University of Groningen

Jackknife instrumental variable estimation with heteroskedasticity

Bekker, Paulus; Crudu, Federico

Published in:
Journal of Econometrics

DOI:
[10.1016/j.jeconom.2014.08.012](https://doi.org/10.1016/j.jeconom.2014.08.012)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Bekker, P., & Crudu, F. (2015). Jackknife instrumental variable estimation with heteroskedasticity. *Journal of Econometrics*, 185(2), 332–342. <https://doi.org/10.1016/j.jeconom.2014.08.012>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

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.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The presence of unknown heteroskedasticity is a common setting in microeconomic 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), Akerberg 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.cru@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.

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 $\mathbf{y}$ and the $n \times g$ matrix $\mathbf{X}$ that satisfy

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (1)$$

$$\mathbf{X} = \mathbf{Z}\boldsymbol{\Pi} + \mathbf{V}, \quad (2)$$

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

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

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

$$\boldsymbol{\omega}_i = \begin{pmatrix} 1 & \boldsymbol{\beta}' \\ \mathbf{0} & \mathbf{I}_g \end{pmatrix} \boldsymbol{\Sigma}_i \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta} & \mathbf{I}_g \end{pmatrix}. \quad (3)$$

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

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

$$Q_{\text{LIML}}(\boldsymbol{\beta}) = \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{P}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{I}_n - \mathbf{P})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}. \quad (4)$$

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

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

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

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

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 $i = 1, \dots, n$, 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

$$Q_{\text{GMM}}(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{Z} \left\{ \sum_{i=1}^n \hat{\sigma}_i^2 \mathbf{Z}_i' \mathbf{Z}_i \right\}^{-1} \mathbf{Z}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad (5)$$

where $\hat{\sigma}_i^2 = (y_i - \mathbf{X}_i\hat{\boldsymbol{\beta}})^2$ 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 $\hat{\sigma}_i^2(\boldsymbol{\beta}) = (y_i - \mathbf{X}_i\boldsymbol{\beta})^2$. 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 $\mathbf{Z}$, 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

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\text{JIVE1}} &= (\tilde{\mathbf{X}}'\mathbf{X})^{-1}\tilde{\mathbf{X}}'\mathbf{y}, \\ \mathbf{e}_i'\tilde{\mathbf{X}} &= \tilde{\mathbf{X}}_i = \frac{\mathbf{Z}_i(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X} - h_i\mathbf{X}_i}{1 - h_i}, \end{aligned} \quad (6)$$

where $h_i = P_{ii}$, and $i = 1, \dots, n$. 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 $\tilde{\mathbf{X}} = (\mathbf{P} - \mathbf{D})\mathbf{X}$ and thus minimizes a 2SLS-like objective function

$$Q_{\text{JIVE2}}(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\{\mathbf{P} - \mathbf{D}\}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad (7)$$

where $\mathbf{D} = \text{Diag}(\mathbf{h})$ is the diagonal matrix formed by the elements of $\mathbf{h} = (h_1, \dots, h_n)'$. JIVE2 is consistent under many instruments asymptotics as has been shown by Akerberg 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

$$Q_{\text{HLIM}}(\boldsymbol{\beta}) = \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\{\mathbf{P} - \mathbf{D}\}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}. \quad (8)$$

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

The estimators are given by

$$\hat{\beta} = \{X'(P - D)X - \hat{\alpha}X'X\}^{-1} \{X'(P - D)y - \hat{\alpha}X'y\}, \quad (9)$$

$$\hat{\alpha} = \frac{(n + c)\tilde{\alpha} - c}{n + c\tilde{\alpha} - c},$$

$$\tilde{\alpha} = \lambda_{\min}[(y, X)'(y, X)]^{-1} (y, X)'(P - D)(y, X).$$

For $c = 0$, $\hat{\beta}_{\text{HLIM}}$ is found, and in particular $c = 1$ produces $\hat{\beta}_{\text{HFUL}}$. Hausman et al. (2012) consider many-instruments and many-weak-instruments asymptotics and show the asymptotic distributions are not affected by deviations from normality. The estimators perform much better than the original jackknife estimators.⁴

Kunitomo (2012) considered modifications of the LIML objective function (4) resulting in many-instruments consistent estimators under heteroskedasticity. The modification amounts to replacing the diagonal elements of P by values whose difference with k/n vanishes asymptotically. The resulting estimators have the same many-instruments asymptotic distribution as HLIM. By replacing the diagonal elements of P by values that converge to their average value the LIML modifications lose signal just as HLIM. As the asymptotic requirement does not define a unique estimator, there are many possibilities to use for finite-sample comparisons. Kunitomo (2012) claims one of these improves on HLIM in finite samples.

3. A method of moments and jackknife estimator

In order to handle heteroskedasticity the grouping estimators use data clustering. In many cases this means information will be lost in the process, although between-group heteroskedasticity is used to improve efficiency. The jackknife approach maintains original instruments to a larger extent, but seems to remove possibly relevant information on β contained in the matrix $(y, X)'D(y, X)$. As an alternative to the objective function Q_{HLIM} , we consider a method-of-moments approach that maintains the signal component in the expectation of $(y, X)'P(y, X)$ and aims at estimating the noise component consistently. Thus we try to maintain the information contained in the data to a larger extent without adding much additional noise. The estimator can be interpreted as a symmetric jackknife estimator.

In order to formulate a method-of-moments estimator similar to LIML, we consider minimizing a criterion given by

$$Q(\beta) = \frac{(y - X\beta)'A(y - X\beta)}{(y - X\beta)'B(y - X\beta)}, \quad (10)$$

where, conditional on the instruments Z , A and B are fixed positive semidefinite matrices. We find the following conditional expectations

$$E\{(y, X)'A(y, X)\} = \begin{pmatrix} \beta' \\ I_g \end{pmatrix} \Pi'Z'AZ\Pi(\beta, I_g) + \sum_{i=1}^n A_{ii}\Omega_i, \quad (11)$$

$$E\{(y, X)'B(y, X)\} = \begin{pmatrix} \beta' \\ I_g \end{pmatrix} \Pi'Z'BZ\Pi(\beta, I_g) + \sum_{i=1}^n B_{ii}\Omega_i. \quad (12)$$

The method-of-moments approach amounts to solving these equations for β . Typically this occurs for solving, using the smallest value of l ,

$$[E\{(y, X)'A(y, X)\} - lE\{(y, X)'B(y, X)\}]\begin{pmatrix} 1 \\ -\beta \end{pmatrix} = 0, \quad (13)$$

⁴ However, notice that the denominator of the objective function (8) weighs the squared errors with equal weights even if the i th row of Z has small squared elements, or even when $Z_i = 0$. If $Z = (Z_1', Z_2')'$, where Z_1 has n_1 rows and $Z_2 = 0$, then the redundant parameters in Z_i , $i = n_1 + 1, \dots, n$, affect both the finite-sample and the many-instruments asymptotic distribution of HLIM.

when the signal matrix $\Pi'Z'(A - lB)Z\Pi$ is positive definite, preferably large, and the noise component vanishes: $\sum_{i=1}^n A_{ii}\Omega_i = l \sum_{i=1}^n B_{ii}\Omega_i$.

For example, LIML uses $A = P$ and $B = I_n - P$, which allows for the solution β if all matrices Ω_i are equal.⁵ In case of heteroskedasticity, however, the noise term does not drop out which renders LIML many-instruments inconsistent.

Hausman et al. (2012) achieve consistency by using $A = P - D$, so $A_{ii} = 0$, $i = 1, \dots, n$ and $B = I_n$. That way (11) alone is enough to solve for β . By removing the diagonal from P , however, the signal contained in $(\beta, I_g)' \Pi'Z'DZ\Pi(\beta, I_g)$ is removed as well.

