Empirical Methods in Applied Economics

Jörn-Steffen Pischke
LSE

October 2007

1 Instrumental Variables

1.1 Basics

A good baseline for thinking about the estimation of causal effects is often the randomized experiment, like a clinical trial. However, for many treatments, even in an experiment it may not be possible to assign the treatment on a random basis. For example, in a drug trial it may only be possible to offer a drug but not to enforce that the patient actually takes the drug (a problem called non-compliance). When we study job training, we may only be able to randomly assign the offer of training. Individuals will then decide whether to participate in training or not. Even those not assigned training in the program under study may obtain the training elsewhere (a problem called non-embargo, training cannot be effectively witheld from the control group). Hence, treatment itself is still a behavioral variable, and only the *intention to treat* has been randomly assigned. The instrumental variables (IV) estimator is a useful tool to evaluate treatment in such a setup.

In order to see how this works, we will start by reviewing the assumptions necessary for the IV estimator to be valid in this case. We need some additional notation for this. Let $z = \{0, 1\}$ the intention to treat variable. $D = \{0, 1\}$ is the treatment again. We can now talk about counterfactual treatments: $D(z)$ is the treatment for the (counterfactual) value of z. I.e. $D(0)$ is the treatment if there was no intention to treat, and $D(1)$ is the treatment if there was an intention to treat. In the job training example, $D(0)$ denotes the training decision of an individual not assigned to the training program, and $D(1)$ denotes the training decision of someone assigned to the program. As before, Y is the outcome. The counterfactual outcome is now $Y(z, D)$ because it may depend on both the treatment choice and the treatment assignment. So there are four counterfactuals for Y.

We can now state three assumptions:

Assumption 1 z is as good as randomly assigned

Assumption 2 $Y(z, D) \; = Y(z', D) \; \forall z, z', D$

This assumption says that the counterfactual outcome only depends on D, and once you know D you do not need to know z. This is the *exclusion restriction*, and it implies that we can write $Y(z, D) = Y(D)$.

There are three causal effects we can define now:

1. The causal effect of z on D is $D(1) - D(0)$.
2. The causal effect of z on Y is $Y(1, D(1)) - Y(0, D(0))$. Given Assumption 2, we can write this as $Y(D(1)) - Y(D(0))$.
3. The causal effect of D on Y is $Y(1) - Y(0)$.

Effects 1. and 2. are *reduced form* effects. 1. is the *first stage* relationship, and 2. is the reduced form for the outcome. 3. is the treatment effect of ultimate interest. Without Assumption 2, it is not clear how to define this effect, beause the causal effect of D on Y would depend on z.

Assumption 3 $E(D(1) - D(0)) \neq 0$.

This assumption says that the variable z has some power to influence the treatment. Without it, z would be of no use to help us learn something about D. It is the existence of a significant first stage.

Assumption 1 is sufficient to estimate the reduced form causal effects 1. and 2. The exclusion restriction and the existence of a first stage are only necessary in order to give these reduced form effects an instrumental variables interpretation. Because of this, it is often useful to see estimates of the reduced forms as well as the IV results in an application.

In order to get some further insights into the workings of the IV estimator, start with the case where the instrument is binary and there are no other covariates. In this case, the IV estimator takes a particularly simple form. Notice that the IV estimator is given by

$$\widehat{\beta}_{IV} = \frac{cov(y_i, z_i)}{cov(D_i, z_i)}.$$

Since z_i is a dummy variable, the first covariance can be written as

$$\begin{aligned}
cov(y_i, z_i) &= E(y_i - \overline{y})(z_i - \overline{z}) \\
&= E y_i z_i - \overline{yz} \\
&= E(y_i|z_i = 1)E(z_i = 1) - \overline{y}E(z_i = 1) \\
&= \{E(y_i|z_i = 1) - \overline{y}\}\, E(z_i = 1) \\
&= \{E(y_i|z_i = 1) - [E(y_i|z_i = 1)E(z_i = 1) + E(y_i|z_i = 0)E(z_i = 0)]\}\, E(z_i = 1) \\
&= \{E(y_i|z_i = 1)E(z_i = 0) - E(y_i|z_i = 0)E(z_i = 0)\}\, E(z_i = 1) \\
&= \{E(y_i|z_i = 1) - E(y_i|z_i = 0)\}\, E(z_i = 1)E(z_i = 0).
\end{aligned}$$

A similar derivation for the denominator leads to

$$\begin{aligned}
\widehat{\beta}_{IV} &= \frac{cov(y_i, Z_i)}{cov(D_i, Z_i)} \\
&= \frac{\{E(y_i|z_i = 1) - E(y_i|z_i = 0)\}\, E(z_i = 1)E(z_i = 0)}{\{E(D_i|z_i = 1) - E(D_i|z_i = 0)\}\, E(z_i = 1)E(z_i = 0)} \\
&= \frac{E(y_i|z_i = 1) - E(y_i|z_i = 0)}{E(D_i|z_i = 1) - E(D_i|z_i = 0)} \\
&= \frac{E(y_i|z_i = 1) - E(y_i|z_i = 0)}{P(D_i = 1|z_i = 1) - P(D_i = 1|z_i = 0)}.
\end{aligned}$$

This formulation of the IV estimator is often referred to as the Wald-estimator. It says that the estimate is given by the difference in outcomes for the groups intended and not intended for treatment divided by the difference in actual treatment for these groups. It is also easy to see that the numerator is the reduced form estimate, also frequently called the intention to treat estimate. The denominator is the first stage estimate. Hence, the Wald-estimator is also the indirect least squares estimator, dividing the reduced form estimate by the first-stage estimate. This has to be true in the just identificed case.

The IV methodology is often useful in actual randomized experiments when the treatment itself cannot be randomly assigned because of the non-compliance and the lack of embargo problems. For example, in the Moving to Opportunity exeriment (Kling et al., 2004), poor households were given housing vouchers to move out of high poverty neighborhood. While the voucher receipt was randomly assigned, whether the household actually ended up moving is not under the control of the experimenter: some households assigned a voucher do not move, but some not assigned a voucher move on their own. Kling et al. (2004) therefore report both estimates of the reduced form effect (*intention to treat estimates* or ITT) and IV estimates (*treatment on the treated* or TOT). Since about half the households with vouchers actually moved, TOT estimates are about twice the size of the ITT estimates. Notice that the ITT estimates are of independent interest. They estimate directly the actual effect of the policy. Nevertheless, the TOT estimates are often of more interest in terms of their economic interpretation.

If an instrument is truely as good as randomly assigned, then IV estimation of the binary model

$$y_i = \alpha + \beta D_i + \varepsilon_i$$

will be sufficient. Often, this assumption is not going to be satisfied. However, an instrument may be as good as randomly assigned conditional on some covariates, so that we can estimate instead

$$y_i = \alpha + \beta D_i + X_i \delta + \varepsilon_i,$$

instrumenting D_i by z_i. The role of the covariates here is to ensure the validity of the IV assumptions. Of course, covariates orthogonal to z_i may also be included simply to reduce the variance of the estimate.

An interesting and controversial example of an IV study is the paper by Angrist and Krueger (1991). They try to estimate the returns to education. The concern is that there may be an omitted variable ("ability") which confounds the OLS estimates. The Angrist and Krueger insight is that US compulsory schooling laws can be used to construct an instrument for the number of years of schooling. US laws specify an age an individual has to reach before being able to drop out of school. This feature, together with the fact that there is only one date of school entry per year means that season of birth affects the length of schooling for dropouts.

