ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 699 October 2011

A cautionary note on tests for overidentifying restrictions

Paulo M.D.C. Parente & J.M.C. Santos Silva

A cautionary note on tests for overidentifying restrictions*

Paulo M.D.C. Parente[†] J.M.C. Santos Silva[‡]

October 13, 2011

Abstract

Tests of overidentifying restrictions are widely used in practice. However, there is often confusion about the nature of their null hypothesis and about the interpretation of their outcome. In this note we argue that these tests give little information on whether the instruments are correlated with the errors of the underlaying economic model and on whether they identify parameters of interest.

JEL classification code: C12, C13, C51, C52.

Key words: GMM, Hansen's J-test, Instrumental variables, Sargan test.

1. INTRODUCTION

Economists doing empirical work often use instrumental variables (IV) and the generalized method of moments (GMM) to identify the parameter of interest.[1] The

*We are grateful to an anonymous referee and to Holger Breinlich, Emanuele Ciani, Daniele Massacci, Patrick Nolen, Matthias Parey, Esmeralda Ramalho, and Joaquim Ramalho for many useful comments and discussions. Santos Silva gratefully acknowledges partial financial support from Fundação para a Ciência e Tecnologia (FEDER/POCI 2010).

[†]Department of Economics, University of Exeter. Email: P.M.Parente@exeter.ac.uk.

[‡]University of Essex and CEMAPRE. E-mail: jmcss@essex.ac.uk.

[1]See, e.g., Wooldridge (2010) for details on these methods.

success of this approach critically depends on the validity of a set of moment conditions and therefore it is not surprising that researchers often try to check whether the assumed moment conditions are valid.

It is well known that when the model is exactly identified it is not possible to check the validity of the moment conditions (e.g., Wooldridge, 2009, p. 529). However, when the model is over identified, researchers often use tests of the overidentifying restrictions to assess the validity of the moment conditions.[2] This practice is misleading because the validity of the overidentifying restrictions is neither sufficient nor necessary for the validity of the moment conditions implied by the underlaying economic model, and therefore provides little information on the possibility of identifying the parameters of interest. Indeed, as noted for example by Deaton (2010), the validity of the moment conditions is an identifying assumption that cannot be tested.

Although this result is known, it is rarely mentioned in the literature. Moreover, when it is mentioned, often no justification is provided for it, or the justification that is provided is either inaccurate or not immediately clear. In this note we present the tests for overidentifying restrictions in a way that makes their nature very transparent, and provide illustrative examples that highlight important characteristics of these tests.

2. TESTS FOR OVERIDENTIFYING RESTRICTIONS

For simplicity, consider a linear model of the form

$$y = x'\beta + u, \tag{1}$$

where x is a k-vector of regressors, β is the k-vector of parameters of interest, and u are the errors of the model, which are correlated with x. If a p-vector of instruments z uncorrelated with u is available and $p \geq k$, β can be consistently estimated from the following moment conditions

$$\mathrm{E}\left[z\left(y - x'\beta\right)\right] = 0. \tag{2}$$

[2] Examples of popular tests for over identifying restrictions are the ones proposed by Sargan (1958) and by Hansen (1982).

Consider a random sample $\{(y_i, x_i', z_i')\}_{i=1}^n$ and let b denote the estimate of β obtained from the sample analog of (2). Then, if $p > k$, tests for overidentifying restrictions check for possible correlation between the residuals $(y_i - x_i'b)$ $i = 1, ..., n$ and the instruments. Hence, these tests are often interpreted as checking the validity of (2).

The crucial point to note is that tests for overidentifying restrictions do not check whether (2) holds but rather whether there is some vector $\beta^* = \text{plim}\,(b)$ such that

$$\mathrm{E}\left[z\left(y - x'\beta^*\right)\right] = 0, \tag{3}$$

where β^* is implicitly defined by the instruments used. Therefore, (2) and (3) imply the orthogonality between the instruments and different errors, and consequently the relation between the validity of the overidentifying restrictions and of the validity of the moment conditions implied by the underlaying economic model is very tenuous.