In our approach we maintain the signal matrix by requiring $Z'AZ = Z'Z$ and $Z'BZ = 0$ and we remove the noise component by requiring only that the diagonal of A is a scalar multiple of the diagonal of B , i.e. $l = \text{tr}(A)/\text{tr}(B)$ and $A_{ii} = lB_{ii}$, $i = 1, \dots, n$. In that case (13) is satisfied and the many-instruments consistency of the minimizer of $Q(\beta)$ would follow easily.

Without loss of generality we use $l = 1$, so A and B should have the same diagonal, and $A = P + \Delta$ where Δ satisfies $Z'\Delta Z = 0$. To stay close to LIML we would like $\sum_{i=1}^n \Delta_{ii}\Omega_i$ to be small. As $A_{ii}\Omega_i = h_i\Omega_i + \Delta_{ii}\Omega_i$, and $E\{(y, X)'B(y, X)\} = \sum_{i=1}^n A_{ii}\Omega_i$, we find $\sum_{i=1}^n \Delta_{ii}\Omega_i$ is small if $(y, X)'B(y, X)$ is a good estimator, with small bias, of $\sum_{i=1}^n h_i\Omega_i$. We formulate such an estimator for Ω_i intuitively as

$$\hat{\Omega}_i = (y, X)' \frac{(I_n - P)e_i e_i'(I_n - P)}{e_i'(I_n - P)e_i} (y, X), \quad (14)$$

which amounts to choosing

$$B = (I_n - P)D(I_n - D)^{-1}(I_n - P). \quad (15)$$

This fixes the diagonal elements of Δ as the difference between the diagonal elements of B and P .⁷ A choice for Δ with such diagonal elements and satisfying $Z\Delta Z = 0$ is given by

$$\Delta = PD(I_n - D)^{-1}P - \frac{1}{2} \{PD(I_n - D)^{-1} + D(I_n - D)^{-1}P\}. \quad (16)$$

To further motivate the choice of Δ we observe that the minimizer of $Q(\beta)$ in (10) can be interpreted as a normalized symmetric jackknife estimator. That is

$$\hat{\beta}_{\text{SJIVE}} = \arg \min_{\beta} \{Q(\beta)\} = \arg \min_{\beta} \{Q_{\text{SJIVE}}(\beta)\}, \quad (17)$$

$$Q_{\text{SJIVE}}(\beta) = \frac{(y - X\beta)'C(y - X\beta)}{(y - X\beta)'B(y - X\beta)}, \quad (18)$$

$$\begin{aligned} C &= A - B, \\ &= P - D(I_n - D)^{-1} + \{PD(I_n - D)^{-1} + (I_n - D)^{-1}DP\}/2 \\ &= (\tilde{P} + \tilde{P}')/2, \end{aligned} \quad (19)$$

$$\begin{aligned} \tilde{P} &= P - D(I_n - D)^{-1} + D(I_n - D)^{-1}P \\ &= (I_n - D)^{-1}(P - D). \end{aligned}$$

Notice the diagonal elements of C equal zero. The jackknife estimator $\beta_{\text{JIVE1}} = (\tilde{X}'X)^{-1}\tilde{X}'y$ in (6) is based on $\tilde{X} = \tilde{P}X$. So, if we define

⁵ Actually, only $\sum_{i=1}^n A_{ii}\Omega_i(1, \beta')' = l \sum_{i=1}^n B_{ii}\Omega_i(1, -\beta')'$ needs to hold, which amounts to $\sum_{i=1}^n A_{ii}\Omega_i^2 = l \sum_{i=1}^n B_{ii}\Omega_i^2$ and $\sum_{i=1}^n A_{ii}\Omega_{21i} = l \sum_{i=1}^n B_{ii}\Omega_{21i}$.

⁶ A less restrictive sufficient requirement is that the covariances between the error term ε_i and all endogenous variables do not vary.

⁷ We find $\Delta_{ii} = -h_i + \sum_{j=1}^n \frac{h_j}{1-h_j} e_i'(I_n - P)e_j e_j'(I_n - P)e_i = -h_i + \frac{h_i}{1-h_i(1-h_i)^2} + \sum_{j \neq i} \frac{h_j}{1-h_j} p_{ij}^2 = -\frac{h_i^2}{1-h_i} + \sum_{j=1}^n \frac{h_j}{1-h_j} p_{ij}^2$. As both terms on the right hand side are bounded, so is $|\Delta_{ii}| \leq h_i \max_i(h_i)/\{1 - \max_i(h_i)\}$. If $\max_i(h_i) \rightarrow 0$, then $\Delta_{ii} \rightarrow 0$, but that will not happen with many instruments if $k/n \rightarrow 0$ does not hold. Yet k/n may be small in practice.

$\tilde{\mathbf{y}} = \tilde{\mathbf{P}}\mathbf{y}$, then we find the numerator of the objective function is given by

$$\begin{aligned} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{C} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) &= \frac{1}{2} (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &\quad + \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}) \\ &= (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \end{aligned}$$

That is to say, jackknife prediction is used for all endogenous variables symmetrically, including the dependent variable. As the statistical problem is basically symmetric in the endogenous variables, it seems a good property the symmetry is maintained in the jackknifing procedure, so that the estimator is invariant with respect to the particular type of normalization.

We find the signal matrix is larger for SJIVE than for HLIM since

$$\begin{aligned} E\{(\mathbf{y}, \mathbf{X})' \mathbf{C} (\mathbf{y}, \mathbf{X})\} &= \begin{pmatrix} \boldsymbol{\beta}' \\ \mathbf{I}_g \end{pmatrix} \Pi \mathbf{Z}' \mathbf{Z} \Pi (\boldsymbol{\beta}, \mathbf{I}_g) \\ &\geq E\{(\mathbf{y}, \mathbf{X})' (\mathbf{P} - \mathbf{D}) (\mathbf{y}, \mathbf{X})\} = \begin{pmatrix} \boldsymbol{\beta}' \\ \mathbf{I}_g \end{pmatrix} \Pi \mathbf{Z}' \mathbf{Z} \Pi (\boldsymbol{\beta}, \mathbf{I}_g) \\ &\quad - \begin{pmatrix} \boldsymbol{\beta}' \\ \mathbf{I}_g \end{pmatrix} \Pi \mathbf{Z}' \mathbf{D} \mathbf{Z} \Pi (\boldsymbol{\beta}, \mathbf{I}_g). \end{aligned} \quad (20)$$

In Section 5 we find as well that the signal component in the many-instruments asymptotic covariance matrix is smaller for SJIVE than for HLIM for cases where the heteroskedasticity is not extreme.

To compute the symmetric jackknife estimator and its Fuller modifications, let $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ and $\mathbf{X}_2 = \mathbf{Z}_2$, where $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$, so the explanatory variables in $\mathbf{X}_2$ are assumed to be exogenous. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2)'$ be partitioned conformably. Let $\mathbf{C}^* = \mathbf{C} - \mathbf{A}\mathbf{X}_2'(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{A}$, then the SJIVE estimator and its Fuller modifications (SJEF) can be computed by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{C}\mathbf{X} - \hat{\lambda}\mathbf{X}'\mathbf{B}\mathbf{X})^{-1} (\mathbf{X}'\mathbf{C}\mathbf{y} - \hat{\lambda}\mathbf{X}'\mathbf{B}\mathbf{y}), \quad (21)$$

$$\hat{\lambda} = \lambda - \alpha / \text{tr}(\mathbf{B}),$$

$$\lambda = \lambda_{\min} \left[\{(\mathbf{y}, \mathbf{X}_1)' \mathbf{B} (\mathbf{y}, \mathbf{X}_1)\}^{-1} (\mathbf{y}, \mathbf{X}_1)' \mathbf{C}^* (\mathbf{y}, \mathbf{X}_1) \right].$$