Suppose, for example, that there are two individuals, Bob and Ron. Bob is born in January, and Bob is born in December of the previous year. So they are almost equal in age. School entry rules typically specify that you are allowed to enter school in summer, if you turned 6 during the previous calendar year. This means that Ron, who turns 6 in December, is allowed to enter school at age 6. Bob will not satisfy this rule and therefore has to wait an additional year and enter when he is 7. At the time of school entry, Bob is 11 months older than Ron. Both can drop out when they reach age 16. Of course, at that age, Bob will have completed 11 months less schooling than Ron, who entered earlier. The situation is illustrated in the following figure.

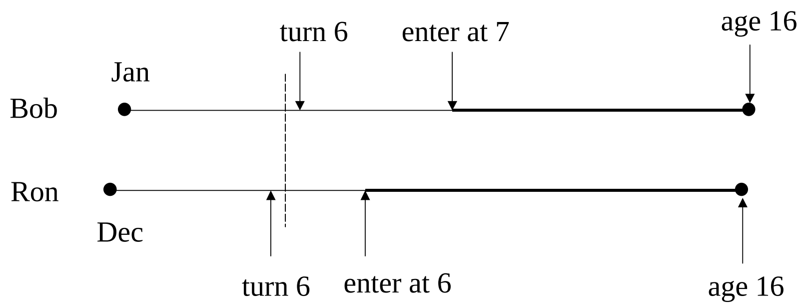

Figure 1: Schooling for Bob and Ron

The idea of the Angrist and Krueger (1991) paper is to use season of birth as an instrument for schooling. So in this application, we have

$$\begin{aligned} y &= \text{log earnings} \\ D &= \text{years of schooling} \\ z &= \text{born in the 1st quarter.} \end{aligned}$$

The first thing to do is to check the three IV assumptions. Assumption 1 (random assignment) is probably close to satisfied. Births are almost uniformly spaced over the year. There is relatively little parental choice over season of birth although there is clearly some. There is some evidence of small differences in the socioeconomic status of parents by season of birth of the child. Assumption 2, the exclusion restriction says that season of birth does not affect earnings directly, only through its effect on schooling. If you are born early in the year you enter school later (like Bob in the example). Hence, age at school entry should not be correlated with earnings. There is some psychology evidence that those who start later do better in school. If this translates into unobserved factors (i.e. anything other than how long you stay in school) which are correlated with earnings, then this will lead to a downward bias in the estimates. Assumption 3, the existence of a first stage is an empirical matter, which we can check in the data.

Figures 1 and 2 in Angrist and Krueger plot years of completed education against cohort of birth (in quarters) for those born between 1930 and 1950. The figures reveal that average education tends to be higher for those born later in the year (quarters 3 and 4) and relatively low for those born in the first quarter, as we would expect. The pattern is more pronounced for the early cohorts, and starts to vanish for the cohorts born in the late 1940s, when average education is higher. This is consisent with the idea that compulsory schooling laws are at the root of this pattern, since fewer and fewer students drop out at the earliest possible date over time.

Table 1 presents numerical estimates of this relationship. It reveals that those born in the first quarter have about 0.1 years less schooling than those born in the fourth quarter, with a slightly weaker relationship for the cohorts born in the 1940s. There is a small effect on high school graduation rates but basically no effect on college graduation or post-graduate education.[1] This pattern suggests that schooling is affected basically only for those with very little schooling. Table 2 compares the quarter of birth effects on school

[1] They also check for the effect on years of education for those with at least a high school degree. This is not really valid, because the conditioning is on an outcome variable (graduating from high school), which they showed to be affected by the instrument.

enrollment at age 16 for those in states with a compulsory schooling age of 16 with states with a higher compolsory schooling age. The enrollment effect is clearly visible for the age 16 states, but not for the states with later dropout ages. The enrollment effect is also concentrated on earlier cohorts when age 16 enrollment rates were lower. This is suggestive that compulsory schooling laws are indeed at the root of the quarter of birth-schooling relationship.

Figure 5 in the paper shows a similar picture to figures 1 and 2, but for earnings. There is again a saw-tooth pattern in quarter of birth, with earnings being higher for those born later in the year. This is consistent with the pattern in schooling and a postive return to schooling. One problem apparent in figure 5 is that age is clearly related to earnings beyond the quarter of birth effect, particularly for those in later cohorts (i.e. for those who are younger at the time of the survey in 1980). It is therefore important to control for this age earnings profile in the estimation, or to restrict the estimates to the cohorts on the relatively flat portion of the age-earnings profile in order to avoid confounding the quarter of birth effect with earnings growth with age. Hence, the exclusion restriction only holds conditional on other covariates in this case, namely adequate controls for the age-earnings profile.

Table 3 presents simple Wald estimates of the returns to schooling estimate for the cohorts in the relatively flat part of their life-cycle profile. Those born in the 1st quarter have about 0.1 fewer years of schooling compared to those born in the remaining quarters of the year. They also have about 1 percent lower earnings. This translates into a Wald estimate of about 0.07 for the 1920s cohort, and 0.10 for the 1930s cohort, compared to OLS estimates of around 0.07 to 0.08.

Tables 4 to 6 present instrumental variables estimates using a full set of quarter of birth times year of birth interactions as instruments. These specifications also control for additional covariates. It can be seen by comparing columns (2) and (4) in the tables that controlling for age is quite important. The IV returns to schooling are either similar to the OLS estimates or above. Finally, table 7 exploits the fact that different states have different compulsory schooling laws and uses 50 state of birth times quarter of birth interactions as instruments in addition to the year of birth times quarter of birth interactions, also controlling for state of birth now. The resulting IV estimates tend to be a bit higher than those in table 5, and the estimates are much more precise.

1.2 The Bias of 2SLS

It is a fortunate fact that the OLS estimator is not just a consistent estimator of the population regression function but also unbiased in small samples. With many other estimators we don't necessarily have such luck, and the 2SLS is no exception: 2SLS is biased in small samples even when the conditions for the consistency of the 2SLS estimator are met. For many years, applied researchers have lived with this fact without the loss of too much sleep. A series of papers in the early 1990s have changed this (Nelson and Startz, 1990a,b and Bound, Jaeger, and Baker, 1995). These papers have pointed out that point estimates and inference may be seriously misleading in cases relevant for empirical practice. A huge literature has since clarified the circumstances when we should worry about this bias, and what can be done about it.

The basic results from this literature are that the 2SLS estimator is biased when the instruments are "weak," meaning the correlation with endogenous regressors is low, and when there are many overidentifying restrictions. In these cases, the 2SLS estimator will be biased towards the OLS estimate. The intuition for these results is easy to see. Suppose you start with a single valid instrument (i.e. one which is as good as randomly assigned and which obeys the exclusion restriction). Now you add more and more instruments to the (overidentified) IV model. As the number of instruments goes to n, the sample size, the first stage R^2 goes to 1, and hence $\widehat{\beta}_{IV} \to \widehat{\beta}_{OLS}$. In any small sample, even a valid instrument will pick up some small amounts of the endogenous variation in x. Adding more and more instruments, the amount of random, and hence endogenous, variation in x will become more and more important. So IV will be more and more biased towards OLS.

