quad E[\hat{X}_{21}'X_1] = E[\pi'Z_1'Z_1\pi] + E[\pi'Z_1'\eta_1] + E[\eta_2'Z_2(Z_2'Z_2)^{-1}Z_1'Z_1\pi] + E[\eta_2'Z_2(Z_2'Z_2)^{-1}Z_1'\eta_1].$$

Because the two samples are independent and η_j is mean-independent of Z_j, only the first term on the right hand side of A5 is non-zero.

<u>Derivation of equation (6)</u>. Recall that $\beta = [\beta_0' \ \beta_1]'$. Write the (p+1) x (p+1) matrix, $\pi'E(Z_iZ_i')\pi$, as a conformably partitioned matrix:

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

where P is p x p, Q is a scalar, and R is p x 1. Also, let $q = c\sigma_\eta^2$. Using the partitioned inversion formula (Theil 1971, p. 18), we have $[\pi'E(Z_iZ_i')\pi]^{-1}$ =

$$\text{(A6)} \quad \begin{bmatrix} P & R \\ R' & Q \end{bmatrix}^{-1} = \begin{bmatrix} P^{-1} + P^{-1}RR'P^{-1}(1/\phi) & -P^{-1}R(1/\phi) \\ R'P^{-1}(1/\phi) & (1/\phi) \end{bmatrix},$$

where $\phi \equiv Q - R'P^{-1}R$ is a scalar. We can use A6 to write

$\{\pi'E(Z_iZ_i')\pi + c\sigma_\eta^2L_1\}^{-1}$ =

$$\begin{bmatrix} P & R \\ R' & Q + q \end{bmatrix}^{-1} = [\phi/(\phi + q)] \cdot$$

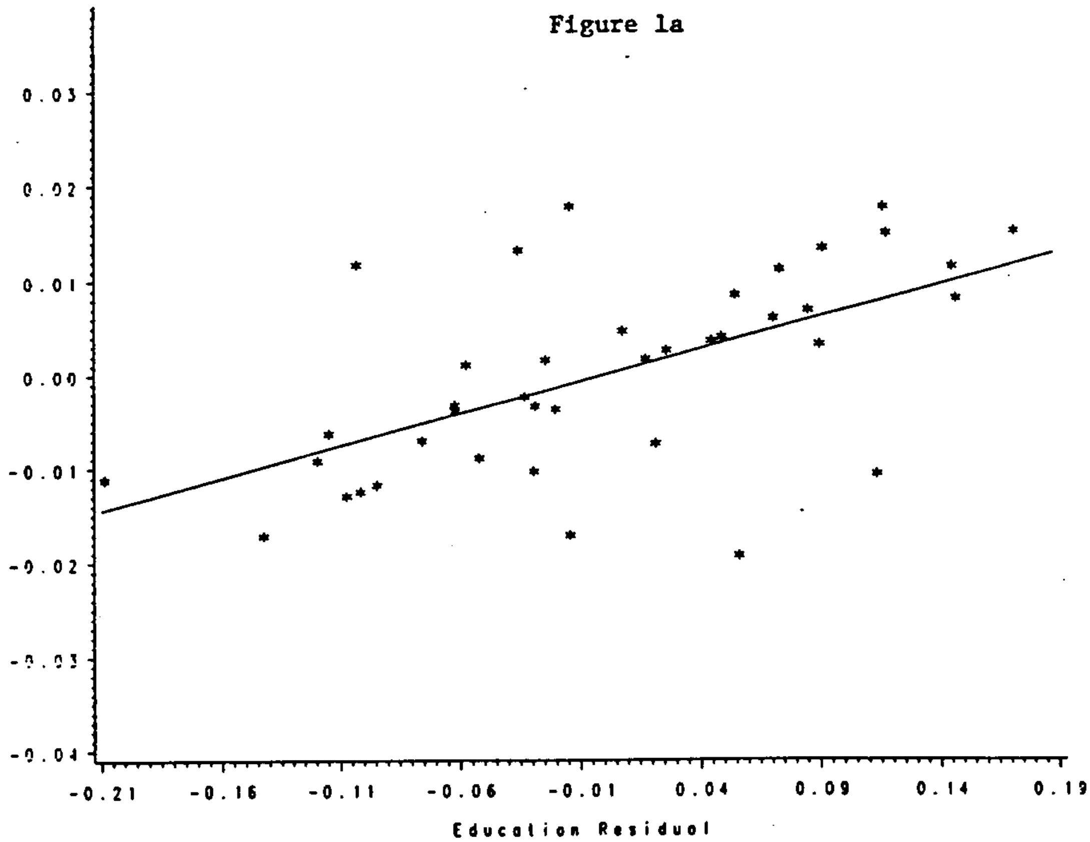
$$\left[\{(1/\phi) \begin{bmatrix} P^{-1}\phi + P^{-1}RR'P^{-1} & -P^{-1}R \\ -P^{-1}R' & 1 \end{bmatrix} \} + \{(1/\phi) \begin{bmatrix} P^{-1}q & 0 \\ 0 & 0 \end{bmatrix} \} \right]$$

