Journal of Econometrics 185 (2015) 332–342

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Jackknife instrumental variable estimation with heteroskedasticity[☆]

Paul A. Bekker[a,1], Federico Crudu[b,c,*]

[a] Faculty of Economics and Business, University of Groningen, The Netherlands
[b] Pontificia Universidad Católica de Valparaíso, Chile
[c] University of Sassari and CRENoS, Italy

ARTICLE INFO

Article history:
Received 18 July 2013
Received in revised form
20 August 2014
Accepted 30 August 2014
Available online 24 November 2014

JEL classification:
C12
C13
C23

Keywords:
Instrumental variables
Heteroskedasticity
Many instruments
Jackknife

ABSTRACT

We present a new jackknife estimator for instrumental variable inference with unknown heteroskedasticity. It weighs observations such that many-instruments consistency is guaranteed while the signal component in the data is maintained. We show that this results in a smaller signal component in the many instruments asymptotic variance when compared to estimators that neglect a part of the signal to achieve consistency. Both many strong instruments and many weak instruments asymptotic distributions are derived using high-level assumptions that allow for instruments with identifying power that varies between explanatory variables. Standard errors are formulated compactly. We review briefly known estimators and show in particular that our symmetric jackknife estimator performs well when compared to the HLIM and HFUL estimators of Hausman et al. in Monte Carlo experiments.

1. Introduction

The presence of unknown heteroskedasticity is a common setting in microeconometric research. Inference based on many instruments asymptotics, as introduced by Kunitomo (1980), Morimune (1983) and Bekker (1994), shows 2SLS is inconsistent under homoskedasticity. Bekker and Van der Ploeg (2005) show in general LIML is many-instruments inconsistent as well under heteroskedasticity. A number of estimators have been considered, including the two step feasible GMM estimator of Hansen (1982), the continuously updated GMM estimator of Hansen et al. (1996), the grouping estimators of Bekker and Van der Ploeg (2005), the jackknife estimators of Angrist et al. (1999), the modified LIML estimators of Kunitomo (2012) and the HLIM and HFUL estimators of Hausman et al. (2012). In particular this last paper has been important for the approach that we present here.

Our starting point is aimed at formulating a consistent estimator for the noise component in the expectation of the sum of squares of disturbances when projected on the space of instruments. That way a method of moments estimator can be formulated similar to the derivation of LIML as a moments estimator as described in Bekker (1994). Surprisingly the estimator can be described as a symmetric jackknife estimator, where 'omit one' fitted values are used not only for the explanatory variables but instead for all endogenous variables including the dependent variable. Influential papers on jackknife estimation include Phillips and Hale (1977), Blomquist and Dahlberg (1999), Angrist et al. (1999), Donald and Newey (2000), Ackerberg and Devereux (2003). Our jackknife estimator shares with LIML the property that the endogenous variables are treated symmetrically in the sense that it is invariant to the type of normalization, as discussed by Anderson (2005).

Hausman et al. (2012) and Chao et al. (2012, 2014) use a LIML version of the JIVE2 estimator of Angrist et al. (1999). In case of homoskedasticity and many weak instruments, while assuming the number of instruments grows slower than the number of observations, the authors show the HLIM estimator is as efficient as LIML.[2] Thus it seems the efficiency problems of jackknife estimators noted in Davidson and MacKinnon (2006) are overcome. Here we show

☆ We thank Tom Wansbeek and Tiemen Woutersen for helpful comments.
* Correspondence to: Avenida Errazuriz 2734, Valparaíso, Chile.
E-mail addresses: p.a.bekker@rug.nl (P.A. Bekker), federico.crudu@gmail.com (F. Crudu).
1 P.O. Box 800, 9700 AV Groningen, The Netherlands.

2 Also the diagonal elements of the projection matrix of the instruments should converge to zero uniformly.

http://dx.doi.org/10.1016/j.jeconom.2014.08.012
0304-4076/

there is room for improvement. The symmetric jackknife estimator has a signal component that is larger than that found for HLIM, resulting in a smaller component in the asymptotic covariance matrix. Monte Carlo experiments show it performs better than HLIM and its Fuller modifications in terms of the bias–variance trade-off.

