Economics Bulletin

Volume 32, Issue 2

Consistently bounding parameter values with one instrument and two endogenous explanatory variables

Richard A Dunn
Texas A&M University

Abstract

The current paper considers a linear regression framework with two endogenous regressors, but only one instrument that is correlated with both. I demonstrate that under reasonable conditions, some of which are testable from the data, these different sources of endogeneity act in opposing directions and hence IV regression can generate economically meaningful bounds.

The author wishes to thank Ivan Canay, John Mullahy, Bobbi Wolfe, Karl Scholz, Richard Woodward, Ximing Wu, David Bessler and seminar participants at the University of Wisconsin, Texas A&M University, Tulane University, University of New Mexico, Towson University and DePauw University. This research was supported by the National Institutes of Health under Ruth L. Kirschstein National Research Service Award No. T32 MH18029-21 from the National Institute of Mental Health.

Citation: Richard A Dunn, (2012) "Consistently bounding parameter values with one instrument and two endogenous explanatory variables", *Economics Bulletin*, Vol. 32 No. 2 pp. 1074-1081.

Contact: Richard A Dunn - dunn.econ@gmail.com.

Submitted: November 17, 2011. **Published:** April 01, 2012.

1. Introduction

Most attempts to overcome endogeneity in applied economic research rely on identifying an instrument that is correlated with the endogenous variable (relevance) but is uncorrelated with the error term (validity). While the former is testable, oftentimes the latter is not and the empirical exercise falters if the researcher is unable to convince the skeptical reader of the instrument's validity.

Yet, this skepticism is wholly warranted. The coefficient estimate for the endogenous explanatory variable from OLS may be closer to the true parameter value than the estimate obtained from using an invalid instrument even if the instrument is 'less endogenous' than the explanatory variable, i.e. less correlated with the error term (Bound, Jaeger and Baker, 1995). The ease of finding some marginally plausible argument to reject the validity of an instrument conspires to create an uncomfortable tension between exploring the myriad important relationships in economics that exhibit endogeneity and deploying an estimation method that may not be robust to untestable departures from validity. The problem is heightened by the lack of alternative methods—designing relevant and ethical laboratory experiments is often unfeasible and so-called 'natural experiments' are both rare and often lack generalizability.

Recently, however, several papers have attempted to define what can be identified through the use of instruments that do not satisfy the standard IV validity assumption (Ashley, 2009; Nevo and Rosen, 2012; Conley, Hansen and Rossi, 2006; Hahn and Hausman, 2003; Manski and Pepper, 2000). Each uses a combination of assumptions about the correlations between the outcome, the instrument, the explanatory variables and/or the error term to generate conditions under which identification of bounds on the true parameter value is achievable. For example, Nevo and Rosen (2012) demonstrate that if the correlation between the instrument and the error term has the same sign as the correlation between the endogenous regressor and the error term, then the coefficient estimates from OLS and IV provide at worst a one-sided bound on the true value.

The current paper is similar in approach, but departs from previous work by considering a different environment in which the use of one instrument fails to achieve point identification—a constant coefficients linear regression framework with two potentially endogenous regressors and only one candidate instrument that is correlated with both. Formally, the relationship of interest is:

$$y_i = \alpha_0 \cdot W_i + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \varepsilon_{0i} \qquad (1)$$

where y_i is the outcome; W_i is a vector of exogenous variables; x_{1i} and x_{2i} are scalars that are potentially endogenous; and ε_{0i}, is a mean zero term that captures unobserved attributes.

The researcher would obviously prefer to have consistent estimates of all parameters, but for policy purposes suppose that the parameter of interest is α_1. Thus, x_1 is termed the *relevant endogenous variable*. The instrument z is proposed for x_1, but there is concern that it is also correlated with x_2, termed the *non-relevant endogenous variable* (as will become evident, x_1 and x_2 are treated asymmetrically because the empirical specification, the set of maintained assumptions and the nature of the possible bounds depends upon which variable is designated as *relevant*). Thus,

$$x_{1i} = \beta_{0,1} \cdot W_i + \beta_{z,1} z_i + \varepsilon_{1i} \quad (2)$$

$$x_{2i} = \beta_{0,2} \cdot W_i + \beta_{z,2} z_i + \varepsilon_{2i} \quad (3)$$

where ε_{1i} and ε_{2i} are mean zero errors.

Thus, the identification problem is similar to that discussed by Leamer (1981) who considers the estimation of demand and supply equations. In this paper, I demonstrate that under reasonable assumptions, some of which are verifiable given the available data, estimating two standard IV specifications—one that includes x_2 in the set of explanatory variables and one that omits x_2—can provide economically meaningful fuzzy bounds on the true value of the parameter of interest, α_1 (fuzzy in the sense that the full covariance structure the regressors is not utlized to further narrow the estimated bounds).

