Resurgence of Instrument Variable Estimation and Fallacy of Endogeneity

Duo Qin

1 Evaluation

This paper provides a critical "rational reconstruction" of the evolution of attitudes to the use of the instrumental variable, IV, estimator in different parts of econometrics. The author, like many statisticians, is sceptical about the usefulness of the technique and in particular about the resurgence of its use in micro-econometrics. Both the critique and the detailed, nicely documented, history of the way that interest in instrumental variables has waxed and waned are very interesting. While the paper is polemical and is likely to be controversial, it raises some relevant questions and should engender some debate, hopefully prompting some interesting responses. I think the paper might be more effective were it less polemical, but I think that with some revision the paper is worth publishing. Below I discuss some of the issues and make a few suggestions.

2 Discussion

The title might be better as "The resurgence of instrumental variables and the fallacy of endogeneity" with instrumental rather than instrument.

It might be worth explaining earlier what the author believes the fallacy of endogeneity to be. I presume it arises from what that the author believes to be the fallacy that unobservable structural error terms exist prior to model specification as is discussed on p27. In common terminology a variable can be endogenous either because it is simultaneously determined within a system or because the exogeneity assumption fails. Implicitly it seems as if the author would prefer a definition within the cited Engle-Hendry-Richard (1983) framework where the definition of exogeneity is is terms of the distribution of the observables not in terms of an unobservable error. This might be made more explicit and similarly the reference to the "non-optimal conditional expectations" on page 3, could also be elaborated.

As the paper notes, statisticians tend to be sceptical about IV because one cannot test all the assumptions necessary for the validity of the technique, in particular because the crucial assumption is about the correlation of an observed variable with an unobservable error. Thus the selection of instruments has to rest on beliefs grounded in knowledge of the subject-matter not on statistical inference. These subject matter beliefs are often inherently controversial, not only in economics. Given that the beliefs cannot be falsified some would regard them as purely metaphysical. Some statisticians would take the even more sceptical position that all one can estimate from observational data are correlations not causal relations.

The cited book by Judea Pearl discusses these issues and also some special cases where one can falsify the assumptions about the instrument. While the assumption about the lack of correlation between an observed variable and an unobserved error cannot generally be tested directly, it can be evaluated indirectly by comparing IV model estimates with experimental estimates, as Lalonde (1986) did. Such opportunties are rare and there may be many aspects of the model, estimation technique and experimental set-up that might cause the divergence.

On p6 it might be worth noting that what motivated Fisher (1965) cited was that given the structure of the Brookings model there were more "exogenous" and predetermined variables than observations. Thus some method of selecting instruments was needed if IV was not to be the same as OLS. The problem of too many weak instruments also occurs in other contexts including some panel data models, e.g. Roodman (2009).

In the example, given as system (1) on p14, surely it is the case that given the structure of the model $\Psi_t = (z_{1t}, z_{2t})'$, since $x_t = \lambda\Psi_t + e_t$ is the reduced form for x_t. Then it is not clear what the point of the first paragraph of p17 is. Ψ_t is well defined by the SEM (1), if the model is accepted. The author suggests that (1) is the simplest example. Below I consider a simpler example, which can be used to illustrate the three problems which may cause the usual exogeneity assumption to fail and to which IV may be seen as a solution: simultaneity, errors in variables and omitted variables.

Consider how these three problems might arise in a traditional consumption function, often used for teaching. For time-series observations $t = 1, 2, ..., T$, c_t is consumption, y_t is income, i_t is exogenous investment and all variables are measured as deviations from their means to remove intercepts. In all cases we make the assumption that the equation error in the consumption function ε_t satisfies: $E(\varepsilon_t) = 0$, $E\left(\varepsilon_t^2\right) = \sigma^2$, $E(\varepsilon_t\varepsilon_s) = 0, s \neq t$;

First, consider the simultaneous equations model

$$\begin{aligned} y_t &= c_t + i_t \\ c_t &= \beta y_t + \varepsilon_t. \end{aligned}$$

