REGRESSION MODELS USING THE INSTRUMENTAL VARIABLES

Prof. univ. dr. Constantin ANGHELACHE
Drd. Cristina SACALĂ
Drd. Emilia STANCIU
Academia de Studii Economice, Bucureşti

Abstract

This article describes the main elements on regression models with instrumental variables. There are reviewed the main theoretical approaches in the literature, asymptotic properties when using sensitive instruments**, t-statistics** with instrumental variables, traditional asymptotic modified **s-statistics** and **t-statistics,** finite sample properties with sensitive instruments.

Keywords: identification, regression, interval, statistics, coefficient

1. Generalities

It is known that standard asymptotic inference techniques to estimate instrumental variable can be extremely weak if we have sensitive instruments. In some cases, failure is very high. False results can be framed in confidence intervals giving the impression of high precision. These punctual estimations give us precise values of these coefficients that have sometimes irreversible consequences. Ultimately, if an equation is mistakenly identified, then they do not give accurate data about the system's parameters. We try to clarify the statistics of the test and the confidence intervals with correct highlights, meaning that they lead to accurate inferences when instruments are sensitive and essentially identical with statistics test with instrumental variables, asymptotic and confidence intervals when correct instruments are available.

Most studies are made on inference in regression models with instrumental variables that focused on simple model which includes endogenous variables. If we consider the general regression model with instrumental variables, in the model results with one endogenous variables included, individual structural coefficient do not apply. The problem occurs when the null value is specified for complete parameter vector, in which case null value estimation can provide us a consistent estimation of the error variance, but the zero value specification of an individual coefficient cannot be specified.

The analysis is focused on obtaining valid values of individual structural coefficients in regression model with instrumental variables.

The approach is similar to the one made by Choi and Philips in 1992, when they have given a finite sample and asymptotic deduction in structural equations partially identified. We are trying to expand the use of Choi and Philips, using sensitive instruments and methods for deducting nonstandard structural coefficients. We take into account cases where tools are sensitive to all structural coefficients and cases in which the instruments are sensitive only to certain coefficients. Also, we consider cases in which the instruments are weak but strong individual coefficients for a linear combination of structural coefficients. We use the asymptotic sensitive instruments Staiger and Stock (1997) to analyze the asymptotic behavior of the estimators and statistics tests of individual structural coefficients. Also, we evaluate the performance of various estimators and finite sample test statistics through a more extensive Monte Carlo set of experiments.

After reviewing the literature on the estimation and inference in recent regression models with instrumental variables with sensitive instruments, we present the standard regression model with instrumental variables for endogenous variables. Then, we review the standard identification conditions and determine how to identify partial and sensitive instruments. Then we refer to methods of estimation and inference methods in regression with instrumental variables, with a focus estimation and inference in the case of individual structural parameters. Finally, we conclude the asymptotic behavior of various estimators and test statistics in a variety of cases with sensitive instruments. We evaluate the performance of finished samples of various statistics through an extensive Monte Carlo set of simulations.

2. The statistical and econometric regression model concept evolution

Several recent papers have examined the distribution of instrumental variable estimator in identifying sensitive and performance related matters asymptotic traditional tests. Among these works are the works of Bekker (1994); Blomquist and Dahlberg (1999); Bound, Jaeger and Baker (1995); Choi and Philips (1992); Hahn and Hausman (2002); Hahn and Inuoe (2002); Hall, Rudebusch and Wilcox (1996); Chamberlain and Imbens (2004); Kleibergen (2000, 2002); Kleibergen and Zivot (2003); Maddie and Jeong (1992); Moreira (2003); Nelson and Startz (1990); Philips (1989); Staiger and Stock (1997); Stock, Wright and Yogo (2002); Stock and Yogo (2004); Wang and Zivot (1998); WNG (1999); Zivot, Startz and Nelson (1998). Dufour (1997) provided overall results for obtaining correct sensitive identification probability levels. Especially, Dufour showed that α nominal test to be valid in a sensitive identification confidence intervals involved in statistical test must be unlimited at least with $1 - \alpha$ value.