It is easy to see that the overidentifying restrictions may be valid even if the instruments are correlated with u. Indeed, substituting (1) in (3) we obtain

$$\begin{aligned} \mathrm{E}\left[z\left(x'\left(\beta - \beta^*\right) + u\right)\right] &= 0, &\text{(4a)} \\ \mathrm{E}\left(zx'\right)\left(\beta^* - \beta\right) &= \mathrm{E}\left(zu\right). &\text{(4b)} \end{aligned}$$

Under the usual assumption that $E(zx')$ has rank equal to k, the expression above shows that (3) will hold as long as it is possible to find a vector $\lambda = (\beta^* - \beta)$ such that $E(zu) = E(zx')\,\lambda$.

When the instruments are valid, $E(zu) = 0$ and the solution to (4b) is $\lambda = 0$. In this case the moment conditions are valid and the estimator identifies the parameters of interest. However, $E(zu) = 0$ is a sufficient but not necessary condition for (4b) to have a solution.[3] Indeed, even if $E(zu) \neq 0$, it may be possible to find a vector β^* such that (4b) holds. In this case, the overidentifying restrictions are still valid but the estimator identifies $\beta^* = \beta + \lambda$ rather than the parameters of interest.

[3] The necessary and sufficient condition for (4b) to have a solution is that $E(uz)$ is in the span of the columns of $E(zx')$, i.e., that rank $\begin{bmatrix} E(zx') & E(uz) \end{bmatrix} = k$.

The result that the validity of the overidentifying restrictions is not sufficient to ensure the identification of the parameters of interest is not entirely new. In his study of the local power of tests for overidentifying restrictions, Newey (1985) noted that these restrictions may be valid even if the instruments are not. Newey's (1985) result for local alternatives is presented in more detail by Hall (2005). In turn, De Blander (2008) considers non-local alternatives and gives somewhat less transparent version of the result presented above. In the same vein, some authors note that when all instruments have the same rational, the fact that the model passes the test of overidentifying restrictions offers little comfort (see, e.g., Murray, 2006. p. 117, and Wooldridge, 2009, p. 529), but do not provide a clear explanation for why this is the case. More recently, Wooldridge (2010, pp. 134-7) and Deaton (2010, pp. 430-2) provided deeper and clearer discussions of this issue.[4]

To see that the validity of the overidentifying restrictions is also not necessary for the parameters of interest to be successfully identified, consider a setting where the population of interest is a mixture of S sub-populations such that $\mathrm{E}\left[z_s\left(y - x'\beta_s\right)\right] = 0$, where expectations are taken over the entire population and β_s are the parameters of interest for sub-population $s = 1, \ldots, S$.[5] With a random sample from the entire population, it is clear that β_s can be estimated by using z_s as instruments. However, if $\beta_s \neq \beta_t\ \exists t \neq s$, in general there will be no value β^* such that $E\left[z_s\left(y - x'\beta^*\right)\right] = 0,\ \forall s$, and therefore the overidentifying restrictions will be invalid if the full set of instruments $z_1, \ldots, z_S$ is used. Hence, the set of overidentifying restrictions may be invalid even if each individual orthogonality condition holds and each instrument identifies a parameter of interest.

[4]It is worth noting that this issue is related to the possibility of existence of observationally equivalent models in the GMM framework. That is, there may be multiple sets of valid moment restrictions, involving different sets of parameters. This issue was studied recently by Hall and Pelletier (2007).

[5]This situation considered by Imbens and Angrist (1994), who have pointed out that different sets of instruments lead to the estimation of different objects.

That the rejection of the overidentifying restrictions can be the result of parameter heterogeneity was pointed out by Angrist, Graddy, and Imbens (2000) and it is also noted by Angrist and Pischke (2009, p. 166), who remark that testing overidentifying restrictions "is out the window in a fully heterogeneous world."