For $\alpha = 0\hat{\boldsymbol{\beta}}_{\text{SJIVE}}$ is found. Based on the Monte Carlo experiments we would use a Fuller modification $\hat{\boldsymbol{\beta}}_{\text{SJEF}}$ with $\alpha = 2$. Using Theorem 1 we compute standard errors as the square root of the diagonal elements of the estimated covariance matrix, which is formulated concisely as

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\widehat{\mathbf{C}}\mathbf{X})^{-1} \tilde{\mathbf{X}}' \{ \mathbf{C}\mathbf{D}_{\hat{\boldsymbol{\beta}}}^2 \mathbf{C} + \mathbf{D}_{\hat{\boldsymbol{\beta}}} (\mathbf{C} * \mathbf{C}) \mathbf{D}_{\hat{\boldsymbol{\beta}}} \} \tilde{\mathbf{X}} (\mathbf{X}'\widehat{\mathbf{C}}\mathbf{X})^{-1}, \quad (22)$$

$$\tilde{\mathbf{X}} = \mathbf{X} - \hat{\sigma}^{-2} \hat{\boldsymbol{\epsilon}} \hat{\sigma}_{12},$$

where $\widehat{\mathbf{C}} = \mathbf{C} - \hat{\lambda}\mathbf{B}$ and $\mathbf{C} * \mathbf{C}$ is the elementwise or Hadamard product. The diagonal matrix $\mathbf{D}_{\hat{\boldsymbol{\beta}}}$ has the residuals $\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ on the diagonal. Finally, $\hat{\sigma}^2$ and $\hat{\sigma}_{21}$ are found based on $\hat{\boldsymbol{\Omega}} = (\mathbf{y}, \mathbf{X})' \mathbf{B} (\mathbf{y}, \mathbf{X}) / k$, which is transformed to $\hat{\boldsymbol{\Sigma}}$ similar to (3).

4. Asymptotic distributions

We consider many strong instruments and many weak instruments parameter sequences to describe the asymptotic distributions of the heteroskedasticity robust estimator SJIVE as given in (21). The number of instruments k_n may depend on n , but to keep the notation simple we will use k as before. Our formulation allows for the presence of both weak and strong instruments within a single model. The derivation is based on high-level regularity conditions, since primitive regularity conditions could be formulated very similar to earlier ones. For example, the ones used by Hausman et al. (2012) could be used, although our results hold more generally.

Assumption 1. The diagonal elements of the hat matrix $\mathbf{P}$ satisfy $\max_i h_i \leq 1 - 1/c_u$.

Assumption 2. The covariance matrices of the disturbances are bounded, $0 \leq \boldsymbol{\Sigma}_i \leq c_u \mathbf{I}_{g+1}$ and satisfy $k^{-1} \sum_{i=1}^n \mathbf{e}_i' \mathbf{B} \mathbf{e}_i \boldsymbol{\Sigma}_i \rightarrow \boldsymbol{\Sigma}$.

We partition $\boldsymbol{\Sigma}$ as we partitioned $\boldsymbol{\Sigma}_i$ and use $\boldsymbol{\Omega}$ defined similar to (3), where $\boldsymbol{\Sigma}$ is used instead of $\boldsymbol{\Sigma}_i$. So we find $\lim_{n \rightarrow \infty} k^{-1} E\{(\mathbf{y}, \mathbf{X})' \mathbf{B} (\mathbf{y}, \mathbf{X})\} = \boldsymbol{\Omega}$, and $E(\mathbf{X}'\mathbf{C}\mathbf{X}) = \mathbf{H}$, where $\mathbf{H}$ is the signal matrix

$$\mathbf{H} = \Pi' \mathbf{Z}' \mathbf{Z} \Pi. \quad (23)$$

Assumption 3. $\text{plim}_{n \rightarrow \infty} k^{-1} (\mathbf{y}, \mathbf{X})' \mathbf{B} (\mathbf{y}, \mathbf{X}) = \boldsymbol{\Omega}$, and $\text{plim}_{n \rightarrow \infty} \mathbf{H}^{-1} \mathbf{X}' \mathbf{C} \mathbf{X} = \mathbf{I}_g$.

Let $r_{\min} = \lambda_{\min}(\mathbf{H})$ be the smallest eigenvalue of the signal matrix and let $r_{\max}$ be the largest eigenvalue. In the original many-instruments and many-weak instruments cases of Bekker (1994) and Chao and Swanson (2005), respectively, the growth rates of $r_{\max}$ and $r_{\min}$ are the same. In the sequences of Hausman et al. (2012) and in our sequences the growth rates may vary. We both assume $r_{\min} \rightarrow \infty$.

Assumption 4. $r_{\min} \rightarrow \infty$.

However, we do not necessarily assume $r_{\max}$ is of order n at most as Hausman et al. (2012) implicitly do. We allow for very strong instruments with super consistent estimators as well. Let $Q_{\text{SJIVE}}^*(\boldsymbol{\beta}) = k Q_{\text{SJIVE}}(\boldsymbol{\beta})$, then the many-strong-instruments asymptotic approximations are based on the following high-level assumption.

Assumption 5 (Many Strong Instruments). $k/r_{\min} \rightarrow \gamma$, where $0 \leq \gamma < \infty$, and

$$\begin{aligned} \left\{ \frac{\partial^2 Q_{\text{SJIVE}}^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right\}^{1/2} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= - \left\{ \frac{\partial^2 Q_{\text{SJIVE}}^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right\}^{-1/2} \\ &\quad \times \frac{\partial Q_{\text{SJIVE}}^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + o_p(1) \stackrel{a}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Phi}_s), \end{aligned}$$

where $\boldsymbol{\Phi}_s = \boldsymbol{\Xi}^{-1/2} \boldsymbol{\Theta} \boldsymbol{\Xi}^{-1/2}$ and

$$\begin{aligned} \boldsymbol{\Xi} &= \text{plim}_{n \rightarrow \infty} \mathbf{H}^{-1/2} \frac{\partial^2 Q_{\text{SJIVE}}^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \mathbf{H}^{-1/2}, \\ \mathbf{H}^{-1/2} \frac{\partial Q_{\text{SJIVE}}^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &\stackrel{a}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Theta}), \\ \boldsymbol{\Theta} &= \lim_{n \rightarrow \infty} \text{Var} \left\{ \mathbf{H}^{-1/2} \frac{\partial Q_{\text{SJIVE}}^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + o_p(1) \right\}. \end{aligned} \quad (24)$$

The $o_p(1)$ term in (24) is defined in (41) in the Appendix. That is to say, as $\delta' \mathbf{S} \delta / \text{tr} \mathbf{B} = \sigma^2 + o_p(1)$, we find using (41) that

$$\begin{aligned} \mathbf{H}^{-1/2} \frac{\partial Q_{\text{SJIVE}}^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= - \frac{2}{\sigma^2} \mathbf{H}^{-1/2} \{ (\mathbf{0}, \mathbf{I}_g) - \sigma^{-2} \sigma_{21} \delta' \} \\ &\quad \times (\mathbf{M} - \mathbf{S}) \delta - o_p(1). \end{aligned}$$

Removing the $o_p(1)$ term leaves the vector given by the first term on right-hand side. It has a finite-sample covariance matrix. We make the mild assumption that these finite-sample covariance matrices converge to the covariance matrix of the limit distribution.