It is also easy to see this formally. For simplicity, consider the case of a single endogenous regressor x, and no other covariates:

$$y = \beta x + \varepsilon$$

and write the first stage as

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

The 2SLS estimator is

$$\widehat{\beta}_{2SLS} = \left(x' P_Z x\right)^{-1} x' P_Z y = \beta + \left(x' P_Z x\right)^{-1} x' P_Z \varepsilon.$$

Using the fact that $P_Z x = Z\pi + P_Z \eta$ and substituting in the first stage we get

$$\widehat{\beta}_{2SLS} - \beta = \left(\pi' Z' Z \pi + \eta' P_Z \eta\right)^{-1} \left(\pi' Z' + \eta'\right) P_Z \varepsilon.$$

To get the small sample bias, we would like to take expectations of this expression. But the expectation of a ratio is not equal to the ratio of the two expectations. A useful approximation to the small sample bias of the estimator is obtained using a group asymptotic argument as suggested by Bekker (1994). In this asymptotic sequence, we let the number instruments go to infinity as we let the number of observations to to infinity but keep the number of observations per instrument constant. This captures the fact that we might have many instruments given the number of observations but it still allows us to use asymptotic theory (see also Angrist and Krueger, 1995). Group assymptotics essentially lets us take the expectations inside the ratio, i.e.

$$\text{plim}\widehat{\beta}_{2SLS} - \beta = \left[E\left(\pi' Z' Z\pi + \eta' P_Z \eta\right)\right]^{-1} E\left(\pi' Z'\varepsilon + \eta' P_Z \varepsilon\right).$$

Further notice that $E\left(Z'\varepsilon\right) = 0$ so that

$$\text{plim}\widehat{\beta}_{2SLS} - \beta = \left[E\left(\pi' Z' Z\pi\right) + \eta' P_Z \eta\right]^{-1} E\left(\eta' P_Z \varepsilon\right). \tag{1}$$

Working out the expectations yields

$$\text{plim}\widehat{\beta}_{2SLS} - \beta = \frac{\sigma_{\varepsilon\eta}}{\sigma_\eta^2}\left[\frac{E\left(\pi' Z' Z\pi\right)/p}{\sigma_\eta^2} + 1\right]^{-1}$$

where p is the number of instruments (see the appendix for a derivation). Notice that the term $(1/\sigma_\eta^2)E\left(\pi' Z' Z\pi\right)/p$ is the population F-statistic for the joint signifcance of all regressors in the first stage regression, call it F, so that

$$\text{plim}\widehat{\beta}_{2SLS} - \beta = \frac{\sigma_{\varepsilon\eta}}{\sigma_\eta^2}\frac{1}{F+1}. \tag{2}$$

Various things are immediately obvious from inspection of (2). First, as the first stage F-statistic gets small, the bias approaches $\sigma_{\varepsilon\eta}/\ \sigma_\eta^2$. Notice that the bias of the OLS estimator is $\sigma_{\varepsilon\eta}/\ \sigma_x^2$, which is equal to $\sigma_{\varepsilon\eta}/\ \sigma_\eta^2$ if $\pi = 0$, i.e. if the instruments are completely uncorrelated with the endogenous regressor. Hence, not surprisingly, unidentified 2SLS will have the same bias as OLS. So the first thing we see is that weakly identified 2SLS will be biased towards OLS.

Now turn things around, and consider the case where F gets large, i.e. the first stage becomes stronger. The 2SLS bias will vanish in the case. So the bias is related to the strength of the first stage as measured by the F-statistic on the excluded instruments.

Finally, look at the F-statistic. The numerator is divided by the degrees of freedom of the test, which is p, the number of excluded instruments. Consider adding completely useless instruments to your 2SLS model. $E(\pi' Z' Z\pi)$ will not change, since the new πs are all zero. Hence, σ_η^2 will also stay the same. So all that changes is that p goes up. The F-statistic becomes smaller as a result. And hence we learn that adding useless or weak instruments will make the bias worse.

Where does the bias of 2SLS come from? In order to get some intution on this, look at (1). If we knew the first stage coefficients π, our predicted values from the first stage would be $\widehat{x}_{true} = Z\pi$. In practice, we do not know π, but have to estimate it (implicitly) in the first stage. Hence, we use $\widehat{x} = P_Z x = Z\pi + P_Z\eta$, which differs from $\widehat{x}_{true}$ by the term $P_Z\eta$. So if we knew π, the numerator in (1) would disappear, and there would be no bias. The bias arises from the fact that we estimate π from the same sample as the second stage. The $P_Z\eta$ term appears twice in (1): Namely in $E(\eta' P_Z\eta) = p\sigma_\eta^2$ and in $E(\eta' P_Z\varepsilon) = p\sigma_{\varepsilon\eta}$. These two expectations give rise to the ratio $\sigma_{\varepsilon\eta}/\ \sigma_\eta^2$ in the formula for the bias (2). So some of the bias in OLS seeps into our 2SLS estimates through the sampling variability in $\widehat{\pi}$.

It is important to note that (2) only applies to overidentified models. Even if $\pi = 0$, i.e. the model is truely unidentified, the expression (2) exists, and is equal to $\sigma_{\varepsilon\eta}/\ \sigma_\eta^2$. For the just identified IV estimator this is not the case. The just identified IV estimator has no moments. With weak instruements, IV is median unbiased, even in small samples, so the weak instrument problem is only one of overidentified models. Just idenitfied IV will tend be very unstable when the instrument is weak because you are almost dividing by zero (the weak covariance of the regressor and the instrument). As a result, the sampling distribution tends to blow up as the instrument becomes weaker.

Fortunately, there are other estimators for overidentified 2SLS which are (approximatedly) unbiased in small samples, even with weak instruments. The LIML estimator has this property, and seems to deliver desirable results in most applications. Other estimators, which have been suggested, like split sample IV (Angrist and Krueger, 1995) or JIVE (Phillips and Hale, 19??; Angrist, Imbens, and Krueger, 1999) are also unbiased but do not perform any better than LIML. In order to illustrate some of the results from the discussion above, the following figures show some Monte Carlo results for various estimators. The simulated data are drawn from the following model

$$
\begin{aligned}
y_i &= \beta x_i + \varepsilon_i \\
x_i &= \sum_{j=1}^{p} \pi_j z_{ij} + \eta_i
\end{aligned}
$$

with $\beta = 1$, $\pi_1 = 0.1$, $\pi_j = 0 \; \forall j > 1$,

$$
\left(\begin{array}{c} \varepsilon_i \\ \eta_i \end{array} \right) \Bigg| Z \sim N \left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} 1 & 0.8 \\ 0.8 & 1 \end{array} \right) \right),
$$

and the z_{ij} are independent, normally distributed random variables with mean zero and unit variance, and there are 1000 observations in each sample.

Figure 2 shows the empirical distribution functions of four estimators: OLS, just identified IV, i.e. $p = 1$ (denoted IV), 2SLS with two instruments, i.e. $p = 2$ (denoted 2SLS), and LIML with two instruments. It is easy to see that the OLS estimator is biased and centerd around a value of β of about 1.79. IV is centered around the true value of β of 1. 2SLS with one weak and one uninformative instrument has some bias towards OLS (the median of 2SLS in the sampling experiment is 1.07). The distribution function for LIML is basically indistinguishable from that for just identified IV. Even though the LIML results are for the case where there is an additional uninformative instrument, LIML effectively discounts this information and delivers the same results as just identified IV in this case.

Figure 3 shows the case where we set $p = 20$, i.e. in addition to the one informative but weak instrument we add 19 instruments with no correlation with the endogenous regressor. We again show OLS, 2SLS, LIML results. It is easy to see that the bias in 2SLS is much worse now (the median is 1.53), and that the sampling distribution of the 2SLS estimator is quite tight. LIML continues to perform well and is centered around the true value of β of 1, with a slightly larger dispersion than with 2 instruments.