In short, whether or not the overidentifying restrictions are valid gives little information on whether the instruments are correlated with the errors of the underlaying economic model, and on whether parameters of interest can be successfully identified. Below we provide simple examples that illustrate this point.

3. ILLUSTRATIVE EXAMPLES

For simplicity, we focus on the case where the researcher wants to estimate the returns to education using a wage equation of the form

$$\ln(w) = \beta_0 + \beta_1 s + u, \tag{5}$$

where w denotes the wage, s is a measure of the level of schooling, and β_1 measures the returns to education and is the parameter of interest. As usual, s is not assumed to be uncorrelated with u, namely due to the possible omission of important regressors. Therefore, consistent estimation of β_1 requires the availability of a vector z of instrumental variables. In what follows, we illustrate the lack of relation between the validity of the instruments and the validity of the overidentifying restrictions using the data studied by Card (1995), which is used by Wooldridge (2009 and 2010) to exemplify the use of instrumental variables estimators.[6]

3.1. Overidentifying restrictions are valid irrespective of the instruments' validity

Suppose that, as for example in Wooldridge (2009, p. 522), the wage equation is estimated by IV using mother's and father's schooling, respectively ms and fs, as instruments for s. That is, z is the vector $\begin{bmatrix} 1 & ms & fs \end{bmatrix}'$.

[6] The data are available at http://www.stata.com/data/jwooldridge/eacsap/card.dta.

Then, assuming that $\mathrm{E}(zz')$ has full rank, the first stage regression consists in estimating the linear projection

$$\mathrm{L}(s|ms, fs) = \pi_0 + \pi_1 ms + \pi_2 fs.$$

Now, because ms and fs essentially measure the same thing, π_1 and π_2 are likely to be similar.[7] Indeed, with the data considered here, the estimates of π_1 and π_2 are, respectively, 0.20 and 0.22, with standard errors of about 0.02. This situation is not particular to this example and it is likely to occur whenever the instruments have essentially the same motivation and are measured on the same scale.

By the same reasoning, it is likely to be the case that ms and fs will have similar coefficients in the linear projection of u on the instruments. That is, it is likely that

$$\mathrm{L}(u|ms, fs) = \gamma_0 + \gamma_1 ms + \gamma_2 fs = \gamma_0 + \gamma_1 (ms + fs),$$

with the instruments being invalid if $\gamma_1 \neq 0$.

Suppose now that indeed $\pi_1 = \pi_2 \neq 0$ and $\gamma_1 = \gamma_2$.[8] Then, it is possible to show that, even if $\gamma_1 = \gamma_2 \neq 0$,

$$\mathrm{E}[(\ln(w) - \beta_0^* - \beta_1^* s)z] = 0$$

for $\beta_0^* = (\beta_0 - \pi_0\gamma_1/\pi_1 + \gamma_0)$ and $\beta_1^* = (\beta_1 + \gamma_1/\pi_1)$. Indeed, from (5) we have that

$$\begin{aligned}\mathrm{E}[(\ln(w) - \beta_0^* - \beta_1^* s)z] &= \mathrm{E}[(u + \beta_0 - \beta_0^* + (\beta_1 - \beta_1^*)\, s)z], \\ &= \mathrm{E}[(u + \pi_0\gamma_1/\pi_1 - \gamma_0 - \gamma_1/\pi_1 s)z].\end{aligned}$$

Now, writing $\varepsilon = s - \mathrm{L}(s|ms, fs)$ and $\eta = u - \mathrm{L}(u|ms, fs)$, we have

$$\begin{aligned}\mathrm{E}[(\ln(w) - \beta_0^* - \beta_1^* s)z] &= \mathrm{E}[(u + \pi_0\gamma_1/\pi_1 - \gamma_0 - \gamma_1/\pi_1(\pi_0 + \pi_1(ms + fs) + \varepsilon))z], \\ &= \mathrm{E}[(\gamma_0 + \gamma_1(ms + fs) + \eta - \gamma_0 - \gamma_1(ms + fs) - \gamma_1/\pi_1\varepsilon)z], \\ &= \mathrm{E}[(\eta - \gamma_1/\pi_1\varepsilon)z] = 0,\end{aligned}$$

[7] See Holmlund, Lindahl and Plug (2010) for a recent study on the effects of parent's education on children's schooling.