Half a century ago, Anderson and Rubin (1949) described the statistic Anderson-Rubin (AR), in which normal conditions it provides an exact test, a small sample of a hypothesis that specifies values for each element of Beta structural parameter vector. Zivot, Startz and Nelson (1998) showed how to use statistics AR to build reliable areas for a single endogenous variables, agreed an improved statistics for maximum likelihood and estimated the method of generalized moments based on tests with LR and LM adjusted degrees of freedom.

Wang and Zivot (1998) gave an asymptotic justification for using Staiger and Stoke (1997) asymptotes for these cases. Recently, Kleinbergen (2002) and Moreira (2003) proposed LM asymptotically tests, as being more accurate than the AR test and a higher probability percentage (likelihood ratio LR) and Lagrange multiplier tests studied by Wang and Zivot (1998).

The analyses of the papers presented above is limited to single endogenous variables or assumptions that specify values for all the vector of coefficients. We handle deduction in the two-point factor variables that extend the results of Choi and Philips (1992) in the case of sensitive instruments. It stressed that Dufour (1997), Wang and Zivot (1998), Dufour and Jasiak (2000) describe the use of projections in several statistics tested correlated to obtain sets of reliable elements β individual but studying these methods in the presence of sensitive instruments. Basically, using projection procedure generally requires a complicated numerical maximization. Recently, Taamouti (2001), Dufour and Taammouti (2003) gave a limited set of results for obtaining reliable analytical sets of coefficients based on projects for individual structures based on certain statistics test.

Stock and Wright (2000) explored to reach wider structural parameters estimated by GMM deduction with sensitive instruments and in the case of the simple linear equations model is based on two-stage small squares or maximum likelihood estimates. If some endogenous variables are thoroughly checked, Stock and Wright suggest focusing on the identified parameters and AR using statistics to identify remaining parameters. However, Stock and Wright expressed that their method of construction of the confidence asymptotically valid intervals vectors are somehow difficult, but the confidence interval can be found by asymptotically conservative design parameters, as suggested by Dufour in 1987.

Kleibergen (2000) gave an alternative to AR concentrate statistics of Stock and Wright in linear model with instrumental variables. A general context alternative in GMM context is a research of Kleibergen (2002).

We consider the linear structure equation with right hand k variables:

$$\begin{aligned} y &= \underset{(n\times k)}{X}\ \underset{(k\times 1)}{\beta} + \underset{(n\times 1)}{u} \\ &= \underset{(n\times 1)}{X_i}\ \underset{(1\times 1)}{\beta_i} + \underset{(n\times(k-1))}{X_{-i}}\ \underset{((k-1)\times 1)}{\beta_{-i}} + \underset{(n\times 1)}{u} \end{aligned} \qquad (1)$$

Where $\mathbf{X}_i$ represents the *i* column of $\mathbf{X}_1$, $\mathbf{X}_{-i}$ is the residual value of **X** and **u** is the deviation error vector. Our attention is focused on inference scalar parameter βi using regression with instrumental variables where variables X are endogenous. The reduced form of population regression model is Y and each column of X over all values of q and exogenous instruments for matrix Z, like:

$$\underset{(n\times 1)}{y} = \underset{(n\times q)}{Z}\ \underset{(q\times 1)}{\theta} + \underset{(n\times 1)}{v} \qquad (2)$$

$$\underset{(n\times k)}{X} = \underset{(n\times q)}{Z}\ \underset{(q\times k)}{\Gamma} + \underset{(n\times k)}{V} \qquad (3)$$

The corresponding equations for endogenous variables $\mathbf{X_i}$ and $\mathbf{X}_{-i}$ are:

$$\underset{(n \times 1)}{X_i} = \underset{(n \times q)}{Z} \underset{(q \times 1)}{\Gamma_i} + \underset{(n \times 1)}{V_i} \qquad (4)$$

$$\underset{(n \times (k-1))}{X_{-i}} = \underset{(n \times q)}{Z} \underset{(q x(k-1))}{\Gamma_{-i}} + \underset{(n \times (k-1))}{V_{-i}} \qquad (5)$$

The model described in equations (1) - (3) is called linear regression model with instrumental variables.