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

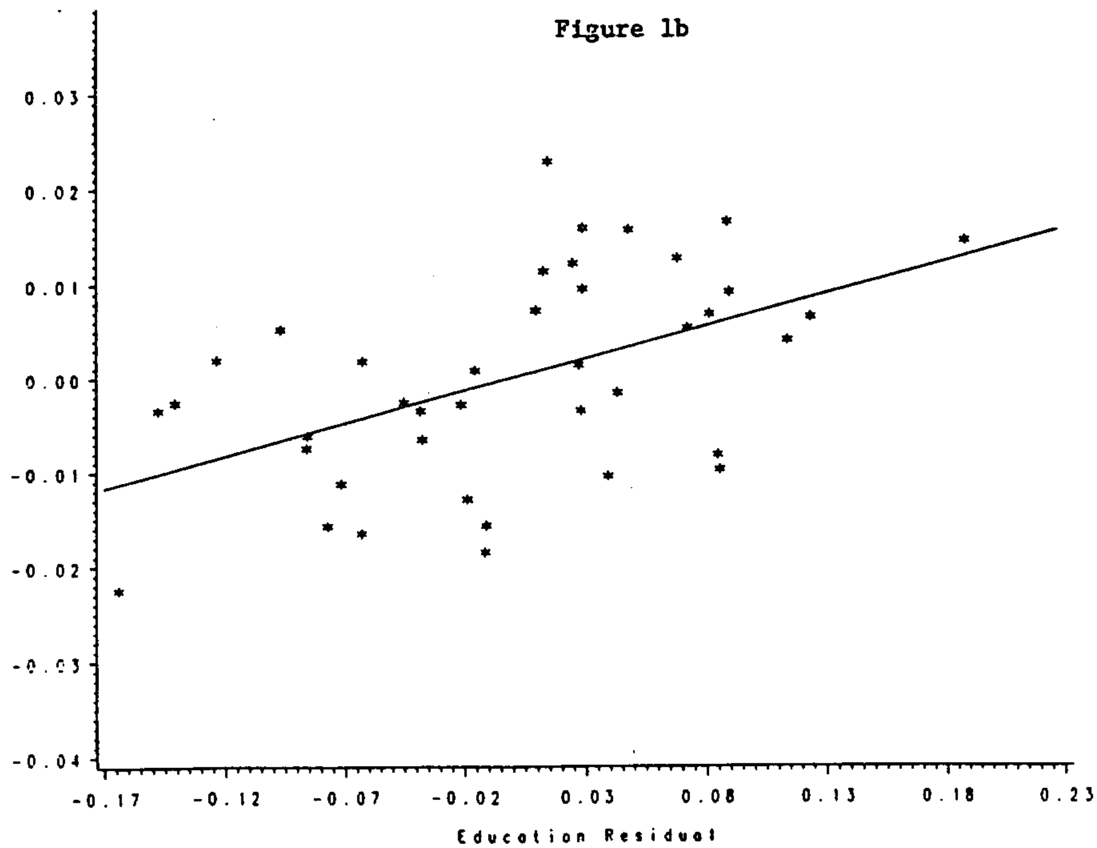
Therefore, $\{\pi' E(Z_i Z_i')\pi + c\sigma_\eta^2 L_1\}^{-1}\{\pi' E(Z_i Z_i')\pi\}$

$$= [\phi/(\phi + q)]I_{p+1} + [q/(\phi + q)] \begin{bmatrix} I_p & P^{-1}R \\ 0 & 0 \end{bmatrix} .$$

Multiplying this times β gives equation (6) in the text. Because $\{\pi' E(Z_i Z_i')\pi\}$ is positive definite, $[\pi' E(Z_i Z_i')\pi]^{-1}$ must also be positive definite. Therefore, $1/\phi \equiv 1/[Q - R'P^{-1}R]$, which is the lower right diagonal element of $[\pi' E(Z_i Z_i')\pi]^{-1}$, must be positive. This implies that the proportional bias in estimates of β_1, $[\phi/(\phi + q)] = [\phi/(\phi + c\sigma_\eta^2)]$, is between 0 and 1.