[8] In fact, all that is needed is that $\frac{\gamma_2}{\gamma_1} = \frac{\pi_2}{\pi_1}$ which implies that evidently the problem persists even if the instruments are measured in different scales.

with the last equality holding because both ε and η are defined as differences between a variable and its linear projection on z, and consequently are uncorrelated with the instruments. Therefore, whether or not z is correlated with u, the overidentifying restrictions will be valid.

Going back to the illustrative data set, using ms and fs as instruments we obtain an estimate of the returns to education of 7% and the (robust) Hansen's J-test statistic has a p-value of 0.22. Of course, given the similarity between the instruments used, this result offers little information on whether or not z is correlated with u.

3.2. Overidentifying restrictions are invalid when each instrument is valid

Suppose now that the effect of education on wages is heterogeneous (see, e.g., Card, 1998, and Ichino and Winter-Ebmer, 1999). More specifically, suppose that the population is divided into two groups and that the returns to schooling for the first group are equal to β_1^h, whereas for the second group the returns are β_1^l, with $\beta_1^h > \beta_1^l$. Using ξ to denote an unobservable indicator that is 1 for individuals in the first group, being zero otherwise, the wage equation can be written as

$$\ln(w) = \beta_0 + \left[\beta_1^h \xi + \beta_1^l (1-\xi)\right] s + u.$$

Consider now the case in which two sets of binary instruments are available: z_h is uncorrelated with u and with $(1-\xi)s$, but is correlated with ξs, whereas z_l is uncorrelated with u and ξs, but is correlated with $(1-\xi)s$. That is, z_l is an instrument that is correlated with the level of schooling only for the individuals for which the returns to education are low, whereas z_h is only correlated with schooling for individuals with high returns to education.

Under standard regularity conditions, the following moment conditions hold

$$\begin{aligned} \mathrm{E}\left[\left(\ln(w) - \beta_0 - \beta_1^h s\right) z_h\right] &= 0, \\ \mathrm{E}\left[\left(\ln(w) - \beta_0 - \beta_1^l s\right) z_l\right] &= 0, \end{aligned}$$

and therefore both instruments are valid in the sense that each of them will allow the identification of a parameter of interest. However, when both instruments are used, the IV or GMM regression of $\ln(w)$ on s generally will only identify a mixture of β_1^h and β_1^l. In this case, in general, the overidentifying restrictions are invalid because there is no single parameter that makes the errors of the model orthogonal to both instruments.

Ichino and Winter-Ebmer (1999) argue that, if they are valid, instruments like parental education will identify β_1^l because they are likely to affect mainly the schooling of individuals with limited ability who may receive more schooling if their parents are highly educated. Therefore, the if ms and fs are valid instruments, the results in Subsection 3.1 would be an estimate of β_1^l. Ichino and Winter-Ebmer (1999) also argue that β_1^h can be estimated using as instruments variables that identify individuals who, thanks to their ability, choose more schooling in the absence extraneous constraints, but drop out of school if constrained. Ichino and Winter-Ebmer (1999) give as an example of such instrument a dummy for the father's participation in WWII.

In the spirit of Ichino and Winter-Ebmer (1999), here we use the variables "living with a single mother at the age of 14" and "living with stepparent at the age of 14" to estimate β_1^h. With these instruments, the returns to education are estimated to be 23% and the (robust) Hansen's J-test statistic has a p-value of 0.30. Therefore, the new set of instruments leads to an estimate of the returns to education that is indeed much larger than the one previously obtained, and again there are no indication that the overidentifying restrictions are violated.