Finally, Figure 4 shows the case where the model is truely unidentified. We again choose 20 instruments but we set $\pi_j = 0$, $j = 1, ..., 20$. Not surprisingly, all estimators are now centered around the same value as OLS. However, we see that the sampling distribution of 2SLS is quite tight while the sampling distribution of LIML becomes very wide. Hence, the LIML standard errors are going to reflect the fact that we have no identifying information.

In terms of prescriptions for the applied researcher, the literature on the small sample bias of 2SLS and the above discussion yields the following results:

1. Report the first stage of your model. The first check is whether the instruments predict the endogenous regressor in the hypothesized way (and sometimes the first stage results are of independent substantive interest). Report the F-statistic on the excluded instruments. Stock, Wright, and Yogo (2002), in a nice survey of these issues, suggest that F-statistics above 10 or 20 are necessary to rule out weak instruments. The p-value on the F-statistic is rather meaningless in this context.

2. Look at the reduced form. There are no small issues with the reduced form: it's OLS so it is unbiased. If there is no effect in the reduced form, or if the estimates are excessively variable, there is going to be no effect in the IV either. And if there is you likely have an overidentified model with a weak instruments probelm.

3. In the overidentified case, if your first stage F-statistics suggest that your instruments may be weak, check the standard 2SLS results with an alternative estimator like LIML. If the results are different, and/or if LIML standard errors are much higher than those of 2SLS, this may indicate weak instruments. Experiment with different subsets of the instruments. Are the results stable?

4. In the just identified case, the IV, LIML, and other related estimators are all the same. So there are no alternatives to check. However, remember that just identified IV is median unbiased, even in small samples.

5. Even if your point estimates seem reliable (sensible reduced form, similar results from different instrument sets, from 2SLS and LIML) but your instruments are on the weak side your standard errors might be biased (downwards, of course). Standard errors based on the Bekker (1994) approximation tend to be more reliable (not implemented anywhere to date).

Bound, Jaeger, and Baker (1995) show that these small sample issues are a real concern in the Angrist and Krueger case, despite the fact that the regressions are being run with 300,000 or more observations. "Small sample" is always a relative concept. Bound et al. show that the IV estimates in the Angrist and Krueger specification move closer to the OLS estimates as more control variables are included, and hence as the first stage F-statistic shrinks (tables 1 and 2). They then go on and completely make up quarters of birth, using a random number generator. Their table 3 shows that the results from this exercise are not very different from their IV estimates reported before. Maybe the most worrying fact is that the standard errors from the random instruments are not much higher than those in the real IV regressions.

In some applications there is more than one endogenous variable, and hence a set of instruments has to predict these multiple endogenous variables. The weak instruments problem can no longer be assessed simply by looking at the F-statistic for each first stage equation alone. For example, consider the case of two endogenous variables and two instruments. Suppose instrument 1 is strong and predicts both endogenous variables well. This will yield high F-statistics in each of the two first stage equations. Nevertheless, the model is underidentified because $\widehat{x}_1$ and $\widehat{x}_2$ will be closely correlated now. With two instruments it is necessary for one to predict the first endogenous variable, and the second the second. In order to assess whether the instruments are weak or strong, it is necessary to look at a matrix version of the F-statistic, which assesses all the first stage equations at once. This is called the Cragg-Donald or minimum eigenvalue statistic . References can be found in Stock, Wright, and Yogo (2002), the statistic is implemented in Stata 10.

1.3 Appendix

Start from equation

$$\begin{aligned}
\text{plim}\widehat{\beta}_{2SLS} - \beta &= \left[E\left(\pi' Z' Z \pi\right) + \eta' P_Z \eta\right]^{-1} E\left(\eta' P_Z \varepsilon\right) \\
&= \left[E\left(\pi' Z' Z \pi\right) + E\left(\eta' P_Z \eta\right)\right]^{-1} E\left(\eta' P_Z \varepsilon\right).
\end{aligned}$$

in the text. In the one regressor case $\eta' P_Z \eta$ is simply a scalar so that

$$\begin{aligned}
E\left(\eta' P_Z \eta\right) &= E\left[\text{tr}\left(\eta' P_Z \eta\right)\right] \\
&= E\left[\text{tr}\left(P_Z \eta \eta'\right)\right] \\
&= \text{tr}\left(P_Z E\left[\eta \eta'\right]\right) \\
&= \text{tr}\left(P_Z \sigma_\eta^2 I\right) \\
&= \sigma_\eta^2 \text{tr}\left(P_Z\right) \\
&= \sigma_\eta^2 p
\end{aligned}$$

where p is the number of instruments, and we have assumed that the ηs are homoskedastic. Applying the trace trick to $\eta' P_Z \varepsilon$ again we can write

$$\begin{aligned}
\text{plim}\widehat{\beta}_{2SLS} - \beta &= \left[E\left(\pi' Z' Z \pi\right) + \sigma_\eta^2 p\right]^{-1} E\left[\text{tr}\left(\eta' P_Z \varepsilon\right)\right] \\
&= \left[E\left(\pi' Z' Z \pi\right) + \sigma_\eta^2 p\right]^{-1} E\left[\text{tr}\left(P_Z \varepsilon \eta'\right)\right] \\
&= \sigma_{\varepsilon\eta} p \left[E\left(\pi' Z' Z \pi\right) + \sigma_\eta^2 p\right]^{-1} \\
&= \frac{\sigma_{\varepsilon\eta}}{\sigma_\eta^2} \left[\frac{E\left(\pi' Z' Z \pi\right)/p}{\sigma_\eta^2} + 1\right]^{-1}
\end{aligned}$$