The asymptotic theory allows for both many instruments and many weak instruments asymptotics. Influential papers in this area include Donald and Newey (2000), Hahn et al. (2004), Hahn (2002), Hahn and Inoue (2002), Chamberlain and Imbens (2004), Chao and Swanson (2005), Stock and Yogo (2005), Han and Phillips (2006), Andrews and Stock (2007), Van Hasselt (2010). Our results are formulated concisely. They are based on high level assumptions where the concentration parameter need not grow at the same rate as the number of observations and the quality of instruments may vary over explanatory variables.

The plan of the paper is as follows. In Section 2 we present the model and some earlier estimators. Section 3 uses a method of moments reasoning to formulate a heteroskedasticity robust estimator that is subsequently interpreted as a symmetric jackknife estimator. Asymptotic assumptions and results are given in Section 4 and proved in the Appendix. Section 5 compares asymptotic distributions and Section 6 compares exact distributions based on Monte Carlo simulations. Section 7 concludes.

2. The model and some estimators

Consider observations in the n vector $\boldsymbol{y}$ and the $n \times g$ matrix $\boldsymbol{X}$ that satisfy

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

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

where the g vector $\boldsymbol{\beta}$ and the $k \times g$ matrix $\boldsymbol{\Pi}$ contain unknown parameters, and $\boldsymbol{Z}$ is an $n \times k$ observed matrix of instruments. Similar to Hausman et al. (2012) we assume $\boldsymbol{Z}$ to be nonrandom, or we could allow $\boldsymbol{Z}$ to be random, but condition on it, as in Chao et al. (2012). The assumption $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{Z}\boldsymbol{\Pi}$ is made for convenience and could be generalized as in Hausman et al. (2012), or as in Bekker (1994). The disturbances in the $n \times (1+g)$ matrix $(\boldsymbol{\varepsilon}, \boldsymbol{V})$ have rows $(\varepsilon_i, \boldsymbol{V}_i)$, which are assumed to be independent, with zero mean and covariance matrices

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \sigma_i^2 & \boldsymbol{\sigma}_{12i} \\ \boldsymbol{\sigma}_{21i} & \boldsymbol{\Sigma}_{22i} \end{pmatrix}.$$

The covariance matrices of the rows $(y_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$, are given by

$$\boldsymbol{\Omega}_i = \begin{pmatrix} 1 & \boldsymbol{\beta}' \\ \boldsymbol{0} & \boldsymbol{I}_g \end{pmatrix} \boldsymbol{\Sigma}_i \begin{pmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{\beta} & \boldsymbol{I}_g \end{pmatrix}. \tag{3}$$

Throughout we use the notation where $\boldsymbol{P} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$ has elements $P_{ij} = \boldsymbol{e}_i'\boldsymbol{P}\boldsymbol{e}_j$, and $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ are conformable unit vectors.

The estimators that we consider are related to LIML which is found by minimizing the objective function

$$\mathrm{Q}_{\mathrm{LIML}}(\boldsymbol{\beta}) = \frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{P}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{I}_n - \boldsymbol{P})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}. \tag{4}$$

The LIML estimator and Fuller (1977) modifications are given by

$$\hat{\boldsymbol{\beta}} = \left\{\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X} - \lambda_f \boldsymbol{X}'(\boldsymbol{I}_n - \boldsymbol{P})\boldsymbol{X}\right\}^{-1} \left\{\boldsymbol{X}'\boldsymbol{P}\boldsymbol{y} - \lambda_f \boldsymbol{X}'(\boldsymbol{I}_n - \boldsymbol{P})\boldsymbol{y}\right\},$$

$$\lambda_f = \lambda - \alpha/(n-k),$$

$$\lambda = 1/\lambda_{\max}[\{(\boldsymbol{y}, \boldsymbol{X})'\boldsymbol{P}(\boldsymbol{y}, \boldsymbol{X})\}^{-1} (\boldsymbol{y}, \boldsymbol{X})'(\boldsymbol{I}_n - \boldsymbol{P})(\boldsymbol{y}, \boldsymbol{X})],$$

where $\lambda_{\max}$ indicates the largest eigenvalue. For $\alpha = 0$ LIML is found, which has no moments under normality. Under normality and homoskedasticity, where the matrices $\boldsymbol{\Sigma}_i$ do not vary over $i = 1, \ldots, n$, the Fuller estimator that is found for $\alpha = 1$ has moments and is nearly unbiased. If one wishes to minimize the mean square error, $\alpha = 4$ would be appropriate. However, as shown by Bekker and Van der Ploeg (2005), LIML is in general inconsistent under many-instruments asymptotics with heteroskedasticity.[3]