Consider the probability convergence vector $\xrightarrow{p}$ and distribution convergence vector $\xrightarrow{d}$. We require the following major assumptions on conditions on exogenous variables and error terms.

- **Z** has *q* value on the entire column and is not correlated with **u** and **V**
- $\mathbf{E[Z_t Z_i^i] = M > 0}$, where $\mathbf{Z_t}$ represents the *t* observation of **Z**
- the error terms u_t and V_t are considered with 0 mean value and are not serial correlated with the positive matrix of covariance.

β vector is usually estimated by the method of instrumental variables (GMM and TSLS equivalent method). The I estimator with instrumental variables is:

$$\hat{\beta}_{IV} = (X' P_z X)^{-1} X' P_{zy} = (\hat{X}' \hat{X})^{-1} \hat{X}' y \qquad (6)$$

Where $\mathbf{P_z = Z(Z^i Z)^{-1} Z'}$ and $X = P_z \hat{X}$.

Using standard regression techniques, we can express βi estimator of the fourth degree:

$$\hat{\beta}_{i\,IV} = (\hat{X}'_i \hat{Q}_{-i} \hat{X}_i)^{-1} \hat{X}'_i \hat{Q}_{-iy} \qquad (7)$$

Where $\hat{X'}_i = P_z X$, $X_{-i} \hat{=} P_z X_{-1}$ si $Q_{-i} = \hat{I}_q - P_x$

We suggest some statistics for individual structures for the deduction of the coefficient regression model with instrumental variables that are rigid sensitive instruments. Some of these methods are based on the estimator with instrumental variables estimator LIML and some of them with IV estimator. We briefly describe these statistics and introduce some new ones.

4. Asymptotic "t" Statistics with instrumental variables

Suppose we want to test Ho: βi = βi0 based on traditional instrumental variables estimation. Standard practice is to use asymptotic statistics t **:**

$$t_{IV} (\beta^0_i) = \frac{\beta_{i,\,IV} - \beta^0_i}{SE(\hat{\beta}_{i,\,IV})} \qquad (9)$$

where $SE(\hat{\beta}_{i,IV}) = \sqrt{\hat{\sigma}_{uu,IV} \hat{H}_{ii}}$, $\sigma_{uu,IV} = n_{-1} (y-X\hat{\beta}_{IV})' (y-X\hat{\beta}_{IV})$

LIML estimator of β maximizes the probability function - log "log likelihood function" referring to Γ and Σ :

$$L^c(\beta) = -n \ln(2\pi) - \frac{n}{2} \ln k(\beta) - \frac{n}{2} \ln \mathrm{I}\, YQ'_Z Y \mathrm{I}\, , \tag{10}$$

where Y = [yX] and

$$k(\beta) = \frac{(y-X\beta)'(y-X\beta)}{(y-X\beta)'Qz\,(y-X\beta)}$$

LIML β estimator minimizes equivalently k (β) and the minimized value, $k(\beta_{LIML}) = k_{LIML}$ can be shown to the smallest root of the equation determining IY'QxY - I kY'QzY. LIML estimator is usually expressed as a k-class estimator

$$\hat{\beta}_{LIML} = [\,X'\,(\,I_n - \hat{k}_{LIML} Qz)X]^{-1} [X'(I_n - \hat{k}_{LIML} Qz)y]$$

For testing $\beta_i = \beta_0$, where ratio „t" – LIML is

$$t_{LIML}(\beta^0_i) = \frac{\hat{\beta}_{i,\,LIML} - \beta^0_i}{SE(\hat{\beta}_{i,\,LIML})} \tag{11}$$

where $\widehat{SE}(\beta_{LIML}) = \sqrt{var(\beta_{i,LIML})} = \sqrt{\sigma_{uu,LIML}\,[X'(In - \hat{k}_{LIML} Qz)X_{ii}^{-1}}$

Also, LR statistics takes the following form:

$$LR_{LIML}(\beta^0_i) = n \ln(\hat{k}_{LIML}(\beta^0_i)) - n \ln(\hat{k}_{LIML}) \tag{12}$$

where $k_{LIML}(\beta^0_i)$ is calculeted with the probability log function, with $\beta_i = \beta^0_i$ restriction and k_{LIML} is calculated from probablility function (10).