Using Assumptions 1–5 the many-instruments asymptotic distribution of SJIVE can be formulated as in Theorem 1. Its derivation is given in the Appendix. In particular (41) and (42) show that Assumption 5 boils down to the asymptotic normality of

$$\mathbf{H}^{1/2} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{H}^{-1/2} (\mathbf{Z}\Pi + \tilde{\mathbf{V}})' \mathbf{C} \boldsymbol{\epsilon} + o_p(1), \quad (25)$$

$$\tilde{\mathbf{V}} = \mathbf{V} - \boldsymbol{\epsilon} \sigma_{12} / \sigma^2. \quad (26)$$

Instead of the high-level Assumption 5, more primitive assumptions can be considered for specific cases that are sufficient for

Assumption 5 to hold true. In fact, apart from the asymptotic behavior of $r_{\max}$ our assumptions are not very different from the specification of Case I in Hausman et al. (2012). For example, consider rather similar primitive conditions, where in addition to Assumptions 1–4, the disturbances have bounded fourth-order moments. Hausman et al. (2012) use the following somewhat complicated specification, as was acknowledged by the authors (p. 220), to accommodate reduced form coefficients with identifying power that depends on n . Let

$$\mathbf{Z}\Pi = n^{-1/2} \mathbf{Z}\tilde{\Pi} \mathbf{D}_{\mu_n} \tilde{\mathbf{S}}, \quad (27)$$

where $\tilde{\mathbf{S}}$ is nonsingular, $\mathbf{D}_{\mu_n} = \text{Diag}(\mu_n)$ and μ_n is a g -vector with elements μ_{jn} that satisfy either $\mu_{jn} = \sqrt{n}$ or $\mu_{jn}/\sqrt{n} \rightarrow 0$. Furthermore, assume a bit more restrictive than Hausman et al. (2012), the existence of $\tilde{\mathbf{H}} = \lim_{n \rightarrow \infty} n^{-1} \tilde{\Pi}' \mathbf{Z}' \mathbf{Z} \tilde{\Pi} > 0$, $\lim_{n \rightarrow \infty} n^{-1} \sum_i \sigma_i^2 \tilde{\Pi}' \mathbf{Z}' \mathbf{C} \mathbf{e}_i \mathbf{e}_i' \mathbf{C} \mathbf{Z} \tilde{\Pi}$, and $\lim_{n \rightarrow \infty} k^{-1} \mathbf{G}$, where

$$\begin{aligned} \mathbf{G} &= \sum_{i=1}^n \sum_{j=1}^n C_{ij}^2 (\sigma_j^2 \tilde{\Sigma}_{i22} + \tilde{\sigma}_{i21} \tilde{\sigma}_{j21}') \\ &= \sum_{i \neq j} C_{ij}^2 (\sigma_j^2 \tilde{\Sigma}_{i22} + \tilde{\sigma}_{i21} \tilde{\sigma}_{j21}'), \end{aligned} \quad (28)$$

$$\tilde{\Sigma}_i = \text{Var}(\varepsilon_i, \tilde{\mathbf{V}}_i). \quad (29)$$

Together with the Case I assumption of Hausman et al. (2012), which amounts to $k/r_{\min} \rightarrow \gamma$, these conditions imply Assumption 5 as is further discussed in Appendix A.4. Of course these conditions are not necessary and other less restrictive specifications may be sufficient as well.

The specification (27) of Hausman et al. (2012) can be avoided since the asymptotic distribution of HLIM and SJIVE can be formulated based on the high-level Assumption 5. For SJIVE we have the following result.

Theorem 1 (Many Strong Instruments). *If Assumptions 1–5 are satisfied, then $\hat{\beta} = \hat{\beta}_{\text{SJIVE}}$ is consistent and $(\mathbf{X}'\hat{\mathbf{C}}\mathbf{X})^{1/2}(\hat{\beta} - \beta) \stackrel{a}{\sim} \mathcal{N}(\mathbf{0}, \Psi_s)$, where $\hat{\mathbf{C}} = \mathbf{C} - \hat{\lambda}\mathbf{B}$ and $\Psi_s = \lim_{n \rightarrow \infty} \mathbf{H}^{-1/2}(\mathbf{F} + \mathbf{G})\mathbf{H}^{-1/2}$, where $\mathbf{F} = \sum_{i=1}^n \sigma_i^2 \Pi' \mathbf{Z}' \mathbf{C} \mathbf{e}_i \mathbf{e}_i' \mathbf{C} \mathbf{Z} \Pi$ and $\mathbf{G}$ is defined in (28).*

To formulate a consistent estimator for Ψ_s we need a condition that guarantees that $\hat{\varepsilon}_i$ converges to ε_i uniformly in probability. This requires uniform convergence of $\|\mathbf{Z}_i \Pi(\hat{\beta} - \beta)\|$ to zero, which is guaranteed by the assumption below. Let d_i denote the i th diagonal element of the projection matrix $\mathbf{P}_{\mathbf{Z}\Pi} = \mathbf{Z}\Pi(\Pi' \mathbf{Z}' \mathbf{Z} \Pi)^{-1} \Pi' \mathbf{Z}'$.⁸

Assumption 6. $\max_{1 \leq i \leq n} d_i \rightarrow 0$.

The restricted specification (27), where $\mathbf{P}_{\mathbf{Z}\Pi} = \mathbf{P}_{\mathbf{Z}\tilde{\Pi}}$ and $\tilde{\mathbf{H}} = \lim_{n \rightarrow \infty} \tilde{\Pi}' \mathbf{Z}' \mathbf{Z} \tilde{\Pi}/n > 0$, is sufficient for Assumption 6 to hold true.

Theorem 2. *If Assumptions 1–6 are satisfied, a consistent estimator for Ψ_s is given by*

$$\begin{aligned} \tilde{\Psi}_s &= (\mathbf{X}'\hat{\mathbf{C}}\mathbf{X})^{-1/2} \tilde{\mathbf{X}}' \{ \mathbf{C} \mathbf{D}_{\hat{\varepsilon}}^2 \mathbf{C} + \mathbf{D}_{\hat{\varepsilon}} (\mathbf{C} * \mathbf{C}) \mathbf{D}_{\hat{\varepsilon}} \} \tilde{\mathbf{X}} (\mathbf{X}'\hat{\mathbf{C}}\mathbf{X})^{-1/2}, \\ \tilde{\mathbf{X}} &= (\mathbf{X} - \hat{\sigma}^{-2} \hat{\varepsilon} \hat{\sigma}_{12}) \end{aligned}$$

where $\tilde{\mathbf{X}}$, $\mathbf{D}_{\hat{\varepsilon}}$, $\hat{\varepsilon} = \mathbf{y} - \mathbf{X}\hat{\beta}$, $\mathbf{C} * \mathbf{C}$, $\hat{\sigma}^2$ and $\hat{\sigma}_{12}$ are defined as in (22).

The proofs of Theorems 1 and 2 are given in the Appendix.

The asymptotic covariance matrix Ψ_s has two terms. Under large-sample asymptotics, when $k/r_{\min} \rightarrow 0$ the second term vanishes. As the second term may be relevant in the finite sample, the

many instruments asymptotic approximation to the finite distribution is usually more accurate than the large-sample approximation as was shown by Bekker (1994). Under homoskedasticity $\tilde{\sigma}_{i21} = \mathbf{0}$, $i = 1, \dots, n$, so the second part of $\mathbf{G}$ drops out.

When instruments are weak the second term may be dominant and the first term may even be negligible. Chao and Swanson (2005) used many-weak-instruments asymptotic sequences, where k grows faster than $r_{\min}$, and showed the first term actually vanishes, when $r_{\min}$ and $r_{\max}$ have the same growth rate, while estimators such as LIML under homoskedasticity are still consistent. Hausman et al. (2012) derived a many-weak-instruments asymptotic distribution of HLIM. However, to accommodate different growth rates for $r_{\max}$ and $r_{\min}$, their formulation allows for degenerate limit distributions when explanatory variables are present that have instruments that are not weak. When there are multiple explanatory variables it will happen when the instruments are weak only with respect to a single one of them. We have a similar result, although in our formulation the asymptotic covariance matrix need not be singular as a result of the inclusion of endogenous variables with different signal strengths. In our formulation we avoid such problems by considering the asymptotic distribution of $k^{-1/2} \mathbf{X}'\hat{\mathbf{C}}\mathbf{X}(\hat{\beta} - \beta)$ which has a larger ‘blow-up factor’ $k^{-1/2} \mathbf{X}'\hat{\mathbf{C}}\mathbf{X}$ than the one considered by Hausman et al. (2012). At the same time we need to bound $r_{\max}/k$, whereas Hausman et al. (2012) only bound $r_{\max}/n$.

