Weak Instruments

"Weak instruments" raises a lot of issues. As weak instruments are quite common, this topic is of obvious importance, but there is no simple "right method" that covers all situations. So this here is a survey of surveys and you should be prepared to do further reading in many cases when you do your own research.

To get the basic intuition, consider the simplest linear model

$$y_i = \beta x_i + \sigma_u u_i \, ,$$

where we suppress the constant, and assume that x_i and u_i are correlated. (I here follow Davidson-MacKinnon, page 326.) Note, that if there are other regressors, we can think of this regression as the one we get after removing the other variables using Frisch-Waugh. Let us assume u_i is normally distributed with variance 1. We also assume that

$$x_i = \pi z_i + \sigma_v v_i \, ,$$

where v_i is normal with variance 1. We assume that the correlation of x with u occurs because the covariance between u and v is ρ, while z is a valid instrument for x if π is non-zero. We will use X, Z, and Y to denote a sample of N observations. We keep N fixed. (We know that IV is consistent.

Here we want to examine small-sample properties.)

The IV (2SLS) estimator is

$$\beta_{IV} = (Z'X)^{-1}Z'Y = (Z'X)^{-1}Z'(\beta X + \sigma_u U)$$

or

$$\beta_{IV} = \beta + \sigma_u (Z'X)^{-1}Z'U$$

or

$$\beta_{IV} - \beta = (Z'(\pi Z + \sigma_v V))^{-1}\sigma_u Z'U\,.$$

Now notice that we can assume $Z'Z = 1$. Why? Because you can always change the units in which you measure your instrument and in this case, it is convenient to re-scale them so the vector has length unity which cuts down on clutter (here we use that we consider the regression for a fixed N). We have

$$\beta_{IV} - \beta = \frac{\sigma_u Z'U}{\pi + \sigma_v Z'V}$$

Here, it is easy to see that the IV estimator is NOT unbiased because the OLS-bias is caused by U being correlated with V (in the OLS case, the denominator is not stochastic, so the expectation of the right-hand side is 0, because we have a bunch of constant stuff times $E\{U\}$).

It is convenient to write $u_i = \rho v_i + u_i^1$ using the formula for conditional normal expectations (so that ρv_i is the expectation of u_i conditional on v_i and u_i^1 is independent of v. We want to show that the expression for $\beta_{IV} - \beta$

can have crazy outliers, which we demonstrate by showing that it does not have a finite expected value. We use the law of iterated expectations and get

$$E\{\beta_{IV} - \beta\} = E\{E\{\beta_{IV} - \beta|V\}\} = E\frac{\sigma_u \rho\, Z'V}{\pi + \sigma_v Z'V} ,$$

where we have taken the expectation $E\{U^1|V\} = 0$ and now have only the expectation over V left. Now we will use our normalization. $Z'V$ is $\Sigma_{i=1}^N z_i v_i$ so it has mean 0 (all the terms have mean zero) and variance $\Sigma_{i=1}^N z_i^2 = 1$ (or course, this is just using the formula that $var(V) = I$ so $var(Z'V) = Z'IZ = 1$. So $Z'V$ is a scalar normal variable with mean 0 and variance 1, which we denote z. So, taking the expectation with respect to V is equivalent to taking the expectation with respect to z.

Now, multiply and divide the constants using elementary algebra and you end up with the expectation

$$E\frac{\rho\sigma_u}{\sigma_v}\frac{z}{\frac{\pi}{\sigma_v} + z}$$

The expectation does not exist as you can verify. (What happens is that there is positive probability of $z \approx -\frac{\pi}{\sigma_v}$ where we are dividing by almost 0, so the ratio gets so big that is does not integrated to a finite number.)

The intuition is simple. We estimate the first-order regression

$$\tilde{x}_i = \hat{\pi} z_i ,$$

and then do the second order regression

$$y_i = \beta(\hat{\pi} z_i) + e_i \,,$$

which implies that the β coefficient is identified from the reduced form

$$y_i = \gamma z_i + \epsilon_i \,,$$

as $\hat{\beta} = \frac{\hat{\gamma}}{\hat{\pi}}$. Here, it is obvious that we can get numerically large values for $\hat{\beta}$ if the estimate of π is very small. You may also have the sign of $\hat{\beta}$ flip. If the true value of π is large, this is unlikely to be a problem in practice, but if the true value of π is near zero, you can get very bad estimates. The situation where π is non-zero (and therefore asymptotically valid), but very small (and therefore giving noisy, often useless, estimates) is referred to as the having "weak instruments." For the case of many instruments, you regress X on Z and get $\tilde{X} = Z\hat{\pi}$ so the IV estimator becomes

$$\hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y = (\hat{\pi}'Z'Z\hat{\pi})^{-1}\hat{\pi}'Z'Y \,,$$

and again you can see that small values of π may make the estimate of β shaky. If the reduced form regression has a clear interpretation, you may be better off focusing on that one.