„S" Statistics and „t" modified Statistics

We consider the formation of a statistical test for H0: βi = β0i so it will be close to 0 even if the estimated deviation from reality is small or if evidence is sensitive for identification:

$$\Psi_i = \hat{\Delta}_i (\beta_{i,IV} - \beta^0_i) \tag{13}$$

$$\text{where } \hat{\Delta}_i = \sqrt{H_{ii}^{-1}} \tag{14}$$

Stock and Wright (2000) found an AR statistic concentrated for testing H0: $\beta_i = \beta^0_i$ for a GMM framework. In linear regression with instrumental variables, this statistic takes the form:

$$AR(\beta^0_i) = \frac{[y - X_i\beta^0_i - X_i\hat{\beta}_{-i}(\beta^0_i)]' Pz[y - X_i\beta^0_i - X_i\hat{\beta}_{-i}(\beta^0_i)]}{[y - X_{-i}\beta^0_i - X_{-i}\hat{\beta}_{-i}(\beta^0_i)]' Qz[y - X_i\beta^0_i - X_{-i}\hat{\beta}_{-i}(\beta^0_i)]/(n-k)} \quad (15)$$

Where $\hat{\beta}_{-i}(\beta^0_i)$ presents IV or LIML β_{-i} estimation that impose $\beta_{i=}$ β_i^0. By estimating the restricted LIML, we minimizes (13), while the restricted estimation with instrumental variables takes the analytical form (19). When $\hat{\beta}_{-i}(\beta^0_i) = \hat{\beta}_{-i,IV}(\beta^0_i)$ we use $AR_{IV}(\beta^0_i)$ and when $\hat{\beta}_{-i}(\beta^0_i) = \hat{\beta}_{-i,LIML}(\beta^0_i)$ we use $AR_{LIML}(\beta^0_i)$.

Kleibergen (2000) proposed a concentrated version of K statistics to test the individual hypothesis H0: βi = β0i which has the form:

$$K(\beta^0_i) = \frac{(y - X_i\beta^0_i - X_{-i}\hat{\beta}_{-i,\ LIML}(\beta^0_i))' Pw_{(\beta 0i)}(y - X_i\beta^0_i - X_{-i}\hat{\beta}_{-i,LIML}(\beta^0_i))}{(y - X_i\beta^0_i - X_{-i}\hat{\beta}_{-i,\ LIML}(\beta^0_i))' Qz(y - X_i\beta^0_i - X_{-i}\hat{\beta}_{-i,LIML}(\beta^0_i))/(n-k)} \quad (16)$$

If we consider $\beta= (\beta_{i,}\beta'_{-2})'$ and apply the hypothesis testing H0: $\beta = \beta^{0,}$ using AR statistics, then:

$$AR(\beta^0) = \frac{(y - X\beta^0)' Pz\ (y - X\beta^0)/ k}{(y - X\beta^0)' Qz\ (y - X\beta^0)/ (n - k)} \quad (17)$$

If the errors are normally distributed, Anderson and Rubin (1949) showed that (17) is distributed in $F_{k,n-k}$ finite samples below zero value.. This result is maintained regardless the quality of the instruments.

5. Asymptotic properties when using sensitive instruments

We evaluate asymptotic properties conditions when using the sensitive instruments for individual regression coefficients deduction of instrumental variables. To simplify asymptotic analysis, we focus on the regression model with instrumental variables (1) - (3) with two endogenous variables, with $\beta = (\beta_1, \beta_2)'$.

Taking into account the results of Staiger and Stock (1997) and Wang and Zivot (1998) we define the framework sensitive instruments using a near zero value. With multiple endogenous variables, characterizing sensitive instruments becomes more complicated because the Z instruments can be sensitive to endogenous variables coefficients at all or only for a subset of coefficients.

Asymptotic distributions of the various estimators and statistic tests in cases with sensitive instruments depend on parameters which measures the issues of X_1 and X_2 endogenous, multivariate normal random vectors scaled and standardized quality measurement tools Z.