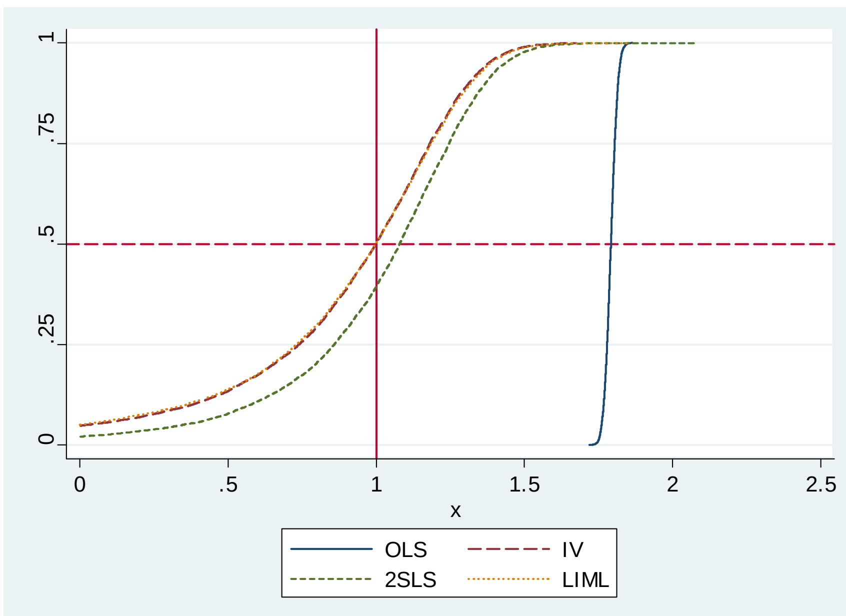

Figure 2: Distribution of the OLS, IV, and 2SLS and LIML estimators

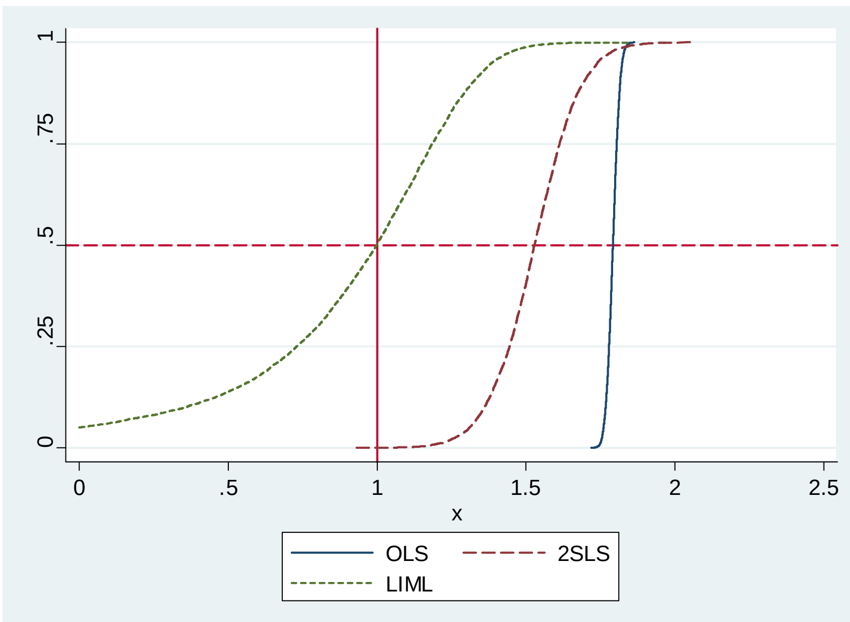

Figure 3: OLS, 2SLS, and LIML estimators with 20 instruments

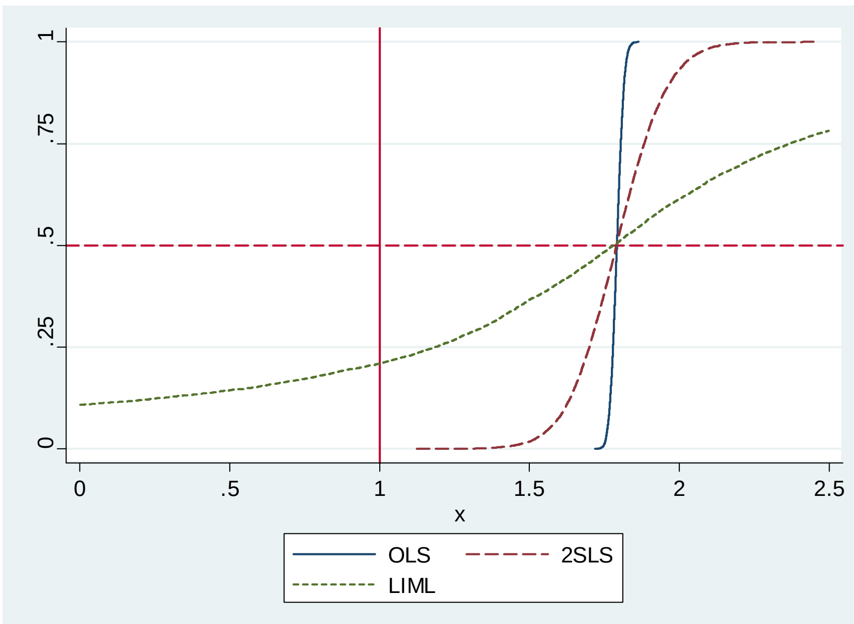

Figure 4: OLS, 2SLS, and LIML estimators with 20 uncorrelated instruments

Angrist and Krueger 1991: Figures 1 and 2

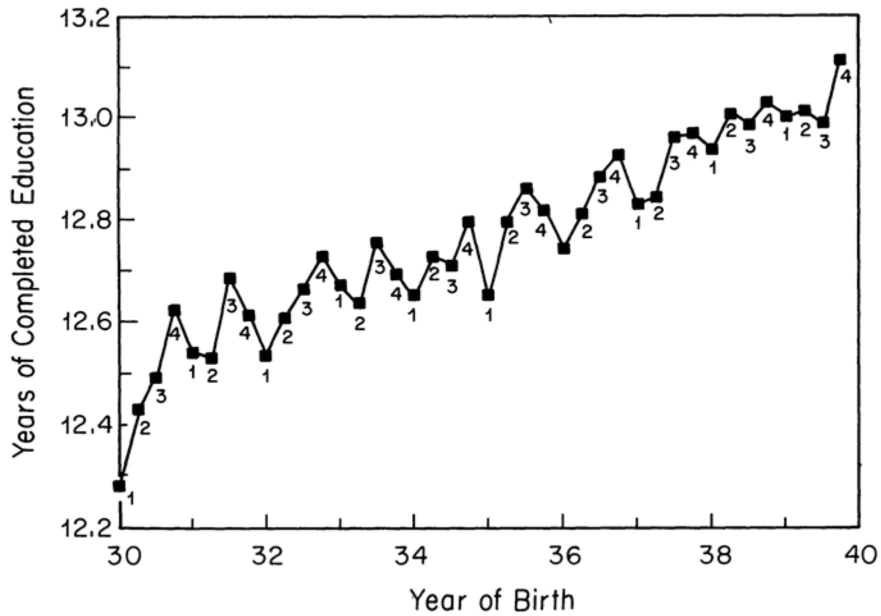

FIGURE I
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

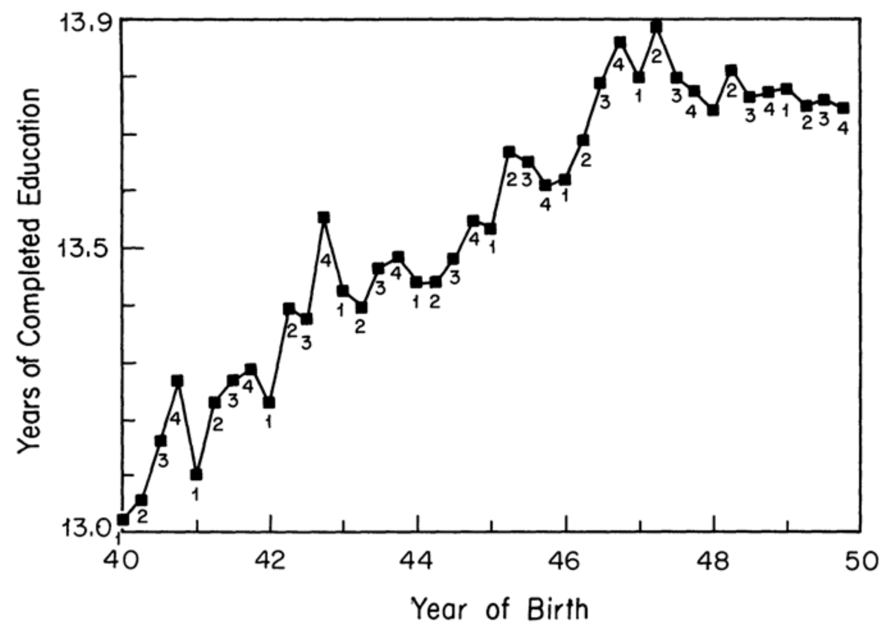

FIGURE II
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

Angrist and Krueger 1991: Table 1

TABLE I
THE EFFECT OF QUARTER OF BIRTH ON VARIOUS EDUCATIONAL OUTCOME VARIABLES

Outcome variable	Birth cohort	Mean	Quarter-of-birth effect[a] I	II	III	F-test[b] [P-value]
Total years of education	1930–1939	12.79	−0.124	−0.086	−0.015	24.9
			(0.017)	(0.017)	(0.016)	[0.0001]
	1940–1949	13.56	−0.085	−0.035	−0.017	18.6
			(0.012)	(0.012)	(0.011)	[0.0001]
High school graduate	1930–1939	0.77	−0.019	−0.020	−0.004	46.4
			(0.002)	(0.002)	(0.002)	[0.0001]
	1940–1949	0.86	−0.015	−0.012	−0.002	54.4
			(0.001)	(0.001)	(0.001)	[0.0001]
Years of educ. for high school graduates	1930–1939	13.99	−0.004	0.051	0.012	5.9
			(0.014)	(0.014)	(0.014)	[0.0006]
	1940–1949	14.28	0.005	0.043	−0.003	7.8
			(0.011)	(0.011)	(0.010)	[0.0017]
College graduate	1930–1939	0.24	−0.005	0.003	0.002	5.0
			(0.002)	(0.002)	(0.002)	[0.0021]
	1940–1949	0.30	−0.003	0.004	0.000	5.0
			(0.002)	(0.002)	(0.002)	[0.0018]
Completed master's degree	1930–1939	0.09	−0.001	0.002	−0.001	1.7
			(0.001)	(0.001)	(0.001)	[0.1599]
	1940–1949	0.11	0.000	0.004	0.001	3.9
			(0.001)	(0.001)	(0.001)	[0.0091]
Completed doctoral degree	1930–1939	0.03	0.002	0.003	0.000	2.9
			(0.001)	(0.001)	(0.001)	[0.0332]
	1940–1949	0.04	−0.002	0.001	−0.001	4.3
			(0.001)	(0.001)	(0.001)	[0.0050]

a. Standard errors are in parentheses. An $MA(+2, -2)$ trend term was subtracted from each dependent variable. The data set contains men from the 1980 Census, 5 percent Public Use Sample. Sample size is 312,718 for 1930–1939 cohort and is 457,181 for 1940–1949 cohort.

b. F-statistic is for a test of the hypothesis that the quarter-of-birth dummies jointly have no effect.

Angrist and Krueger 1991: Table 2

TABLE II

PERCENTAGE OF AGE GROUP ENROLLED IN SCHOOL BY BIRTHDAY AND LEGAL DROPOUT AGE[a]

Date of birth	Type of state law[b]: School-leaving age: 16 (1)	Type of state law[b]: School-leaving age: 17 or 18 (2)	Column (1) − (2)
	Percent enrolled April 1, 1960		