Men born 1920-29 in 1970 census
Wages and education by quarter of birth, residuals from year effects



Men born 1930-39 in 1980 census
Wages and education by quarter of birth, residuals from year effects

Table 1

Comparison of Estimates of Education Coefficient

from Angrist and Krueger (1991)

Parameter	Excluded Instruments: First Quarter (1)	Year*QOB (2)	State*QOB Year*QOB (3)
	1980 Census, Men Born 1930-39		
ß	.102 (.024)	.089 (.016)	.093 (.009)
First-stage F-statistic	67.94	4.75	2.43
	1970 Census, Men Born 1920-29		
ß	.072 (.022)	.077 (.015)	NA
First-stage F-statistic	65.60	4.54	NA
Number of excluded instruments	1	30	180

Source: Angrist and Krueger (1991); the estimates in column (1) are the Wald estimates from Table III, other estimates are from Tables IV (column 2), Table V (column 2), and Table VII (column 2). Column (1) includes no exogenous regressors, column (2) includes 9 year of birth dummies as exogenous regressors, and column (3) includes 9 year of birth and 50 state of birth dummies as exogenous regressors. Standard errors are shown in parentheses.

Table 2

Various Estimates of Model with 30 Quarter of Birth x Year of Birth Interactions used as Excluded Instruments

Parameter	Type of Estimator OLS (1)	IV (2)	SSIV (3)	USSIV (4)
	1980 Census, Men Born 1930-39			
ß	.063 (.0003)	.081 (.016)	.070 (.023)	.089 (.030)
θ	--	--	.780 (.118)	--
First-stage F	--	4.75	2.41	2.41
	1970 Census, Men Born 1920-29			
ß	.070 (.0004)	.069 (.015)	.059 (.023)	.063 (.024)
θ	--	--	.934 (.127)	--
First-stage F	--	4.54	2.03	2.03

Notes: 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 IV 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 1970 sample for OLS and IV 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.

Table 3

Various Estimates of Model with 30 Quarter of Birth x Year of Birth Interactions and 150 and Quarter of Birth x State of Birth Interactions used as Excluded Instruments

Parameter	Type of Estimator			
	OLS (1)	IV (2)	SSIV (3)	USSIV (4)
	1980 Census, Men Born 1930-39			
ß	.063	.083	.031	.076
	(.0003)	(.009)	(.011)	(.028)
θ	--	--	.408	--
			(.057)	
First-stage F	--	2.43	1.70	1.70

Notes: Models include 9 year of birth dummies, 50 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 IV 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 4

Results of 31 Monte Carlo Replications of Split

Summary Statistics for Education Coefficient

	Actual Instruments		Random Instruments	
	SSIV	USSIV	SSIV	USSIV
	(1)	(2)	(3)	(4)
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

Notes: Models include 9 year-of-birth dummies and 50 state-of-birth dummies as exogenous regressors. The conventional IV estimate and standard error with the random instruments is .057 (.014).

References

Altonji, Joseph, and Segal, "Small-Sample Bias in GMM Estimation of Covariance Structures," Mimeo., Northwestern University.

Angrist, Joshua, and Alan B. Krueger (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?," Quarterly Journal of Economics, 106, 979-1014.

Angrist, Joshua, and Alan B. Krueger (1992), "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples," JASA, 87, 328-336.

Bekker, Paul A. (1992), "Alternative Approximations to the Distributions of Instrumental Variables Estimators," Department of Econometrics Research Memorandum No. 512, University of Groningen.

Bound, John, David Jaeger, and Regina Baker, "The Cure Can Be Worse than the Disease: A Cautionary Tale Regarding Instrumental Variables," Cambridge, MA, NBER Technical Working Paper No. 137, 1993.

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

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

Fuller, Wayne A., "Regression Analysis for Sample Surveys," Sankhya: The Indian Journal of Statistics C37, 312-326.

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, XX.

Sargan, J.D. (1974), "The Validity of Nagar's Expansion for the Moments of Econometric Estimators," Econometrica 42(1), 169-176.

Staiger, D., and J.H. Stock (1993), "Asymptotics for Instrumental Variables Regressions with Weakly Correlated Instruments," Kennedy School of Government, mimeo, July.

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