Similarly, the Hansen (1982) two-step GMM estimator is inconsistent under many-instruments asymptotics. It is found by minimizing

$$\mathrm{Q}_{\mathrm{GMM}}(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{Z} \left\{\sum_{i=1}^{n} \hat{\sigma}_i^2 \boldsymbol{Z}_i'\boldsymbol{Z}_i\right\}^{-1} \boldsymbol{Z}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{5}$$

where $\hat{\sigma}_i^2 = (y_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}})^2$ and $\hat{\boldsymbol{\beta}}$ is a first stage IV estimator such as 2SLS or LIML. A many-instruments consistent version is given by the continuously updated GMM estimator of Hansen et al. (1996), which is found by minimizing the objective function (5) where $\hat{\sigma}_i^2$ is replaced by $\hat{\sigma}_i^2(\boldsymbol{\beta}) = (y_i - \boldsymbol{X}_i\boldsymbol{\beta})^2$. Newey and Windmeijer (2009) showed this estimator and other generalized empirical likelihood estimators are asymptotically robust to heteroskedasticity and many weak instruments. Donald and Newey (2000) gave a jackknife interpretation. However, the efficiency depends on using a heteroskedastic consistent weighting matrix that can degrade the finite sample performance with many instruments as was shown by Hausman et al. (2012) in Monte Carlo experiments.

To reduce problems related to the consistent estimation of the weighting matrix Bekker and Van der Ploeg (2005) use clustering of observations. If this clustering, or grouping, is formulated as a function of $\boldsymbol{Z}$, it is exogenous and continuously updated GMM estimation can be formulated conditional on it. Bekker and Van der Ploeg (2005) give standard errors that are consistent for sequences where the number of groups grows at the same rate as the number of observations. Contrary to LIML, the asymptotic distribution is not affected by deviations from normality. It uses the between group heteroskedasticity to gain efficiency, yet it loses efficiency due to within group sample variance of the instruments.

Another way to avoid problems of heteroskedasticity is to use the jackknife approach. The jackknife estimator, suggested by Phillips and Hale (1977) and later by Angrist et al. (1999) and Blomquist and Dahlberg (1999) uses the omit-one-observation approach to reduce the bias of 2SLS in a homoskedastic context. The JIVE1 estimator of Angrist et al. (1999) is given by

$$\hat{\boldsymbol{\beta}}_{\mathrm{JIVE1}} = (\tilde{\boldsymbol{X}}'\boldsymbol{X})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{y},$$
$$\boldsymbol{e}_i'\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{X}}_i = \frac{\boldsymbol{Z}_i(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X} - h_i\boldsymbol{X}_i}{1 - h_i}, \tag{6}$$

where $h_i = P_{ii}$, and $i = 1, \ldots, n$. It is robust against heteroskedasticity and many-instruments consistent. The JIVE2 estimator of Angrist et al. (1999) shares the many-instruments consistency property with JIVE1. It uses $\tilde{\boldsymbol{X}} = (\boldsymbol{P} - \boldsymbol{D})\boldsymbol{X}$ and thus minimizes a 2SLS-like objective function

$$\mathrm{Q}_{\mathrm{JIVE2}}(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\{\boldsymbol{P} - \boldsymbol{D}\}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{7}$$

where $\boldsymbol{D} = \mathrm{Diag}(\boldsymbol{h})$ is the diagonal matrix formed by the elements of $\boldsymbol{h} = (h_1, \ldots, h_n)'$. JIVE2 is consistent under many instruments asymptotics as has been shown by Ackerberg and Devereux (2003). However, Davidson and MacKinnon (2006) have shown that the jackknife estimators can have low efficiency relative to LIML under homoskedasticity.

Therefore, Hausman et al. (2012) consider jackknife versions of LIML and the Fuller (1977) estimator by using the objective function

$$\mathrm{Q}_{\mathrm{HLIM}}(\boldsymbol{\beta}) = \frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\{\boldsymbol{P} - \boldsymbol{D}\}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}. \tag{8}$$

[3] When dummy instruments indicate groups and group sizes are equal, LIML is many-instruments consistent even under heteroskedasticity.