If the instruments are weak for all structural coefficients we cannot get a valid asymptotic inference using any of the proposed statistic tests. However, although valid asymptotically but conservative confidence sets for individual coefficients can be calculated using AR's Dufour and Taamouti projection sets. In case of very sensitive instruments, these sets are endless with probability close to the coverage probability mentioned.

Most asymptotic results for estimators and statistic tests were based until now on the case with sensitive instruments. In the third case with sensitive instruments, β_1 and β_2 are also sensitive but the individual linear combination of $\alpha = \beta_1 + \beta_2$ is strong. To determine the limiting distributions of β_1, β_2 and α we return to Choi and Philips (1992) and consider a rotation of the regression model. An asymptotically valid set for β_i scalar confidence level based on the reversal of α 1 T (β_i^0) is defined by the relationship:

$$C_{\beta i}(\alpha) = (\beta^0_i : T(\beta^0_i)' \leq C_\alpha \qquad (18)$$

Where C_α is the 1- α limiting quintile distribution of T (β_i^0). Calculation of the set (18) requires finding values of βi0 so T $(\beta_i^0) < C_\alpha$. In general, the process involves a numerical expression. However, using the conclusions of Dufour (1997), Zivot, Startz, Nelson (1998), Dufour and Jasiak (2001) we can find out that T $(\beta_i^0) \leq C_\alpha$ inequality can be rewritten as a quadratic inequality, as:

$$a(\beta_i^0)^2 + b\beta_i^0 + c \leq 0,$$

where the values of a, b, c, depend on the information provided and on cα, then the regions of trust defined (18) have closed shape convenient expression and can take one of four forms: a period connected form $(\beta_i^L, \beta_i^{H; L}$ = lower; H= upper); union of the two rays $(-\infty, \beta_i^L)$ U (β_i^H, ∞), the entire real line, or empty set.

Finite sample properties with sensitive instruments

In this section we evaluate the properties of the finished sample statistics deduction competition for individual structural coefficients using a comprehensive set of Monte Carlo experiments. Some authors have considered Monte Carlo regression models with varying instruments with sensitive instruments are the most important works of Choi and Philips (1992); Hall, Rudebusch and Wilcox (1997). Most studies have focused on Monte Carlo estimation methods and deductive performance in this sensitive instruments based on models with one right endogenous variable. Choi and Philips (1992), Flores-Lagunes (2000) and Kleibergen 2000 used models with two endogenous variables and the results of their work shows that it is misleading to extrapolate results from one case with one variable from one case multivariables.

Staiger-Stock sensitive instruments show varying distributions of estimators instruments and test statistics depend on three key parameters issues:

- The degree of endogeneity as it is measured by the correlation coefficients p_{u1}, p_{u2}, p_{12};
- The number of instruments, q;
- The relevance instruments measured by Λ'Λ / q. The instruments are irrelevant when Λ'Λ / q = 0.

For the endogenous variable Steiger's simulation experiments show that the instruments are essentially Stock sensitive when 0 <Λ'Λ / q <10.

The tools are pretty good when Λ'Λ / q> 10. In the case of multiple endogenous variables, $\Lambda'\Lambda$ / q is a weak matrix and tools are characterized by a minimum of $\Lambda'\Lambda$ / q. In addition, Steiger and Stock shows that the performance standard inference methods with sensitive instruments is insignificant in models with many irrelevant tools (high value of q and $\Gamma \approx 0$) and have high degrees of endogeneity.