Hahn and Hausman shows the following (see the JEP survey by Michael Murray—I follow his simple exposition here, except I keep the notation from above). Let the equation for x be more general with q regressors (again suppressing the constant).

$$x_i = z_i \pi + \sigma_v v_i \,,$$

Let R^2 be the R-squared from this regression. One can show, to a second-order approximation, for N observations (although we will no have time for derivations):

$$E\{\beta_{IV}\} - \beta = \frac{q\,\rho(1 - R^2)}{NR^2} \,.$$

As you can tell, this bias is near infinity if R^2 is near 0. The bias is worse the larger ρ is, and there is no bias if $\rho = 0$ (but then OLS is BLUE). More surprising, if you add instruments that do not increase R^2 the bias double with the number of instruments. Intuitively, what happens is that by randomness they will capture some of the variation in X—we know that if we have N instruments, then we have a perfect fit in the first stage, and IV is the same as OLS. You are expected to remember this formula, or at least the content of it.

Hahn and Hausman also show that

$$\frac{Bias(\beta_{IV})}{Bias(\beta_{OLS})} \approx \frac{q}{NR^2} \,.$$

from which you can tell that the IV estimator with many instruments can be worse than OLS (also for the case of $q = 1$, if the R^2 is really low).

In general, "what to do" is not fully settled, in particular if you have more than one endogenous variable.

Advice on dealing with potentially weak instruments.

1. Always display the results from the reduced form estimation. In cases where you have one instrument, this sometimes gives a clear answer. Example: you want to estimated the marginal propensity to consume using a change in taxes as an instrument (here we assume that you can find such an exogenous change, even if that is not so easy). The reduced form will tell you if consumption reacts to the taxes and give you a confidence interval, and this may answer most of what you want to know.

2. You always have to show the first-stage estimation. This may be informative, but you need to know if you your instrument(s) is(are) weak.

3. Andrews, Stock, and Sun (2018) (ASS) provide practical advice on testing for weak instruments:

   (a) Common rule-of-thumb: Based on Stock and Yogo (2005), it has become a "standard" rule-of-thumb for one instrument that the first stage F-test should be larger than 10, so you would see that used in papers. For two or three instruments, the rule of thumb may be 20. You should consult Stock and Yogo (2005) for critical values. However, Stock and Yogo's critical values are assuming homoskedasticity, which is often not reasonable.

(b) ASS suggest that you use the recent adjusted "Efficient F-test" of Montiel Olea and Pflueger instead. The Montiel Olea and Pflueger (2013) adjusted F-test is

$$\frac{q\hat{\sigma}_v^2}{\text{Trace}[\hat{Var}(\hat{\pi})N(Z'Z)]}F\,,$$

where k is the number of instruments and $\hat{Var}(\hat{\pi})$ is the heteroskedasticity robust variance estimator. You can notice that for one instrument, when there is no heteroskedasticy the variance of $\hat{\pi}$ is $\sigma_v^2(Z'Z)^{-1}$ (the usual OLS variance) and this reduces to F, which is the standard F statistic for testing all the instruments being 0. You will have to look up the critical values but for one instrument they are equal to the Stock and Yogo ones and the rule of thumb of a value over 10 is probably good.

(c) If you have many endogenous variables, you can test if the instruments are weak for each endogenous variable one-by-one. The literature is not very explicit on this point.

(d) Non-parametric bootstrap tests are not valid in the weak instrument case.

(e) You can do a parametric bootstrap to gain insights, but you would then have to have a model for potential heteroskedasticity. (Although, currently it seems that "everybody" wants to use estimators that allow for heteroskedasticity of unknown form.)

4. If you find clear evidence against weakness of your instruments, you continue as usual. But note that you want to have much larger values of the F-tests to feel comfortable. So how do you continue in the second stage?

   (a) If you suspect/cannot clearly reject weak instruments, use the LIML estimator (or the slightly better Fuller estimator). You may compare to the 2SLS estimator, but if they differ, LIML is likely to be less biased. (The recent focus on weak instruments is the reason why LIML is getting significant attention again.)

   (b) In a recent useful paper, Lee, McCrary, Moreira, and Porter (AER 2022) (LMMP), show that—for the case of one instrument and one endogenous variable—you can explicitly adjust your t-statistic based on the first stage F-test. Possibly using (the same!) robust variance estimator for both stages. Their paper give adjustment factors (larger than 1) based on the first-stage F-test that you can use to adjust the length of the second stage confidence interval. The adjustment factor is non-neglible even of F-values quite a bit larger than the Stock-Yogo rule-of-thumb.

   (c) In general, there is no consensus about how to deal with many instruments in the non-homoskedastic case and ASS has no clear recommendation. If you are in the situation with more than one, but weak, instruments, look up the tests suggest in ASS. If you have a model for potential heteroskedasticity (for example, the variance goes down with size of the unit) you can do the first stage GLS transformation (in this case, just weighing the variables) and get back to the homoskedastic situation.

(d) For the case of many endogenous variables and weak instruments, life is harder. If one endogenous variable has strong instruments, you can use a Frisch-Waugh procedure, but if you have more than one weak instrument, it is not obvious how to construct second-stage valid t- or F-statistics. ASS discuss this case, but as they do not give any clear recommendation and you will need to put in work to decide what do to. Maybe the best you can do is to show robustness. However, this weak IV literature is constantly evolving (I didn't know about the useful LMMP paper a year ago!) so ask, Google, and look at Web-pages of people like Stock and the Andrews'es etc.) for recent work.