In particular, the critical assumptions required for identification of two-sided fuzzy bounds on the coefficient of the relevant endogenous regressor are: 1) the instrument is uncorrelated with the error term; 2) the instrument is *strong* in the sense that it is highly correlated with the relevant regressor, but is not redundant to the other variables (this is formalized subsequently); and 3) the researcher can sign the relationship between the non-relevant regressor and both the outcome and the unobservables in the outcome equation, i.e. there is prior knowledge of the signs of α_2, and $cov(x_2,\varepsilon_0)$. Notice, however, that no prior knowledge of the analogous relationships for the relevant regressor is required. Furthermore, even when the data do not allow for two-sided bounds, a one-sided bound may still be recoverable.

Although the case of linear specifications with homogenous treatment effects initially appears limited in focus, the subsequent bounding results are nevertheless useful given their continued popularity of such models in economics and other fields. Moreover, the approach is attractive because the bounds are easy to compute using standard IV methods that can be found in modern statistical software packages. Finally, the exceedingly straight-forward estimation technique and minimal computational requirements belie the many empirical problems in economics where such a procedure can be applied, such as estimation of demand systems; output substitution; firm entry, exit and location choice; and household production.

For example, a researcher may wish to undertake a demand analysis where quantity demanded (y) is regressed on own-price (x_1) and the price of either a complement or substitute (x_2) with a supply shock that affected the prices of both goods as the candidate instrument (z). Since both prices are likely endogenous, using the supply shock to identify the own-price elasticity would not be valid. The subsequent result provides conditions under which bounds could still be attained.

2. The basic assumptions

The first four assumptions required for the subsequent bounding result are identical to those that would be necessary for point identification under standard IV estimation with one endogenous variable:

Assumption A1 (exogeneity of W): E[$W\varepsilon_0$]=$\mathbf{0}$.
Assumption A2 (z as an instrument for x_1): E[$z\varepsilon_1$] $= 0$, $\beta_{z,1} \neq 0$.
Assumption A3 (conditional exogeneity of z): $cov(\varepsilon_0,z) = 0$.
Assumption A4 (finite covariances): E[$(x_1\ x_2\ W)'(x_1\ x_2\ W)$] exists and is finite.

As is standard, consider the case without the exogenous variables W, which is equivalent to estimation after projecting y, x_1 and x_2 onto the vector space spanned by W. Thus, the following results hold with additional exogenous variables when interpreted as conditional covariances, i.e. the covariances of the residuals after regressing on W.

$$y_i = \alpha_1 x_{1i} + \alpha_2 x_{2i} + \varepsilon_{0i} \quad (4)$$

$$x_{1i} = \beta_{z,1} z_i + \varepsilon_{1i} \quad (5)$$

$$x_{2i} = \beta_{z,2} z_i + \varepsilon_{2i} \quad (6)$$

3. Bounding result

Consider the following specifications in which x_1 is instrumented with z:

$$y_i = \alpha_1 x_{1i} + v_{1i} \quad (s1)$$

$$y_i = \alpha_1 x_{1i} + \alpha_2 x_{2i} + v_{2i} \quad (s2)$$

where $v_{1i}=\alpha_2 x_{2i}+v_{2i}$ since ($s2$) is assumed to be the true model. Let $\sigma_{1,2}$=cov(x_1,x_2), $\sigma_{z,1}$=cov(x_1,z) and $\sigma_{z,2}$=cov(x_2,z) be scalars. Denote $\alpha_{1,s1}$ and $\alpha_{1,s2}$ as the estimators of α_1 under specifications $s1$ and $s2$. If **A1**-**A4** hold, then their respectively probability limits are:

$$\text{plim } \alpha_{1,s1} = \alpha_1 + \alpha_2(\sigma_{z,2}/\sigma_{z,1}) \quad (7)$$

$$\text{plim } \alpha_{1,s2} = \alpha_1 - \Gamma\sigma_{z,2}\, cov(x_2,v_2) \quad (8)$$

where $\Gamma^{-1}=\sigma_{z,1}\sigma_2^{\,2} - \sigma_{z,2}\sigma_{2,1}$. If x_2 were not endogenous, then $cov(x_2,v_2)=0$ and $\alpha_{1,s2}$, would be a consistent estimate of α_1. Alternatively, if either z were uncorrelated with x_2 ($\sigma_{z,2}=0$) or x_2 did not affect the outcome ($\alpha_2=0$) then $\alpha_{1,s1}$ would be a consistent estimate of α_1. To make the problem interesting, however, these trivial cases are ignored.