1. Jan 1–Mar 31, 1944	87.6	91.0	−3.4
(age 16)	(0.6)	(0.9)	(1.1)
2. Apr 1–Dec 31, 1944	92.1	91.6	0.5
(age 15)	(0.3)	(0.5)	(0.6)
3. Within-state diff.	−4.5	−0.6	−4.0
(row 1 − row 2)	(0.7)	(1.0)	(1.2)
	Percent enrolled April 1, 1970		
4. Jan 1–Mar 31, 1954	94.2	95.8	−1.6
(age 16)	(0.3)	(0.5)	(0.6)
5. Apr 1–Dec 31, 1954	96.1	95.7	0.4
(age 15)	(0.1)	(0.3)	(0.3)
6. Within-state diff.	−1.9	0.1	−2.0
(row 1 − row 2)	(0.3)	(0.6)	(0.6)
	Percent enrolled April 1, 1980		
7. Jan 1–Mar 31, 1964	95.0	96.2	−1.2
(age 16)	(0.1)	(0.2)	(0.2)
8. Apr 1–Dec 31, 1964	97.0	97.7	−0.7
(age 15)	(0.1)	(0.1)	(0.1)
9. Within-state diff.	−2.0	−1.5	0.5
(row 1 − row 2)	(0.1)	(0.2)	(0.3)

a. Standard errors are in parentheses.

b. Data set used to compute rows 1–3 is the 1960 Census, 1 percent Public Use Sample; data set used to compute rows 4–6 is 1970 Census, 1 percent State Public Use Sample (15 percent form); data set used to compute rows 7–9 is the 1980 Census, 5 percent Public Use Sample. Each sample contains both boys and girls. Sample sizes are 4,153 for row 1; 12,512 for row 2; 7,758 for row 4; 24,636 for row 5; 42,740 for row 7; and 131,020 for row 8.

Angrist and Krueger 1991: Figure 5

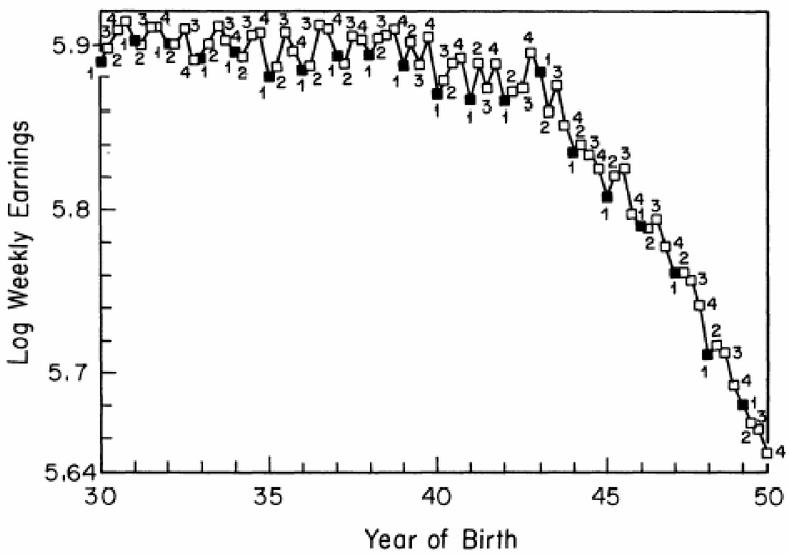

FIGURE V
Mean Log Weekly Wage, by Quarter of Birth
All Men Born 1930–1949; 1980 Census

Angrist and Krueger 1991: Table 3

TABLE III

Panel A: Wald Estimates for 1970 Census—Men Born 1920–1929[a]

	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year	(3) Difference (std. error) (1) − (2)
ln (wkly. wage)	5.1484	5.1574	−0.00898
			(0.00301)
Education	11.3996	11.5252	−0.1256
			(0.0155)
Wald est. of return to education			0.0715
			(0.0219)
OLS return to education[b]			0.0801
			(0.0004)

Panel B: Wald Estimates for 1980 Census—Men Born 1930–1939

	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year	(3) Difference (std. error) (1) − (2)
ln (wkly. wage)	5.8916	5.9027	−0.01110
			(0.00274)
Education	12.6881	12.7969	−0.1088
			(0.0132)
Wald est. of return to education			0.1020
			(0.0239)
OLS return to education			0.0709
			(0.0003)

a. The sample size is 247,199 in Panel A, and 327,509 in Panel B. Each sample consists of males born in the United States who had positive earnings in the year preceding the survey. The 1980 Census sample is drawn from the 5 percent sample, and the 1970 Census sample is from the State, County, and Neighborhoods 1 percent samples.

b. The OLS return to education was estimated from a bivariate regression of log weekly earnings on years of education.

