Available online at www.sciencedirect.com
ScienceDirect
Economics Letters 100 (2008) 68 – 71
economics letters
www.elsevier.com/locate/econbase

The reduced form: A simple approach to inference with weak instruments

Victor Chernozhukov [a], Christian Hansen [b,*]

[a] MIT, United States
[b] University of Chicago, Graduate School of Business, 5807 S. Woodlawn Ave., Chicago, IL 60637, United States

Received 30 July 2007; received in revised form 19 October 2007; accepted 14 November 2007
Available online 28 November 2007

Abstract

In this paper, we show that conventional heteroskedasticity and autocorrelation robust inference procedures based on the reduced form provide tests and confidence intervals for structural parameters that are valid when instruments are strongly or weakly correlated to the endogenous variables.

Keywords: Heteroskedasticity; Autocorrelation; Weak identification

JEL classification: C12C30

1. Introduction

We consider a practical approach to performing inference that will be valid under weak identification.[1] The approach provides tests with the correct size and confidence regions with correct coverage in the presence of weak instruments that can be made robust to heteroskedasticity and autocorrelation through the use of conventional robust covariance matrix estimators. The procedure uses only OLS estimation and inference and so may be easily implemented using standard inference tools available in any regression software.

2. Testing procedure

Suppose we are interested in the parameters of a structural equation in a simultaneous equation model that can be represented in limited information form as

$$Y = X\beta + \epsilon \tag{1}$$

$$X = Z\Pi + V \tag{2}$$

where Y is an $n\times 1$ vector of outcomes, X is an $n\times s$ matrix of endogenous regressors, and Z is an $n\times r$ matrix of excluded instruments where $r\geq s$.[2] The condition that $r\geq s$ insures that there are, in principle, sufficient instruments to identify the parameters of interest.

The inference procedure we consider may then be motivated as follows. Suppose we are interested in testing the null that $\beta=0$. When Z is highly correlated to X, we can proceed by estimating β by 2SLS and using the corresponding asymptotic distribution to test the hypothesis. However, when the instruments are weakly correlated with X, the usual asymptotics may provide a poor approximation to the actual sampling distributions of the estimator and test statistics. An alternative test based on the reduced form for Y is also available. Under the null hypothesis, the exclusion restriction implies that the coefficients on the instruments in the reduced form for Y, defined by substituting Eq. (2) into Eq. (1) to obtain $Y=Z\gamma+U$, should equal zero. Thus, testing that these coefficients equal 0 tests the hypothesis that $\beta=0$. This procedure will be robust to weak instruments since no information about the correlation between X and Z is required to test that there is no relationship between the outcome and the instruments.[3]

* Corresponding author.
E-mail address: chansen1@gsb.uchicago.edu (C. Hansen).

[1] See Stock et al. (2002) and Andrews and Stock (2005) for surveys of the weak instruments literature.

[2] The model allows for additional exogenous regressors which have been "partialed out".

[3] Angrist and Krueger (2001) advocate exactly this approach to testing the hypothesis that $\beta=0$ when one is worried about weak-identification.

0165-1765/$ - see front matter © 2008 Published by Elsevier B.V.
doi:10.1016/j.econlet.2007.11.012

Table 1
Summary Statistics for Survey of IV Papers from AER, JPE, and QJE 1999–2004

	N*	Mean	10th	25th	50th	75th	90th
A. One Observation per First Stage							
Sample Size	220	39,491	60	106	1577	55,495	152,742
Number of Instruments	261	2.89	1	1	1	3	10
First-Stage Wald Statistic	221	140.31	7.22	11.56	22.02	67.65	637.97
Correlation	126	0.375	0.021	0.045	0.333	0.652	0.812
B. One Observation per Paper							
Sample Size	38	7272	64	200	700	4965	17,649
Number of Instruments	56	3.55	1	1	2	4	9
First-Stage Wald Statistic	28	142.06	7.53	12.51	22.58	117.32	722.27
Correlation	22	0.293	0.026	0.041	0.267	0.529	0.559

Note: This table reports summary statistics for a survey of linear IV papers published in various issues of the *American Economic Review*, *Journal of Political Economy*, and *Quarterly Journal of Economics* in 1999–2004. The exact issues are given in the main text. The top panel reports results where each unique first stage is treated as an observation, and the bottom panel reports results which treat the within paper medians as unique observations. In each panel, we report summary statistics for the sample size used in estimating the first-stage relationship, the number of instruments, the first-stage Wald statistic for testing for relevance of the excluded instruments, and the correlation between the first-stage and structural error terms. For each variable, we report the number of nonmissing observations, N∗, as well as the mean and various percentiles of the distribution.