To generate bounds, it is obvious that some stand on the signs of α_2, $\sigma_{z,1}$, Γ and $cov(x_2,v_2)$ is necessary. Notice that $\sigma_{z,1}$ enters the expression of Γ. This fact leads to the definition of a *strong instrument* for the purposes of the current paper.

Definition (strong instrument): *The variable z is an strong instrument iff* $sign(\sigma_{z,1})=sign(\Gamma)$.

Notice that $sign(\Gamma)=sign(\rho_{z,1}-\rho_{z,2}\rho_{2,1})$. Thus, a sufficient (but not necessary) condition for $sign(\Gamma)=sign(\beta_{z,1})$ is $|\rho_{z,1}|>|\rho_{z,2}|$, i.e. the partial correlation of the instrument with the relevant endogenous regressor is be larger than with the non-relevant regressor. This leads to the bounding result.

Proposition 1 (two-sided bounds): *If **A1-A4** hold; z is a strong instrument; and $sign(cov(x_2,v_2))=sign(\alpha_2)$, then $max\{\alpha_{1,S1},\alpha_{1,S2}\}$ is an upper-bound on α_1 and $min\{\alpha_{1,s1},\alpha_{1,s2}\}$ is a lower-bound on α_1.*

Proof: *By inspection of (7) and (8).*■

Unlike the standard relevancy requirement for IV estimation that only requires that the correlation between the instrument and the endogenous variable is non-zero, Proposition 1 requires that the correlation is sufficiently large, i.e. *z* is a *strong instrument*. In addition, the researcher must have some prior belief on the relationship between the non-relevant regressor and both the outcome and the unobservables. Specifically, if x_2 is positively(negatively) related to the outcome, it must also be positively(negatively) related to the unobservables.

The need for additional restrictions to achieve some degree of identification follows naturally since the researcher is implicitly relaxing the necessary assumption for point identification that a second exclusion restriction exists. For example, the restriction that $sign(cov(x_2,v_2)) = sign(\alpha_2)$ is similar in spirit to the monotone treatment selection and monotone treatment response properties used by Manski and Pepper (2000) to attain bounds on the treatment effect when no instrument is available for a single endogenous regressor. In their application, education is assumed to increase labor income ($\alpha_2>0$) and individuals who benefit most from additional schooling select into high education levels ($cov(x_2,v_2)>0$). For some applications, it may be more plausible to make this type of assumption about the non-relevant endogenous variable than about the relevant one. It may also be more reasonable than assuming the second independent exclusion restriction that is necessary for point identification.

It is worth noting that in other applications, selection and treatment may work in opposite directions for the non-relevant regressor. For example, an increase in the mortality risk associated with an occupation will tend to increase the wage rate, but individuals who are the least risk averse and who therefore require the lowest risk premium will tend to select into these occupations. Although Proposition 1 does not cover this situation, one-sided bounds may still be possible as reported in Table 1.

Finally, the construction of confidence intervals, and therefore inference, is possible by recognizing that the bounded set of parameter values is constructed as the intersection of two intervals (Chernozhukov, Lee and Rosen, 2008; Nevo and Rosen, 2012; Imbens and Manski, 2004).

3.1 Bounding the non-relevant regressor

It is also possible to use the estimate of $\alpha_{2,S2}$ to provide bounds on the true parameter value of α_2 if **A1-A4** hold. The probability limit for $\alpha_{2,S2}$ is straight-forward to calculate:

$$\text{plim } \alpha_{2,S2} = \alpha_2 + \Gamma\sigma_{z,1}\, cov(x_2,v_2)$$

Proposition 2: *Suppose **A1-A4** hold and z is a strong instrument for* x_1*. If* $sign(\alpha_2)=sign(cov(x_2,v_2))$, then $\min\{0, \alpha_{2,S2}\}$ is a lower-bound on α_2 and $\max\{0, \alpha_{2,S2}\}$ is an upper-bound on α_2. *If* $sign(\alpha_2)\neq sign(cov(x_2,v_2))$ and $\alpha_2<0$, then $\alpha_2<\alpha_{2,S2}$. *If* $sign(\alpha_2)\neq sign(cov(x_2,v_2))$ and $\alpha_2>0$, then $\alpha_2>\alpha_{2,S2}$.

