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. This survey-of-surveys is also shorter than I would have liked because we are about out of time.

Consider the 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.) 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 (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):

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

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

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

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).

Some advice on estimation: I suggest that you always display the results for the reduced form estimation. (You always have to show the first-stage 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.

If you suspect 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.)

Andrews, Stock, and Sun (2018) (ASS) provide practical advice on testing for weak instruments. The "standard" rule of thumb for one instrument is that the first stage F-test should be larger than 10. For two or three instruments, you should probably use 20. You should consult Stock and Yogo (2005) for critical values. Although, you might want to do a Monte Carlo study yourself, you can adapt the code from the home works, it is not hard, and it can buy you a lot of credibility. But note that the test-statistics in Stock and Yogo (2005) are based on an assumption of homoskedasticity. ASS points out that many papers in AER use the Stock-Yogo rule-of-thumb but then proceed to use White heteroskedasticity-robust or clustered standard errors (which doesn't square with using the F-test that relies on homoskedasticity). ASS suggest that you use the Montiel Olea and Pflueger (2013) adjusted F-test:

$$\frac{k\hat{\sigma}_v^2}{\text{Trace}[\hat{Var}(\hat{\pi})N(Z'Z)^{-1}]}F\,,$$

where k is the number of instruments and $\hat{Var}(\hat{\pi})$ is the hetroskedasticity 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.

I would point out that in many cases, you may have a model for potential heteroskedasticity (for example, the variance goes down with size of the unit) and in this case you can do the first stage GLS transformation (in this case, just weighing the variables) and get back to the homoskedastic situation.

ASS suggest that you use Anderson-Rubin confidence intervals for your parameters of interest in just identified models. (We won't have time to discuss the logic of those, but consult ASS when you need this in your research.) For the over-identified homoskedastic case, they suggest using the likelihood ratio statistic of Moreira (2003).

For the case of many endogenous variables and weak instruments, life is harder. ASS discuss this case, but you will need to put in work to decide what do to. But make sure to show your first-stage and reduced-form regressions. Frisch-Waugh partial correlation plots are always good at signaling outliers and Monte Carlo studies are good (although, you cannot do those with heteroskedasticity of unknown form, you have to choose the model to simulate.)

The final advice is for your future work is to search for recent literature on weak instruments to make sure you know the most recent progress from top econometricians like Jim Stock.