This basic intuition may be extended to testing the more general hypothesis that $\beta=\beta_0$. Under the null hypothesis, we may write (1) as

$$Y - X\beta_0 = Z\alpha + \epsilon, \tag{3}$$

and the exclusion restriction implies that $\alpha=0$. Thus, a test of $\alpha=0$ in Eq. (3) tests the null that $\beta=\beta_0$. Letting $W_S(\beta_0)$ denote the conventional Wald statistic for testing $\alpha=0$, we note that the use of a robust covariance matrix estimator, e.g. a HAC estimator, in forming $W_S(\beta_0)$ will result in a robust statistic for testing $\beta=\beta_0$ that will be asymptotically distributed as a χ_r^2 regardless of the strength of the instruments.

Repeating the testing procedure mentioned above for multiple values of β_0 allows the construction of confidence intervals which are robust to weak instruments and heteroskedasticity or autocorrelation through a series of conventional least squares regressions. The procedure for constructing a confidence interval is as follows:

1. Select a set, $\mathcal{B}$, of potential values for β.
2. For each $b\in\mathcal{B}$, construct $\tilde{Y}=Y-Xb$ and regress $\tilde{Y}$ on Z to obtain $\hat{a}$. Use $\hat{a}$ and the corresponding estimated covariance matrix of $\hat{a}$, $\widehat{Var}(\hat{\alpha})$, to construct the Wald statistic for testing $\alpha=0$, $W_S(b) = \hat{\alpha}'\left[\widehat{Var}(\hat{\alpha})\right]^{-1}\hat{\alpha}$. The use of a robust covariance matrix estimator in forming $\widehat{Var}(\hat{\alpha})$ will result in tests and confidence intervals robust to both weak instruments and heteroskedasticity and/or autocorrelation.
3. Construct the $1-p$ level confidence region as the set of b such that $W_S(b)\le c(1-p)$ where $c(1-p)$ is the $(1-p)^{th}$ percentile of a χ_r^2 distribution.

That $W_S(\beta_0)\xrightarrow{d}\chi_r^2$ under the null hypothesis that $\beta=\beta_0$ follows under usual regularity conditions for consistency and asymptotic normality of $\hat{\alpha}$ and consistent estimation of $\widehat{Var}(\hat{\alpha})$ in Eq. (3). Such regularity conditions will only require a rank condition on the instrument matrix, Z, and will impose no restrictions on the relationship between X and Z. Thus, the resulting inference statements will be correct regardless of the strength of the instruments. One can also show under standard regularity conditions that $W_S(\cdot)$ is asymptotically equivalent to the S-statistic of Stock and Wright (2000).

It is important to note that when there are more instruments than endogenous regressors, the simple testing procedure outlined above may inefficiently use information. We can see this by noting that we obtain an r degree of freedom test for testing s hypotheses. Many of the testing procedures discussed in the surveys of Stock, Wright, and Yogo (2002) and Andrews and Stock (2005) were designed to deal with this problem. Also, the test statistic defined above jointly tests model specification and the hypothesis that $\beta=\beta_0$ when $r>s$ and so may result in empty confidence sets in cases where one would reject the hypothesis of correct specification using the over identifying restrictions.[4]

3. Survey of IV Papers

To provide evidence on the relevance of the weak-instrument problem in practice, we examined the March 1999 to March 2004 issues of the *American Economic Review*, the February 1999 to June 2004 issues of the *Journal of Political Economy*, and the February 1999 to February 2004 issues of the *Quarterly Journal of Economics*. We summarize the results in Table 1. The statistics we report are based on papers which estimated a linear instrumental variable (IV) model with only one righthand side

[4] In a working version of this paper (Chernozhukov and Hansen, 2005), we show how the above procedure may be modified to produce s degree of freedom tests in the presence of overidentifying restrictions but note that much of the ease of implementation is lost. In the working paper, we also include simulation results that suggest that the above procedure has correct size. The simulation results also suggest the power loss is small when there are not many more instruments than endogenous regressors. These results are not presented for brevity but are available upon request.