Assumption 7 (Many Weak Instruments). $r_{\max}/k \rightarrow \gamma$, where $0 \leq \gamma < \infty$, $r_{\min}/k \rightarrow 0$ and $k^{1/2}/r_{\min} \rightarrow 0$ and

$$k^{-1/2} \frac{\partial^2 Q_{\text{SJIVE}}^*(\beta)}{\partial \beta \partial \beta'} (\hat{\beta} - \beta) = k^{-1/2} \frac{\partial Q_{\text{SJIVE}}^*(\beta)}{\partial \beta} + o_p(1) \stackrel{a}{\sim} \mathcal{N}(\mathbf{0}, \Phi_w),$$

$$\Phi_w = \lim_{n \rightarrow \infty} \text{Var} \left\{ k^{-1/2} \frac{\partial Q_{\text{SJIVE}}^*(\beta)}{\partial \beta} + o_p(1) \right\}. \quad (30)$$

The $o_p(1)$ term in (30) is defined explicitly in (43) in the Appendix. Instead of the normality of (25), Assumption 7 boils down to the asymptotic normality of

$$\begin{aligned} k^{-1/2} \mathbf{H}(\hat{\beta} - \beta) &= k^{-1/2} (\mathbf{Z}\Pi + \tilde{\mathbf{V}})' \mathbf{C} \varepsilon + o_p(1) \\ &= k^{-1/2} \tilde{\mathbf{V}}' \mathbf{C} \varepsilon + o_p(1). \end{aligned}$$

Theorem 3 (Many Weak Instruments). *If Assumptions 1–4 and 7 are satisfied, then $\hat{\beta} = \hat{\beta}_{\text{SJIVE}}$ is consistent and $k^{-1/2} \mathbf{X}'\hat{\mathbf{C}}\mathbf{X}(\hat{\beta} - \beta) \stackrel{a}{\sim} \mathcal{N}(\mathbf{0}, \Psi_w)$, where $\Psi_w = \lim_{n \rightarrow \infty} k^{-1}(\mathbf{F} + \mathbf{G})$, and $\mathbf{F}$ and $\mathbf{G}$ are defined in Theorem 1.*

Notice that the signal component $\lim_{n \rightarrow \infty} k^{-1} \mathbf{F}$ is singular since $r_{\min}/k \rightarrow 0$. If $\gamma = 0$ in Assumption 7, then $\lim_{n \rightarrow \infty} k^{-1} \mathbf{F} = \mathbf{0}$, so the signal component vanishes completely. Also notice that the rank of the noise component $\lim_{n \rightarrow \infty} k^{-1} \mathbf{G}$ is not affected by the inclusion of endogenous variables with different signal strengths. For the actual computation of standard errors the many weak instruments asymptotic distribution is not needed, since the many-instruments standard errors remain consistent.

5. A comparison of asymptotic distributions

Comparing many-instruments asymptotic distributions usually does not give a clear-cut ordering of asymptotic covariance matrices with positive semidefinite differences. Many-instrument asymptotic distributions are good for detecting inconsistency and to formulate standard errors, but to use them for efficiency comparisons would be too restrictive. Only in restrictive classes such efficiency exists. Typically such classes, as in Kunitomo (2012), exclude relevant alternatives, which makes them less useful for making relevant comparisons. For example, contrary to large-sample

⁸ Notice that assuming $\lim_{n \rightarrow \infty} \Pi' \mathbf{Z}' \mathbf{Z} \Pi/n \rightarrow \mathbf{C} > 0$, would imply $\max_{1 \leq i \leq n} d_i \rightarrow 0$.

asymptotic approximations, there is no guarantee that using more instruments increases asymptotic efficiency. Dropping a weak instrument may well increase efficiency. Consequently, it might happen that estimators that treat high-signal instruments poorly perform quite well when instruments have low signal strength. In parts of the high-dimensional parameter space one estimator may perform well, while in other parts other estimators dominate.

Under homoskedasticity Anderson et al. (2010) described a class of estimators for which LIML is many-instruments asymptotically efficient. However, in an interesting example Hausman et al. (2012, p. 224) show that under homoskedasticity LIML is not many-instruments asymptotically efficient relative to HLIM. Similar results can be found in Van der Ploeg and Bekker (1995), Hahn (2002), and Chioda and Jansson (2009). With unknown heteroskedasticity things become more complicated, since instruments and variances may covary in ways that affect the asymptotic distributions substantially. Under heteroskedasticity Kunitomo (2012) describes a class for which a modified LIML estimator is asymptotically optimal. However, similar to the case of homoskedasticity, it is rather restrictive as it excludes relevant alternatives. Interestingly, under homoskedasticity LIML is not many-instruments asymptotically efficient relative to its heteroskedastic modification. So, the property of asymptotic efficiency as it is commonly required with large-sample asymptotic theory is simply too strong for inferential comparisons based on many-instruments asymptotic theory.

In addition to these difficulties, the accuracy of the asymptotic approximation to the finite-sample distributions is relevant as well. For example, the continuously updated GMM estimator may perform quite well when compared to the other estimators in terms of asymptotic variances, but its actual performance in the finite sample shows a different picture. So comparing the performance of the estimators and the accuracy of the asymptotic approximations in Monte Carlo simulations remains important.

Before turning to the Monte Carlo experiments, however, we can make a comparison of the signal components in the many-instruments asymptotic covariance matrices of SJIVE and HLIM. To make a comparison we assume the heteroskedasticity is not extreme so that a positive definite difference found under homoskedasticity is present as well in a neighborhood of moderate heteroskedasticity. Under homoskedasticity and normality we compare the asymptotic distributions of LIML, SJIVE and HLIM. We do need to distinguish between strong and weak instruments asymptotics since the asymptotic approximations to the finite-sample distributions are the same.

For both SJIVE and HLIM the many-instruments asymptotic approximation to the finite distribution of the estimators can be formulated as a normal distribution with mean β and covariance matrix $H^{-1}(C^*) \{F(C^*) + G(C^*, \Sigma^*)\} H^{-1}(C^*)$ where

$$H(C^*) = \Pi'Z'C^*Z\Pi, \quad (31)$$

$$F(C^*) = \sum_{i=1}^n \sigma_i^2 \Pi'Z'C^*e_i e_i' C^*Z\Pi, \quad (32)$$

$$G(C^*, \Sigma^*) = \sum_{i=1}^n \sum_{j=1}^n C_{ij}^{*2} (\sigma_j^2 \tilde{\Sigma}_{i22} + \tilde{\sigma}_{i21} \tilde{\sigma}_{j21}'), \quad (33)$$

$$\tilde{V}_j = V_j - \varepsilon_j \sigma_{12}^* / \sigma^{2*},$$

and the submatrices of $\tilde{\Sigma}_i$ have been computed as in (29). For SJIVE we have $C^* = C_{SJIVE} = C$, and $\Sigma^* = \Sigma$ as defined in (19) and Assumption 2, respectively. For HLIM $C^* = C_{HLIM} = P - D$ and $\Sigma^* = n^{-1} \sum_{i=1}^n \Sigma_i$ should be used. Furthermore, when we assume homoskedasticity and normality, the many-instruments asymptotic distribution for LIML is found for $C^* = C_{LIML} = P - k(n - k)^{-1}(I_n - P)$.