Bibliografie

1. Anderson, T. W. and H. Rubin (1949). "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Annals of Mathematical Statistics,* 20, 46-63.
2. Bekker, P. A. (1994). "Alternative approximations to the distributions of the instrumental variables estimators," *Econometrica,* 62, 657-681.
3. Blomquist, S. and M. Dahlberg (1999). "Small sample properties of LIML and jackknife IV estimators: experiments with weak instruments," *Journal of Applied Econometrics,* 14(1), 69-88.
4. Bound, I, D. A. Jaeger, and Baker R. M. (1995). "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak," *Journal of the American Statistical Association,* 90,443-450.
5. Chamberlain, G. and G. Imbens (2004). "Random effects estimators with many instrumental variables," *Econometrica,* 72, 295-306.
6. Choi, I. and P. C. B. Phillips (1992). "Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations," *Journal of Econometrics,* 51,113-150.
7. Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics,* New York: Oxford University Press.
8. Dufour, J.-M. (1997). "Some impossibility theorems in econometrics with applications to structural and dynamic models," *Econometrica,* 65,1365-88.
9. Dufour, J.-M. and J. Jasiak (2001). "Finite sample limited information inference methods for structural equations and models with unobserved and generated regressors," *International Economic Review,* 42, 815-843.
10. Dufour, J.-M. and L. Khalaf (1997). "Simulation based finite and large sample inference methods in simultaneous equations," unpublished manuscript, Universite de Montreal.
11. Dufour, J.-M. and M. Taamouti (2003). "Projection-based statistical inference in linear structural models with possibly weak instruments," unpublished manuscript, Department of Economics, Universite de Montreal.
12. Flores-Lagunes, A. (2000). "Estimation methods robust to weak instruments," unpublished manuscript, Department of Economics, Ohio State University.
13. Hall, A. R., G. D. Rudebusch and D. W. Wilcox (1996). "Judging instrument relevance in instrumental variables estimation," *International Economic Review,* 37,283-289.
14. Hahn, J. and J. A. Hausman (2002), "A new specification test for the validity of instrumental variables," *Econometrica,* 70,163-189.
15. Hahn, J. and A. Inoue (2002). "A Monte Carlo comparison of various asymptotic approximations to the distribution of instrumental variables estimators," *Econometric Reviews,* 21, 309-336.
16. Kleibergen, F. (2000). "Pivotal statistics for testing subsets of structural parameters in the IV regression model," Tinbergen Institute Discussion Paper 2000-88/4, University of Amsterdam.
17. Kleibergen, F. (2002). "Pivotal statistics for testing structural parameters in instrumental variables regression," *Econometrica,* 70, 1781-1803,
18. Maddala, G. S. and J. Jeong (1992). "On the exact small sample distribution of the IV estimator," *Econometrica,* 60,181-83.
19. Moreira, M. (2003). "A conditional likelihood ratio test for structural models," *Econometrica,* 71,1027-1048.
20. Nelson, C. R. and R. Startz (1990a). "Some further results on the exact small sample properties of the instrumental variables estimator," *Econometrica,* 58, 967-976.

21. Nelson, C. R. and R. Startz (1990b): "The distribution of the instrumental variables estimator and its /-ratio when the instrument is a poor one," *Journal of Business,* 63, S125-S140.
22. Phillips, P. C. B. (1989). "Partially identified econometric models," *Econometric Theory,* 5,181-240.
23. Shea, J. (1997). "Instrument relevance in multivariate linear models: A simple measure," *The Review of Economics and Statistics,* Vol. LXXIX, No. 2, 348-352.
24. Staiger, D. and J. H. Stock (1997): "Instrumental variables regressions with weak instruments," *Econometrica,* 65, 557-586.
25. Stock, J. H. and J. Wright (2000), "GMM with weak identification," *Econometrica,* 68,1055-1096.
26. Stock, J. H, J. Wright, and M. Yogo (2002). "A survey of weak instruments and weak identification in generalized method of moments," *Journal of Business and Economic Statistics,* 20, 518-529.
27. Stock, J. H. and M. Yogo (2004). "Testing for Weak Instruments in Linear IV Regression," in D. W. K. Andrews and J. H. Stock, eds., *Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg,* Cambridge: Cambridge University Press.
28. Taamouti, M. (2001). "Techniques d'inference exact dans les modeles structurels avec applications macroeconomiques," Ph.D. thesis, Departemente de sciences economiques, Universite de Montreal.
29. Wang, J. and E. Zivot (1998): "Inference on structural parameters in instrumental variables regressions with weak instruments," *Econometrica ,* Vol. 66, No. 6,1389-1404.
30. Wong, K.-F. (1999). "A simulation comparison of inference for instrumental variable estimators," unpublished manuscript, Department of Economics, Chinese University of Hong Kong.
31. Zivot, E., R. Startz, and C. R. Nelson (1998). "Valid confidence intervals and inference in the presence of weak instruments," *International Economic Review,* Vol. 39, No. 4, pp. 1119-1144