Proof: *If z is a strong instrument for* x_1*, then* $sign(\Gamma)=sign(\sigma_{z,1})$ *and the result then follows.* ■

Proposition 2 is also useful as a specification check if the researcher feels comfortable signing α_2 beforehand. For example, if α_2 is assumed positive and it is also assumed that $sign(\alpha_2)=sign(cov(x_2,v_2))$, then an estimated coefficient which is negative reveals that one of the two assumptions must be mistaken. Either both the signs of α_2 and $cov(x_2,v_2)$ are negative and Proposition 1 still follows or the signs are not the same and Proposition 1 no longer applies.

Moreover, Proposition 2 along with the probability limits for $\alpha_{1,S1}$ and $\alpha_{1,S2}$ actually yield a collection of one-sided bounds on the relevant endogenous regressor even when the conditions of Proposition 1 are not met. These are summarized in Table 1.

Table 1: Matrix of bounding results based on prior assumptions and data attributes

Prior assumptions		From data		Implication		
α_2	$cov(x_2,v_2)$	$\lvert\rho_{z,1}\rvert>\lvert\rho_{z,2}\rho_{2,1}\rvert$	$sign(\sigma_{z,1})=sign(\sigma_{z,2})$	lower bound on α_1	upper bound on α_1	
+	+	Y	Y	$\alpha_{1,S2}$	$\alpha_{1,S1}$	†
+	+	Y	N	$\alpha_{1,S1}$	$\alpha_{1,S2}$	†
+	+	N	Y	$-\infty$	$min\{\alpha_{1,S1},\alpha_{1,S2}\}$	
+	+	N	N	$\alpha_{1,S1}$	∞	
+	-	Y	Y	$-\infty$	$min\{\alpha_{1,S1},\alpha_{1,S2}\}$	
+	-	Y	N	$max\{\alpha_{1,S1},\alpha_{1,S2}\}$	∞	
+	-	N	Y	$-\infty$	$\alpha_{1,S1}$	
+	-	N	N	$\alpha_{1,S1}$	∞	
-	+	Y	Y	$max\{\alpha_{1,S1},\alpha_{1,S2}\}$	∞	
-	+	Y	N	$-\infty$	$min\{\alpha_{1,S1},\alpha_{1,S2}\}$	
-	+	N	Y	$\alpha_{1,S1}$	∞	
-	+	N	N	$-\infty$	$\alpha_{1,S1}$	
-	-	Y	Y	$\alpha_{1,S1}$	$\alpha_{1,S2}$	†
-	-	Y	N	$\alpha_{1,S2}$	$\alpha_{1,S1}$	†
-	-	N	Y	$max\{\alpha_{1,S1},\alpha_{1,S2}\}$	∞	
-	-	N	N	$-\infty$	$\alpha_{1,S1}$	

Notes: † denotes Proposition 1 applies. All bounds assume A1-A4 hold.

4. Discussion

This paper has demonstrated how to find upper and lower bounds on a policy relevant endogenous variable when the proposed instrument is also correlated with a second endogenous variable. The only *a priori* knowledge that is required to estimate these bounds involves assumptions on the non-relevant endogenous regressor. In empirical work, these may be more reasonable than assumptions about the relevant regressor itself or assuming additional exclusion restrictions.

References

Ashley R. (2009) “Assessing the credibility of instrumental variables inference with imperfect instruments via sensitivity analysis” *Journal of Applied Econometrics* **24**, 325-337.

Bound, J., D.A. Jaeger and R.M. Baker (1995) “Problems with Instrumental Variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak” *Journal of the American Statistical Association* **90**, 443-450.

Chernozhukov, V., S. Lee and A. Rosen (2008) “Intersection bounds, estimation and inference” Working paper, MIT.

Conley, T.C., C. Hansen and P.E. Rossi (2008). “Plausibly exogenous” Working Paper, Chicago Graduate School of Business.

Hahn, J. and J. Hausman (2003) “IV estimation with valid and invalid instruments” Working paper, MIT.

Imbens, G. and C.F. Manski (2004) “Confidence intervals for partially identified parameters” *Econometrica* **72**, 1845-1857.

Leamer, E.E. (1981)“Is it a demand curve, or is it a supply curve? Partial identification through inequality constraints” *The Review of Economics and Statistics* **63**, 319-327.

Manksi, C.F. and J.V. Pepper (2000) “Monotone instrumental variables: with an application to the returns to schooling” *Econometrica* **68**, 997-1010.

Nevo, A. and A.M. Rosen (2012) “Identification with imperfect instruments” *The Review of Economics and Statistics*, forthcoming

Staiger, D. and J.H. Stock (1997) "Instrumental variables regression with weak instruments" *Econometrica* **65**, 557-86.