The signal matrices $H(C^*)$ for LIML, SJIVE and HLIM can be ordered. Due to $PZ = Z$ and $CZ = \{I_n - (1/2)(I_n - P)(I_n - D)^{-1}\}Z$, we find $Z'CZ = Z'Z$ and so

$$\begin{aligned} H_{LIML} &= \Pi'Z'C_{LIML}Z\Pi = \Pi'Z'Z\Pi = \\ H_{SJIVE} &= \Pi'Z'C_{SJIVE}Z\Pi = \Pi'Z'CZ\Pi = \Pi'Z'Z\Pi > \\ H_{HLIM} &= \Pi'Z'C_{HLIM}Z\Pi \\ &= \Pi'Z'(P - D)Z\Pi = \Pi'Z'Z\Pi - \Pi'Z'DZ\Pi. \end{aligned}$$

Under homoskedasticity the signal components $H^{-1}(C^*)F(C^*)$ are given by

$$\begin{aligned} H_{LIML}^{-1} F_{LIML} H_{LIML}^{-1} &= \sigma^2 (\Pi'Z'Z\Pi)^{-1}, \\ H_{SJIVE}^{-1} F_{SJIVE} H_{SJIVE}^{-1} &= \sigma^2 (\Pi'Z'Z\Pi)^{-1} \Pi'Z'C_{SJIVE}^2 Z\Pi (\Pi'Z'Z\Pi)^{-1}, \\ H_{HLIM}^{-1} F_{HLIM} H_{HLIM}^{-1} &= \sigma^2 \{\Pi'Z'(I_n - D)Z\Pi\}^{-1} \Pi'Z'(I_n - D)^2 \\ &\quad \times Z\Pi \{\Pi'Z'(I_n - D)Z\Pi\}^{-1}. \end{aligned}$$

As

$$\begin{aligned} Z'C_{SJIVE}^2 Z &= Z'Z + (1/4)Z'(I_n - D)^{-1}(I_n - P)(I_n - D)^{-1}Z \\ &\geq Z'Z, \end{aligned} \quad (34)$$

and

$$\begin{aligned} (\Pi'Z'Z\Pi)^{-1} &\leq \{\Pi'Z'(I_n - D)Z\Pi\}^{-1} \Pi'Z'(I_n - D)^2 \\ &\quad \times Z\Pi \{\Pi'Z'(I_n - D)Z\Pi\}^{-1}, \end{aligned}$$

we find the signal component in the asymptotic covariance matrix of LIML is smaller than the signal components of both SJIVE and HLIM. When the diagonal elements h_i of P are not too big, the signal components of SJIVE and HLIM are ordered as well, as follows from Theorem 4, which is proven in Appendix A.5.

Theorem 4. Let the difference between the signal components in the many-instruments asymptotic covariance matrices of HLIM and SJIVE under homoskedasticity be given by $\sigma^2 Y$, where $Y = H^{-1}(C^*) \Pi'Z'C^{*2} Z\Pi H^{-1}(C^*) - H^{-1}(C) \Pi'Z'C^2 Z\Pi H^{-1}(C)$, and $C^* = P - D$. If $D < I_n/2$ then $Y \geq 0$ and $\text{rank}(Y) = \text{rank}\{(I_n - P_{Z\Pi})DZ\Pi\}$, where $P_{Z\Pi} = Z\Pi(\Pi'Z'Z\Pi)^{-1}\Pi'Z'$.

Under homoskedasticity and normality, when $D < I_n/2$, the signal components in the asymptotic variances of LIML, SJIVE and HLIM range from small to large.⁹ Under heteroskedasticity LIML becomes inconsistent in general. SJIVE and LIML remain consistent and if the difference between their signal components is positive definite, the ranking of the signal components also holds when heteroskedasticity is sufficiently moderate.

For the noise components in the asymptotic covariance matrices of LIML, SJIVE and HLIM there does not seem to be a clear ordering. Under homoskedasticity the matrices $G(C^*, \Sigma^*) = R \sum_{i=1}^n \sum_{j=1}^n C_{ij}^{*2}$ are scalar multiples of the covariance matrix $R = \sigma^2 \tilde{\Sigma}_{22} + \tilde{\sigma}_{21} \tilde{\sigma}_{21}'$. We find

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n C_{LIML,ij}^2 &= \frac{nk}{n-k}, \\ \sum_{i=1}^n \sum_{j=1}^n C_{SJIVE,ij}^2 &= \sum_{i \neq j} C_{ij}^2 \\ &= \sum_{i \neq j} P_{ij}^2 \left(1 + \frac{h_i}{2(1-h_i)} + \frac{h_j}{2(1-h_j)} \right)^2 \\ &> \sum_{i \neq j} P_{ij}^2, \\ \sum_{i=1}^n \sum_{j=1}^n C_{HLIM,ij}^2 &= \sum_{i \neq j} P_{ij}^2 = k - \sum_{i=1}^n h_i^2 \leq k - \sum_{i=1}^n \frac{k^2}{n^2} \\ &= \frac{k(n-k)}{n} < \frac{nk}{n-k}, \end{aligned}$$

⁹ The restriction $D < I_n/2$, which implies $\max_i P_{ii} < 1/2$ is not very restrictive since in most applications the diagonal elements of P are of order k/n . If the instruments are poorly balanced and the restriction is violated there is ground to doubt the accuracy of all asymptotic approximations.

Fig. 1. $R^2_{\varepsilon^2|z} = 0$: Median bias against the Nine decile range of HFUL with $c = 0, 1, 2, 3, 4, 5$ from right to left, and SJEF for $\alpha = 0, 1, 2, 3, 4, 5$ from right to left, based on 20,000 replications.

so $\sum_i \sum_j C^2_{\text{HLIM},ij}$ is smaller than both $\sum_i \sum_j C^2_{\text{SJIVE},ij}$ and $\sum_i \sum_j C^2_{\text{LIML},ij}$.¹⁰ However, that does not imply that the noise component in the asymptotic covariance matrix of HLIM, $\mathbf{H}^{-1}_{\text{HLIM}} \mathbf{R} \mathbf{H}^{-1}_{\text{HLIM}} \sum_i \sum_j C^2_{\text{HLIM},ij}$, is smaller than the noise components of either SJIVE or LIML. As noted by Hausman et al. (2012) with respect to LIML and HLIM, the differences will disappear if $k/n \rightarrow 0$ and $\max_{1 \leq i \leq n} h_i \rightarrow 0$, which holds for SJIVE as well.

We thus find, under homoskedasticity and normality, that although the signal components in the asymptotic covariance matrices favor LIML over SJIVE and HLIM, there is no clear ordering in terms of many-instruments asymptotic efficiency. Indeed Hausman et al. (2012) showed LIML is not efficient relative to HLIM. Still, if normality and homoskedasticity are guaranteed, we would prefer to use LIML. Similarly, under mild heteroskedasticity, when LIML is inconsistent, the signal components in the asymptotic covariance matrices favor SJIVE over HLIM.