If the model is estimated using simultaneously both sets of instruments, however, the (robust) Hansen's J-test statistic has a p-value of 0.01, which would lead to the rejection of the null at the usual 5% level. At first sight this result is puzzling because if each sub-set of overidentifying restrictions is valid, the full set should also be valid. However, this naïve interpretation is flawed because the residual whose orthogonality to the instruments is checked by the test for overidentifying restrictions depends on the chosen set of instruments, and therefore the set of restrictions tested when z_h and

z_l are used together is not the union of the sets of restrictions tested when z_h and z_l are considered separately.

Overall, the three test statistics computed in this example suggest that z_h and z_l estimate different sets of parameters, but are mute about the ability of these sets of instruments to identify parameters of interest.

4. CONCLUDING REMARKS

The examples presented in this note clearly illustrate a number of interesting points. First and foremost, they show that the validity of the overidentifying restrictions provides little information on the ability of the instruments to identify the parameter of interest. It is important to note that this is not a finite sample limitation of the test, but rather it is one of its intrinsic characteristics. Second, these examples show that, contrarily to what is often stated, the interpretation of the outcome of a test for overidentifying restrictions does not depend on the presence of enough valid instruments. Finally, these examples suggest that it is more appropriate to interpret tests for overidentifying restrictions as checks for whether or not all the instruments identify the same vector of parameters, as proposed by Hausman (1983). That is, the tests check the coherency of the instruments rather than their validity.

REFERENCES

Angrist, J.D., Graddy, K., and Imbens, G.W. (2000). "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *The Review of Economic Studies*, 67, 499-527.

Angrist, J. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton (NJ): Princeton University Press.

Card, D. (1995). "Using Geographic Variation in College Proximity to Estimate the Return to Schooling." In: Christophides, L.N., Grant, E.K., and Swidinsky, R. (Eds), *Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp*. Toronto: University of Toronto Press.

Card, D. (1998). "The Causal Effect of Education on Earnings". In: Ashenfelter, O., Card, D., (Eds.), *Handbook of Labor Economics*, vol. 3. Amsterdam: North-Holland.

Deaton, A. (2010). "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature*, 48, 424–55.

De Blander, R. (2008). "Which Null Hypothesis do Overidentification Restrictions Actually Test?". *Economics Bulletin*, 3, 1-9.

Hall, A. (2005). *Generalized Method of Moments*, New York (NY): Oxford University Press.

Hall, A. R. and D. Pelletier (2007): "Non-Nested Testing in Models Estimated via Generalized Method of Moments," forthcoming in *Econometric Theory.*

Hausman, J.A. (1983). "Specification and Estimation of Simultaneous Models." In: Griliches, Z. and Intriligator. M., (Eds.), *Handbook of Econometrics*, vol. 1. Amsterdam: North-Holland.

Holmlund, H., Lindahl, M., and Plug, E. (2010). *The Causal Effect of Parents' Schooling on Children's Schooling - A Comparison of Estimation Methods*, CESifo Working Paper Series 3234, CESifo Group Munich.

Ichino, A. and Winter-Ebmer, R. (1999). "Lower and Upper Bounds of Returns to Schooling: An Exercise in IV Estimation with Different Instruments," *European Economic Review*, 43, 889-901.

Imbens, G.W. and Angrist, J.D. (1994). "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-475.

Murray, M. P. (2006). "Avoiding Invalid Instruments and Coping with Weak Instruments," *Journal of Economic Perspectives*, Vol. 20, 111-132.

Newey, W. (1985). "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29, 229-256.

Sargan, J.D. (1958). "The Estimation of Economic Relationships using Instrumental Variables," *Econometrica*, 26, 393-415.

Wooldridge, J.M. (2009). *Introductory Econometrics, A Modern Approach*, 4th ed., Mason (OH): South-Western.

Wooldridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., Cambridge (MA): MIT Press.