If we also have: Assumption 1: $E(i_t\varepsilon_t) = 0$, then few would disagree that the OLS estimator

$$\widehat{\beta} = \left(\sum_{t=1}^{T} c_t y_t\right) / \left(\sum_{t=1}^{T} y_t^2\right) \tag{1}$$

is inconsistent and the consistent estimator for β is

$$\widetilde{\beta} = \left(\sum_{t=1}^{T} c_t i_t\right) / \left(\sum_{t=1}^{T} y_t i_t\right). \tag{2}$$

This requires $\sum y_t i_t \neq 0$, which given the structure of the model should be the case and, in any case, can be checked empirically. This check generalises to tests for weak instruments. This estimator can be motivated as IV, indirect least squares or 2 stage least squares, which in just identified models like this are identical. As the author notes, conditional on Assumption 1 we can do a Durbin-Wu-Hausman type test of whether the difference $\left(\widetilde{\beta} - \widehat{\beta}\right)$ is significant, though the test might not perform well with weak instruments.

If there were more instruments, e.g. if $y_t = c_t + i_t + g_t$ and $E(g_t \varepsilon_t) = 0$, we could do a Sargen-Bassman-Hansen type test of the over-identifying restriction. These additional tests may influence our faith in the model but do not test assumption 1.

Second consider the errors in variable model, where consumption is determined by permanent income, y_t^p; actual income is the sum of permanent and transitory components,

$$\begin{aligned} c_t &= \beta y_t^p + \varepsilon_t, \\ y_t &= y_t^p + y_t^\tau. \end{aligned}$$

If in addition to Assumption 1 we have Assumption 2 $E(y_t^p i_t) \neq 0$, $E(y_t^\tau i_t) = 0$, $E(\varepsilon_t y_t^\tau) = 0$; then again i_t is an valid instrument, so that (2) consistently estimates β.

Third consider the ommitted variable model, where there is an unobserved variable, z_t, say the state of confidence, that may be correlated with y_t and appears in the consumption function. So

$$c_t = \beta y_t + \gamma z_t + \varepsilon_t.$$

If in addition to Assumption 1, we also have Assumption 3: $E(z_t i_t) = 0$; then again i_t is an valid instrument, so that (2) estimates β. Whereas in case 1, the instrument investment is directly suggested by the structure of model, this is not true in cases 2 and 3.

While given the assumptions of each model, $\widetilde{\beta}$ consistently estimates β, the issue becomes the plausibility of the assumptions of each model. While they are pedagogically useful, none of these models are plausible. Not only can Assumption 1, be questioned, but the specifications of the first and last models ignore dynamics and inter-temporal optimisation. In the second, Friedman used the errors in variables structure of the permanent income model very effectively to interpret a wide range of evidence but never used Assumption 1 or instrumental variables to estimate the model. In case 3 the instrument is required to be uncorrelated with both the unobserved error and the unobserved omitted variable, two unfalsifiable assumptions.

If we all agree that the instrument influences the dependent variable exclusively through its significant impact on the right hand side variable that is correlated with the error term in a structural equation, then IV works. But we rarely can all agree about that.

As the paper notes, time series analysts tend to worry about the "incredible identifying assumptions" involved. The classic text, Hamilton (1994, p253 and 426), is sceptical about finding instruments and in particular whether lagged values will constitute valid instruments. Hamilton comments positively on the use by Angrist (1990) cited in this paper of the Vietnam draft lottery as an instrument, which has been criticised by others. But he then says "Unfortunately, it is unusual to be able to find such a compelling instrument for many questions that one would like to ask of the data.", p253. The main workhorse of empirical macroeconomics, the DSGE model, tends to be estimated by full information systems methods, typically Bayesian, rather than limited information IV or GMM methods. The systems approach takes advantage of the cross-equation restrictions implied by theory.

However, whether the macro-econometric concerns can be translated into the micro-econometic context, where the aim is to approximate experimental conditions for observational data, may be less clear-cut than the author suggests.

3 References

Hamilton, J. D (1994) *Time Series Analysis*, Princeton University Press.

Lalonde, R.J. (1986) Evaluating the Econometric Evaluations of Training Programs with Experimental Data, *American Economic Review*, 76 604-620.

Roodman, D (2009) A note on the theme of too many instruments, *Oxford Bulletin of Economics and Statistics*, 71(1) 135-158.