6. Monte Carlo simulations

We compare the finite sample properties of the HLIM and SJIVE and their Fuller modifications given by (9) and (21), respectively. We use the same Monte Carlo setup as Hausman et al. (2012). The data generating process is given by $\mathbf{y} = \gamma + \mathbf{x}\beta + \varepsilon$ and $\mathbf{x} = \mathbf{z}\pi + \mathbf{v}$, where $n = 800$, $\gamma = \beta = 0$. The strength of the instruments is varied by using two values $\pi = 0.1$ or $\pi = 0.2$, so that $\mu^2 = n\pi^2 = 8$ and $\mu^2 = 32$, respectively. Furthermore, $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ and independently $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. The disturbances ε are generated by

$$\varepsilon = \mathbf{v}\rho + \sqrt{\frac{1-\rho^2}{\phi^2 + \psi^4}}(\phi\mathbf{w}_1 + \psi\mathbf{w}_2),$$

where $\rho = 0.3$, $\psi = 0.86$ and conditional on $\mathbf{z}$, independent of $\mathbf{v}$, $\mathbf{w}_1 \sim \mathcal{N}(\mathbf{0}, \text{Diag}(\mathbf{z}^2))$ and $\mathbf{w}_2 \sim \mathcal{N}(\mathbf{0}, \psi^2 \mathbf{I}_n)$. The values $\phi = 0$ and $\phi = 1.38072$ are chosen such that the R-squared between ε^2_i

and the instruments equals 0 and 0.2, respectively.¹¹ The instruments $\mathbf{Z}$ are given for $k = 2$, $k = 5$ and $k = 15$ by matrices with rows $(1, z_i)$, $(1, z_i, z_i^2, z_i^3, z_i^4)$ and $(1, z_i, z_i^2, z_i^3, z_i^4, z_i b_{1i}, \dots, z_i b_{10i})$, respectively, where independent of other random variables, the elements $b_{1i}, \dots, b_{10i}$ are i.i.d. Bernoulli distributed with $p = 1/2$. We used 20,000 simulations of estimates of β .

Fig. 1 plots the nine decile ranges – between the 5th and 95th percentiles – and the median bias of Fuller modifications HFUL for $c = 0, 1, 2, 3, 4, 5$, and SJEF for $\alpha = 0, 1, 2, 3, 4, 5$, when $R^2_{\varepsilon^2|z} = 0$. As observed by Hausman et al. (2012), LIML is many-instruments consistent for this case and no big differences were found between HLIM and LIML. Here we see that the HFUL and SJEF estimators are very similar as well and the differences are due mainly to the degree of Fullerization.

When $R^2_{\varepsilon^2|z} = 0.2$ this situation changes. Table 1 compares the outcomes for HFUL when $c = 1$ and SJEF when $\alpha = 2$. We see that SJEF dominates HFUL in terms of median bias and nine decile range. The rejection rates of SJEF are smaller than the ones found for HFUL, indicating that confidence sets based on SJEF are more conservative. Fig. 2 plots the median bias and nine-decile ranges for all Fullerizations when $R^2_{\varepsilon^2|z} = 0.2$. We find SJEF performs better for this setup than HFUL. That is to say, HFUL with $c = 1$ is dominated by SJEF with $\alpha = 2$. Other choices of these coefficients result in estimators with varying location and spread characteristics. When instruments are weak Fullerization is useful since allowing for some bias reduces the spread considerably. We find that this trade-off is better for SJEF than for HFUL.

Concerning the accuracy of the asymptotic approximations we find that in particular HFUL is more spread out than its asymptotic approximation if $k > 2$.¹² This holds for both strong and weak instruments. Fig. 3 gives QQ plots for the heteroskedastic case, $R^2_{\varepsilon^2|z} = 0.2$, with weak instruments, $\mu^2 = 8$. It plots quantiles of the empirical distributions of HFUL with $c = 1$ and SJEF with $\alpha =$

¹¹ $R^2_{\varepsilon^2|z} = \text{var}\{E(\varepsilon^2|z)\} / [\text{var}\{E(\varepsilon^2|z)\} + E\{\text{var}(\varepsilon^2|z)\}]$.

¹² Kunitomo (2012) noticed the difference between the finite-sample distribution of HLIM and its asymptotic approximation as well.

¹⁰ We thank an anonymous referee for pointing to the fact that $\sum_{i \neq j} p^2_{ij} < \sum_{i \neq j} c^2_{ij}$.

Fig. 2. $R^2_{\varepsilon^2|Z} = 0.2$: Median bias against the Nine decile range of HFUL with $c = 0, 1, 2, 3, 4, 5$ from right to left, and SJEF for $\alpha = 0, 1, 2, 3, 4, 5$ from right to left, based on 20,000 replications.

Fig. 3. $R^2_{\varepsilon^2|Z} = 0.2$, $\mu^2 = 8$: QQ plots for quantiles of HFUL with $c = 1$ and SJEF with $\alpha = 2$ based on 5000 replications against quantiles of their asymptotic normal approximations.

2, based on 5000 replications, against quantiles of the normal distribution with a variance given by element (2, 2) of the covariance matrix $\{H(C^*) + f\Sigma^*\}^{-1}\{F(C^*) + G(C^*, \Sigma^*)\}\{H(C^*) + f\Sigma^*\}^{-1}$, where the matrix functions have been defined in (31)–(33), and for SJEF we used $C^* = C$, $\Sigma^* = \Sigma$ and $f = \alpha = 2$. For HFUL we used $C^* = P - D$, $\Sigma^* = n^{-1} \sum_{i=1}^n \Sigma_i$ and $f = nc/(n - c)$, where $c = 1$. For the choice $c = \alpha = 0$ the asymptotic normal approximation has a larger variance but still the empirical distribution of

HFUL has fatter tails than the normal. For SJEF we find the asymptotic approximation is more accurate.

7. Conclusion

We considered instrumental variable estimation that is robust against heteroskedasticity. A new estimator has been based on a method-of-moments reasoning and interpreted as a symmetric

Table 1

$R^2_{\varepsilon^2|z} = 0.2$: Median bias, Nine decile range and 5% Rejection rates for HFUL ($c = 1$) and SJEF ($\alpha = 2$) based on 20,000 replications.

μ^2	k	Median bias		Nine decile range		Rejection rates	
		HFUL	SJEF	HFUL	SJEF	HFUL	SJEF
8	2	0.071	0.067	1.484	1.393	0.033	0.026
8	5	0.093	0.058	2.211	1.959	0.042	0.021
8	15	0.104	0.069	2.853	2.464	0.044	0.030
32	2	0.017	0.016	0.850	0.846	0.048	0.043
32	5	0.022	0.011	0.972	0.947	0.045	0.036
32	15	0.018	0.009	1.171	1.099	0.047	0.039

jackknife estimator. It preserves the signal component in the data. Asymptotic theory based on high level assumptions, which allow for both many strong instruments and many weak instruments, resulted in a concise formulation of non-degenerate asymptotic distributions and standard errors. We found a smaller signal component in the asymptotic variance when compared to estimators that neglect a part of the signal to achieve consistency. For the noise component such ordering was not found. Sufficient primitive conditions were given as well. A Monte Carlo comparison with the HFUL estimator of Hausman et al. (2012) showed the Fuller modification of symmetric jackknife estimator performs better in terms of trade-off between bias and spread. It showed as well the asymptotic approximation of the finite-sample distribution of the symmetric jackknife estimator is more accurate.

Acknowledgments

Financial support from Marie Curie Excellence Grant MEXT-CT-2006-042471 and Regione Autonoma della Sardegna Master and Back Grant PRR-MAB-A2011-24192 is gratefully acknowledged.