Table 2
IV Regressions of Log GDP per Capita on Institutions

	Base Sample	Base Sample	Base Sample without Neo-Europes	Base Sample without Neo-Europes	Base Sample without Africa	Base Sample without Africa	Base Sample	Base Sample
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
N	64	64	60	60	37	37	64	64
W	16.322	12.110	6.917	6.150	36.240	27.563	5.244	3.648
Latitude		Y		Y		Y		Y
Continent Dummies							Y	Y
2SLS Results								
β_{2SLS}	0.924	0.969	1.236	1.219	0.578	0.576	0.932	1.026
Asymptotic Interval	(0.580,1.267)	(0.544,1.395)	(0.461,2.011)	(0.424,2.013)	(0.408,0.749)	(0.384,0.768)	(0.281,1.583)	(0.114,1.938)
Weak-Instrument Interval	(0.662,1.667)	(0.653,2.066)	(0.731,4]	(0.709,4]	(0.418,0.816)	(0.378,0.832)	(0.443,4]	(0.357,4]

Note: This table reports the 2SLS estimate of the effect of institutions on GDP (β_{2SLS}). Asymptotic Interval reports the 95% confidence interval constructed using the usual asymptotic approximation, and Weak-Instrument Interval reports the 95% confidence interval constructed using the weak-instrument robust statistic. Both intervals account for possible heteroskedasticity. W is the first-stage Wald statistic for testing instrument irrelevance, and N is the sample size. Even numbered columns include latitude as an explanatory variable, and columns (7) and (8) include continent dummy variables. Columns (3) and (4) omit Australia, Canada, New Zealand, and the USA. Columns (5) and (6) omit all African countries.

endogenous variable.[5] We use the Wald statistic for testing the relevance of the instruments in the first stage as our measure of the strength of instruments.[6] In cases where we had sufficient information, we also constructed an estimate of the correlation between the reduced form error and the structural error, ρ. The estimates of ρ were made under assumptions of strong instruments and homoskedasticity and so should be regarded with caution.

There are a number of features that are worth pointing out. It is unsurprising that the bulk of estimated IV models use exactly as many instruments as endogenous regressors and that there are relatively few examples where there are more than two or three over identifying restrictions. This finding suggests that the simple procedure outlined in this paper will be adequate in many settings encountered in practice. It is interesting that the median value of ρ is around .3, which suggests that the degree of correlation between the structural and first-stage errors is quite modest in many cases. The most interesting results concern the first-stage Wald statistics. It is apparent that researchers are unwilling to report IV estimates when the first-stage relationship is not statistically significant. Assuming that the 10th percentile value of approximately 7 corresponds to a case with a single instrument, the p-value for testing the null hypothesis that the coefficient on the instrument is equal to 0 is approximately .01. By the time we reach the median value of the Wald statistic, the p-value is negligible. However, statistical significance is not sufficient to guarantee that inference based on the usual asymptotic approximation has good performance. Hansen, Hausman, and Newey (2005) suggest that a useful rule of thumb is to not use the usual asymptotic approximation if the Wald statistic is not in the mid-thirties or higher, and this cutoff value is increasing in the number of instruments.[7] Using this rule of thumb, we see that the asymptotic approximation is suspect in more than 50% of the cases examined. Overall, our survey suggests that the weak-instrument problem may be practically relevant and that having a simple and intuitive approach to producing valid inference statements may be quite valuable.

4. Empirical example

We illustrate the use of weak-instrument robust statistic in a simple empirical example. The data are drawn from Acemoglu, Johnson, and Robinson (2001) which examines the effect of institutions on economic performance using mortality rates among European colonists as an instrument for current institutions. This paper provides a useful case to consider in that there is some variation in the strength of the instruments across specifications. The model is just-identified and so is typical in the sense that most empirical analyses are based on just-identified empirical models.