Angrist and Krueger 1991: Table 4

TABLE IV
OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR MEN BORN 1920–1929: 1970 CENSUS[a]

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0802	0.0769	0.0802	0.1310	0.0701	0.0669	0.0701	0.1007
	(0.0004)	(0.0150)	(0.0004)	(0.0334)	(0.0004)	(0.0151)	(0.0004)	(0.0334)
Race (1 = black)	—	—	—	—	0.2980	−0.3055	−0.2980	−0.2271
					(0.0043)	(0.0353)	(0.0043)	(0.0776)
SMSA (1 = center city)	—	—	—	—	0.1343	0.1362	0.1343	0.1163
					(0.0026)	(0.0092)	(0.0026)	(0.0198)
Married (1 = married)	—	—	—	—	0.2928	0.2941	0.2928	0.2804
					(0.0037)	(0.0072)	(0.0037)	(0.0141)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region of residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
Age	—	—	0.1446	0.1409	—	—	0.1162	0.1170
			(0.0676)	(0.0704)			(0.0652)	(0.0662)
Age-squared	—	—	−0.0015	−0.0014	—	—	−0.0013	−0.0012
			(0.0007)	(0.0008)			(0.0007)	(0.0007)
χ^2 [dof]	—	36.0 [29]	—	25.6 [27]	—	34.2 [29]	—	28.8 [27]

a. Standard errors are in parentheses. Sample size is 247,199. Instruments are a full set of quarter-of-birth times year-of-birth interactions. The sample consists of males born in the United States. The sample is drawn from the State, County, and Neighborhoods 1 percent samples of the 1970 Census (15 percent form). The dependent variable is the log of weekly earnings. Age and age-squared are measured in quarters of years. Each equation also includes an intercept.

Angrist and Krueger 1991: Table 5

TABLE V
OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR MEN BORN 1930–1939: 1980 CENSUS[a]

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0711	0.0891	0.0711	0.0760	0.0632	0.0806	0.0632	0.0600
	(0.0003)	(0.0161)	(0.0003)	(0.0290)	(0.0003)	(0.0164)	(0.0003)	(0.0299)
Race (1 = black)	—	—	—	—	−0.2575	−0.2302	−0.2575	−0.2626
					(0.0040)	(0.0261)	(0.0040)	(0.0458)
SMSA (1 = center city)	—	—	—	—	0.1763	0.1581	0.1763	0.1797
					(0.0029)	(0.0174)	(0.0029)	(0.0305)
Married (1 = married)	—	—	—	—	0.2479	0.2440	0.2479	0.2486
					(0.0032)	(0.0049)	(0.0032)	(0.0073)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region-of-residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
Age	—	—	−0.0772	−0.0801	—	—	−0.0760	−0.0741
			(0.0621)	(0.0645)			(0.0604)	(0.0626)
Age-squared	—	—	0.0008	0.0008	—	—	0.0008	0.0007
			(0.0007)	(0.0007)			(0.0007)	(0.0007)
χ^2 [dof]	—	25.4 [29]	—	23.1 [27]	—	22.5 [29]	—	19.6 [27]

a. Standard errors are in parentheses. Sample size is 329,509. Instruments are a full set of quarter-of-birth times year-of-birth interactions. The sample consists of males born in the United States. The sample is drawn from the 5 percent sample of the 1980 Census. The dependent variable is the log of weekly earnings. Age and age-squared are measured in quarters of years. Each equation also includes an intercept.

Angrist and Krueger 1991: Table 6

TABLE VI

OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR MEN BORN 1940–1949: 1980 CENSUS[a]

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0573 (0.0003)	0.0553 (0.0138)	0.0573 (0.0003)	0.0948 (0.0223)	0.0520 (0.0003)	0.0393 (0.0145)	0.0521 (0.0003)	0.0779 (0.0239)
Race (1 = black)	—	—	—	—	−0.2107 (0.0032)	−0.2266 (0.0183)	−0.2108 (0.0032)	−0.1786 (0.0296)
SMSA (1 = center city)	—	—	—	—	0.1418 (0.0023)	0.1535 (0.0135)	0.1419 (0.0023)	0.1182 (0.0220)
Married (1 = married)	—	—	—	—	0.2445 (0.0022)	0.2442 (0.0022)	0.2444 (0.0022)	0.2450 (0.0023)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region-of-residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
Age	—	—	0.1800 (0.0389)	0.1325 (0.0486)	—	—	0.1518 (0.0379)	0.1215 (0.0474)
Age-squared	—	—	0.0023 (0.0006)	0.0016 (0.0007)	—	—	0.0019 (0.0005)	0.0015 (0.0007)
χ^2 [dof]	—	101.6 [29]	—	49.1 [27]	—	93.6 [29]	—	50.6 [27]

a. Standard errors are in parentheses. Sample size is 486,926. Instruments are a full set of quarter-of-birth times year-of-birth interactions. Sample consists of males born in the United States. The sample is drawn from the 5 percent samples of the 1980 Census. The dependent variable is the log of weekly earnings. Age and age-squared are measured in quarters of years. Each equation also includes an intercept.

Angrist and Krueger 1991: Table 7

TABLE VII

OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR MEN BORN 1930–1939: 1980 CENSUS[a]

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0673 (0.0003)	0.0928 (0.0093)	0.0673 (0.0003)	0.0907 (0.0107)	0.0628 (0.0003)	0.0831 (0.0095)	0.0628 (0.0003)	0.0811 (0.0109)
Race (1 = black)	—	—	—	—	−0.2547 (0.0043)	−0.2333 (0.0109)	−0.2547 (0.0043)	−0.2354 (0.0122)
SMSA (1 = center city)	—	—	—	—	0.1705 (0.0029)	0.1511 (0.0095)	0.1705 (0.0029)	0.1531 (0.0107)
Married (1 = married)	—	—	—	—	0.2487 (0.0032)	0.2435 (0.0040)	0.2487 (0.0032)	0.2441 (0.0042)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region-of-residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
50 State-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	—	—	−0.0757 (0.0617)	−0.0880 (0.0624)	—	—	−0.0778 (0.0603)	−0.0876 (0.0609)
Age-squared	—	—	0.0008 (0.0007)	0.0009 (0.0007)	—	—	0.0008 (0.0007)	0.0009 (0.0007)
χ^2 [dof]	—	163 [179]	—	161 [177]	—	164 [179]	—	162 [177]

a. Standard errors are in parentheses. Excluded instruments are 30 quarter-of-birth times year-of-birth dummies and 150 quarter-of-birth times state-of-birth interactions. Age and age-squared are measured in quarters of years. Each equation also includes an intercept term. The sample is the same as in Table VI. Sample size is 329,509.

Bound et al. 1995: Table 1

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

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

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509. All specifications include Race (1 = black), SMSA (1 = central city), Married (1 = married, living with spouse), and 8 Regional dummies as control variables. *F* (first stage) and partial R^2 are for the instruments in the first stage of IV estimation. *F* (overidentification) is that suggested by Basmann (1960).

Bound et al. 1995: Table 2

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

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

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509. All specifications include Race (1 = black), SMSA (1 = central city), Married (1 = married, living with spouse), 8 Regional dummies, and 50 State of Birth dummies as control variables. *F* (first stage) and partial R^2 are for the instruments in the first stage of IV estimation. *F* (overidentification) is that suggested by Basmann (1960).

Bound et al. 1995: Table 3

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

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

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

Monte Carlo Design

y = 1 + x*b + e   b = 1

Instrument vector z includes one instrument with various correlations with x and k garbage instruments (with no correlation with x).
All experiments use samples with 100 observations and no other regressors.

10000 replications
0 garbage instruments

	ols	2sls	2sls	2sls	2sls	2sls
corr of z and x		0.000	0.025	0.050	0.100	0.200
mean of b	1.707	6.803	-0.822	0.631	0.919	0.957
10th perc. of b	1.549	-1.938	-0.777	-0.281	0.227	0.509
25th perc. of b	1.623	0.497	0.405	0.512	0.658	0.763
median of b	1.708	1.735	1.117	1.020	0.999	0.997
75th perc. of b	1.791	2.959	1.634	1.400	1.286	1.202
90th perc. of b	1.866	5.597	2.288	1.722	1.516	1.363
median of std err	0.123	2.630	0.935	0.652	0.451	0.320
std dev of b	0.124	371.588	152.595	7.684	2.734	0.366
se of mean of b	0.001	3.716	1.526	0.077	0.027	0.004

10000 replications
1 garbage instruments

	ols	2sls	2sls	2sls	2sls	2sls
corr of z and x		0.000	0.025	0.050	0.100	0.200
mean of b	1.707	1.719	1.184	1.057	1.010	0.999
10th perc. of b	1.550	0.101	0.089	0.217	0.402	0.565
25th perc. of b	1.625	1.024	0.758	0.729	0.763	0.810
median of b	1.707	1.722	1.266	1.141	1.069	1.037
75th perc. of b	1.790	2.411	1.710	1.490	1.332	1.237
90th perc. of b	1.865	3.315	2.185	1.811	1.549	1.393
median of std err	0.123	1.232	0.746	0.574	0.424	0.307
std dev of b	0.123	2.938	1.631	1.104	0.685	0.354
se of mean of b	0.001	0.029	0.016	0.011	0.007	0.004

10000 replications
2 garbage instruments

	ols	2sls	2sls	2sls	2sls	2sls
corr of z and x		0.000	0.025	0.050	0.100	0.200
mean of b	1.707	1.730	1.334	1.197	1.081	1.038
10th perc. of b	1.553	0.561	0.434	0.483	0.528	0.625
25th perc. of b	1.626	1.169	0.936	0.879	0.842	0.856
median of b	1.707	1.712	1.368	1.242	1.129	1.066
75th perc. of b	1.788	2.253	1.757	1.553	1.378	1.258
90th perc. of b	1.864	2.889	2.175	1.868	1.591	1.417
median of std err	0.123	0.895	0.644	0.517	0.398	0.299
std dev of b	0.123	1.475	0.885	0.693	0.485	0.323
se of mean of b	0.001	0.015	0.009	0.007	0.005	0.003

5000 replications
4 garbage instruments

	ols	2sls	2sls	2sls	2sls	2sls
corr of z and x		0.000	0.025	0.050	0.100	0.200
mean of b	1.707	1.700	1.467	1.336	1.201	1.103
10th perc. of b	1.549	0.874	0.782	0.743	0.712	0.723
25th perc. of b	1.624	1.317	1.144	1.049	0.980	0.928
median of b	1.707	1.709	1.470	1.352	1.228	1.128
75th perc. of b	1.790	2.089	1.788	1.638	1.452	1.308
90th perc. of b	1.864	2.510	2.133	1.897	1.652	1.454
median of std err	0.123	0.628	0.511	0.445	0.362	0.283
std dev of b	0.123	0.718	0.589	0.487	0.393	0.299
se of mean of b	0.002	0.010	0.008	0.007	0.006	0.004

5000 replications
8 garbage instruments

	ols	2sls	2sls	2sls	2sls	2sls
corr of z and x		0.000	0.025	0.050	0.100	0.200
mean of b	1.706	1.700	1.561	1.458	1.327	1.209
10th perc. of b	1.543	1.116	1.042	0.978	0.908	0.863
25th perc. of b	1.621	1.407	1.291	1.212	1.118	1.040
median of b	1.707	1.699	1.568	1.454	1.346	1.223
75th perc. of b	1.790	1.993	1.822	1.710	1.552	1.391
90th perc. of b	1.868	2.270	2.069	1.943	1.731	1.534
median of std err	0.123	0.444	0.395	0.362	0.311	0.258
std dev of b	0.128	0.472	0.422	0.395	0.336	0.264
se of mean of b	0.002	0.007	0.006	0.006	0.005	0.004

2500 replications
16 garbage instruments

	ols	2sls	2sls	2sls	2sls	2sls
corr of z and x		0.000	0.025	0.050	0.100	0.200
mean of b	1.705	1.705	1.630	1.556	1.462	1.347
10th perc. of b	1.549	1.314	1.255	1.190	1.129	1.057
25th perc. of b	1.623	1.491	1.426	1.362	1.299	1.199
median of b	1.704	1.715	1.633	1.554	1.469	1.353
75th perc. of b	1.785	1.910	1.838	1.749	1.633	1.502
90th perc. of b	1.860	2.112	2.016	1.926	1.792	1.622
median of std err	0.123	0.310	0.292	0.278	0.253	0.224
std dev of b	0.124	0.317	0.306	0.292	0.264	0.223
se of mean of b	0.002	0.006	0.006	0.006	0.005	0.004

2500 replications
32 garbage instruments

	ols	2sls	2sls	2sls	2sls	2sls
corr of z and x		0.000	0.025	0.050	0.100	0.200
mean of b	1.701	1.703	1.665	1.630	1.583	1.496
10th perc. of b	1.539	1.422	1.387	1.365	1.316	1.245
25th perc. of b	1.618	1.555	1.518	1.483	1.444	1.370
median of b	1.702	1.705	1.668	1.626	1.585	1.501
75th perc. of b	1.787	1.855	1.808	1.770	1.718	1.629
90th perc. of b	1.859	1.978	1.938	1.900	1.845	1.735
median of std err	0.123	0.217	0.211	0.207	0.198	0.184
std dev of b	0.126	0.220	0.216	0.213	0.208	0.193
se of mean of b	0.003	0.004	0.004	0.004	0.004	0.004

Table 4.6.2
Alternative IV estimates of the economic returns to schooling

	(1)	(2)	(3)	(4)	(5)	(6)
2SLS	.105	.435	.089	.076	.093	.091
	(.020)	(.450)	(.016)	(.029)	(.009)	(.011)
LIML	.106	.539	.093	.081	.106	.110
	(.020)	(.627)	(.018)	(.041)	(.012)	(.015)
F-statistic (excluded instruments)	32.27	.42	4.91	1.61	2.58	1.97
Controls						
Year of birth	✓	✓	✓	✓	✓	✓
State of birth					✓	✓
Age, age squared		✓		✓		✓
Excluded instruments						
Quarter-of-birth dummies	✓	✓				
Quarter of birth*year of birth			✓	✓	✓	✓
Quarter of birth*state of birth					✓	✓
Number of excluded instruments	3	2	30	28	180	178

Notes: The table compares 2SLS and LIML estimates using alternative sets of instruments and controls. The age and age squared variables measure age in quarters. The OLS estimate corresponding to the models reported in columns 1–4 is .071; the OLS estimate corresponding to the models reported in columns 5 and 6 is .067. Data are from the Angrist and Krueger (1991) 1980 census sample. The sample size is 329,509. Standard errors are reported in parentheses.