We focus on the results from Table 4 of Acemoglu, Johnson, and Robinson (2001). These results correspond to a simple linear IV model

$$Y = X\beta + W\gamma + \epsilon$$
$$X = Z\Pi + W\Gamma + V$$

where Y is the log of PPP adjusted GDP per capita in 1995, X is average protection against expropriation risk from 1985 to 1995 which provides a measure of institutions and well-enforced property rights, and W is a set of additional covariates which

[5] In total we found 108 articles that estimated a linear IV model of which 91 reported some results using only one right hand side endogenous variable.

[6] The Wald statistic equals $r * F$ where F is the first-stage F-statistic for testing the relevance of the instruments. It is closely related to the concentration parameter, which may be estimated by $r(F-1)$ in the one endogenous regressor case; see, for example, Hansen, Hausman, and Newey (2005).

[7] Hansen, Hausman, and Newey (2005) suggest that a cutoff value in the mid-thirties provides a useful rule of thumb for the use of the asymptotic approximation of Bekker (1994).

varies across specifications and may include a normalized measure of distance from the equator (latitude) as well as dummy variables for a country's continent. Detailed descriptive statistics and data descriptions are found in Acemoglu, Johnson, and Robinson (2001).

Table 2 reports the 2SLS estimate of β. Heteroskedasticity robust confidence intervals constructed from both the standard asymptotic distribution and the weak-instrument robust statistics are provided.[8] The first-stage Wald statistic for the test that $\Pi = 0$ is also reported at the top of the table. There is substantial variation in the strength of the instruments across specifications. W ranges from 3.65 to 36.24. For low values of W, we would expect that the usual asymptotic confidence intervals and the weak-instrument robust confidence intervals to be different, while for values of W in the 30's, we would expect that the usual intervals and the weak-instrument robust intervals to be quite close.

Looking at the interval estimates, we observe the expected pattern in the actual results. In columns (5) and (6), where the relationship between the instrument and endogenous variable is the strongest (W above 30), the usual 95% confidence intervals and the weak-instrument intervals are quite similar. As W decreases, the differences become much larger. In particular, in the four cases with $W < 10$, the upper limit of the confidence interval is equal to the upper limit of the set we consider for β_0 and is much larger than the upper limit of the usual confidence intervals. In the two remaining cases with $W \approx 12$, the upper bounds of the intervals are much larger than those of the usual intervals, but remain within the interval considered for β_0.

It is interesting that in all cases the weak-instrument intervals provide a sharp lower bound on the value of β. Even when W is low and the 95% weak-instrument robust intervals appear to be unbounded on one side, the weak-instrument intervals have a positive lower bound that is removed from zero. This finding provides strong evidence that institutions do matter for GDP and that the lower bound on the effect is substantial. Thus, we see that even in cases where instruments are weak, the data may still inform us about parameter values and allow us to make economically useful inferences. This example also illustrates that there may be large differences in the confidence intervals which could dramatically alter the inference.

References

Acemoglu, D., Johnson, S., Robinson, J.A., 2001. The Colonial Origins of Comparative Development: An Empirical Investigation. American Economic Review 91 (5), 1369–1401.

Andrews, D.W.K., Stock, J.H., 2005. Inference with Weak Instruments. Cowles Foundation Discussion Paper No. 1530.

Angrist, J.D., Krueger, A.B., 2001. Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. Journal of Economic Perspectives 15, 69–85.

Bekker, P.A., 1994. Alternative Approximations to the Distributions of Instrumental Variables Estimators. Econometrica 63, 657–681.

Chernozhukov, V., Hansen, C., 2005. The Reduced Form: A Simple Approach to Inference with Weak Instruments. Working Paper. Available at SSRN: http://ssrn.com/abstract=937943.

Hansen, C., Hausman, J., and Newey, W.K., 2005. Estimation with Many Instrumental Variables, mimeo.

Stock, J.H., Wright, J.H., 2000. GMM withWeak Identification. Econometrica 68, 1055–1096.

Stock, J.H., Wright, J.H., Yogo, M., 2002. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. Journal of Business and Economic Statistics 20 (4), 518–529.

[8] For the purposes of this analysis, we construct the weak-instrument intervals by considering β_0 equally spaced at 0.001 unit intervals in the range $[-1,4]$.