Appendix

A.1. Derivation of Theorem 1

To derive Theorem 1 we use the notation $\delta = (1, -\beta')'$, $\mathbf{M} = (\mathbf{y}, \mathbf{X})' \mathbf{A}(\mathbf{y}, \mathbf{X})$ and $\mathbf{S} = (\mathbf{y}, \mathbf{X})' \mathbf{B}(\mathbf{y}, \mathbf{X})$. So $(\mathbf{y}, \mathbf{X})' \mathbf{C} \varepsilon = (\mathbf{M} - \mathbf{S})\delta$. We find

$$E(\mathbf{M} - \mathbf{S})\delta = \mathbf{0}, \quad (35)$$

$$\text{Var}\{(\mathbf{M} - \mathbf{S})\delta\} = E\{(\mathbf{y}, \mathbf{X})' \mathbf{C}(\mathbf{y}, \mathbf{X}) \delta \delta' (\mathbf{y}, \mathbf{X})' \mathbf{C}(\mathbf{y}, \mathbf{X})\}$$

$$\begin{aligned} &= E \sum_{i=1}^n \sum_{j=1}^n (\mathbf{y}, \mathbf{X})' \mathbf{C} \varepsilon_i \varepsilon_j' (\mathbf{y}, \mathbf{X}) \delta \delta' (\mathbf{y}, \mathbf{X})' \varepsilon_j \varepsilon_i' \mathbf{C}(\mathbf{y}, \mathbf{X}) \\ &= E \sum_i \varepsilon_i^2 (\mathbf{y}, \mathbf{X})' \mathbf{C} \varepsilon_i \varepsilon_i' \mathbf{C}(\mathbf{y}, \mathbf{X}) \\ &\quad + E \sum_{i=1}^n \sum_{j=1}^n \varepsilon_i \varepsilon_j C_{ij}^2 (\mathbf{y}, \mathbf{X})' \varepsilon_j \varepsilon_i' (\mathbf{y}, \mathbf{X}), \end{aligned} \quad (36)$$

$$\begin{aligned} &= \sum_{i=1}^n \sigma_i^2 \begin{pmatrix} \beta' \\ \mathbf{I}_g \end{pmatrix} \Pi' \mathbf{Z}' \mathbf{C} \varepsilon_i \varepsilon_i' \mathbf{C} \mathbf{Z} \Pi(\beta, \mathbf{I}_g) \\ &\quad + \sum_{i=1}^n \sum_{j=1}^n C_{ij}^2 (\sigma_j^2 \Omega_i + \Omega_i \delta \delta' \Omega_j), \end{aligned} \quad (37)$$

where we use the notation $C_{ij}^2 = (C_{ij})^2$. Using Assumption 2 we find $\Omega_i \leq (1 + \delta' \delta) c_u^2 \mathbf{I}_{g+1}$ and

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n C_{ij}^2 (\sigma_j^2 \Omega_i + \Omega_i \delta \delta' \Omega_j) &\leq 2c_u^2 (1 + \delta' \delta) \sum_{i=1}^n \sum_{j=1}^n C_{ij}^2 \mathbf{I}_{g+1} \\ &= 2c_u^2 (1 + \delta' \delta) \text{tr}(\mathbf{C}^2) \mathbf{I}_{g+1}, \end{aligned}$$

which is of order $O(k)$, since $C_{ii} = 0$ and, for $i \neq j$,

$$C_{ij}^2 = P_{ij}^2 \{(1 - h_i)^{-1} + (1 - h_j)^{-1}\}^2 \leq 4c_u^2 P_{ij}^2 \quad (38)$$

for $i, j = 1, \dots, n$ by Assumption 1 and $\text{tr}(\mathbf{P}^2) = k$. Consequently,

$$\delta'(\mathbf{M} - \mathbf{S})\delta = O_p(k^{1/2}). \quad (39)$$

For the first term we find

$$\begin{aligned} \sum_{i=1}^n \sigma_i^2 \begin{pmatrix} \beta' \\ \mathbf{I}_g \end{pmatrix} \Pi' \mathbf{Z}' \mathbf{C} \varepsilon_i \varepsilon_i' \mathbf{C} \mathbf{Z} \Pi(\beta, \mathbf{I}_g) &\leq c_u \begin{pmatrix} \beta' \\ \mathbf{I}_g \end{pmatrix} \Pi' \mathbf{Z}' \mathbf{C}^2 \mathbf{Z} \Pi(\beta, \mathbf{I}_g) \\ &= c_u \begin{pmatrix} \beta' \\ \mathbf{I}_g \end{pmatrix} \Pi' \mathbf{Z}' \left\{ \mathbf{I}_n + \frac{1}{4} \mathbf{D}^2 (\mathbf{I}_n - \mathbf{D})^{-2} \right\} \mathbf{Z} \Pi(\beta, \mathbf{I}_g) \\ &\leq c_u \left(1 + \frac{c_u^2}{4} \right) \begin{pmatrix} \beta' \\ \mathbf{I}_g \end{pmatrix} \Pi' \mathbf{Z}' \mathbf{Z} \Pi(\beta, \mathbf{I}_g), \end{aligned}$$

so by Assumption 5, where $k/r_{\min} \rightarrow \gamma$ we find, using $\mathbf{H} = \Pi' \mathbf{Z}' \mathbf{Z} \Pi$ as in (23),

$$\mathbf{H}^{-1/2}(\mathbf{0}, \mathbf{I}_g)(\mathbf{M} - \mathbf{S})\delta = O_p(1). \quad (40)$$

The first derivative of the objective function is given by

$$\begin{aligned} \frac{\partial Q_{\text{SJIVE}}^*(\beta)}{\partial \beta} &= -2 \left\{ \frac{\delta' \mathbf{S} \delta}{\text{tr}(\mathbf{B})} \right\}^{-1} (\mathbf{0}, \mathbf{I}_g) \left\{ \mathbf{M} \delta - \left(\frac{\delta' \mathbf{M} \delta}{\delta' \mathbf{S} \delta} \right) \mathbf{S} \delta \right\} \\ &= -2 \left\{ \frac{\delta' \mathbf{S} \delta}{\text{tr}(\mathbf{B})} \right\}^{-1} (\mathbf{0}, \mathbf{I}_g) \left(\mathbf{I}_{g+1} - \frac{\mathbf{S} \delta \delta'}{\delta' \mathbf{S} \delta} \right) (\mathbf{M} - \mathbf{S}) \delta. \end{aligned}$$

Using Assumptions 3 and 5 we find,

$$\begin{aligned} &- \frac{1}{2} \left\{ \frac{\delta' \mathbf{S} \delta}{\text{tr}(\mathbf{B})} \right\} \mathbf{H}^{-1/2} \frac{\partial Q_{\text{SJIVE}}^*(\beta)}{\partial \beta} \\ &= \mathbf{H}^{-1/2} \{ (\mathbf{0}, \mathbf{I}_g) - \sigma^{-2} \sigma_{21} \delta' \} (\mathbf{M} - \mathbf{S}) \delta \\ &\quad - \left(\frac{k}{r_{\min}} \right)^{1/2} \left(\frac{\mathbf{H}}{r_{\min}} \right)^{-1/2} \left\{ \frac{(\mathbf{0}, \mathbf{I}_g) \mathbf{S} \delta}{\delta' \mathbf{S} \delta} - \frac{\sigma_{21}}{\sigma^2} \right\} \\ &\quad \times k^{-1/2} \delta' (\mathbf{M} - \mathbf{S}) \delta \\ &= \mathbf{H}^{-1/2} \{ (\mathbf{0}, \mathbf{I}_g) - \sigma^{-2} \sigma_{21} \delta' \} (\mathbf{M} - \mathbf{S}) \delta + o_p(1). \end{aligned} \quad (41)$$

The second derivative of the objective function is given by

$$\begin{aligned} \frac{\partial^2 Q_{\text{SJIVE}}^*(\beta)}{\partial \beta \partial \beta'} &= 2 \left\{ \frac{\delta' \mathbf{S} \delta}{\text{tr}(\mathbf{B})} \right\}^{-1} (\mathbf{0}, \mathbf{I}_g) \left(2 \frac{\mathbf{S} \delta \delta'}{\delta' \mathbf{S} \delta} - \mathbf{I}_{g+1} \right) \\ &\quad \times \left\{ \mathbf{M} - \left(\frac{\delta' \mathbf{M} \delta}{\delta' \mathbf{S} \delta} \right) \mathbf{S} \right\} \times \left(2 \frac{\delta \delta' \mathbf{S}}{\delta' \mathbf{S} \delta} - \mathbf{I}_{g+1} \right) \begin{pmatrix} \mathbf{0} \\ \mathbf{I}_g \end{pmatrix} \\ &= 2 \left\{ \frac{\delta' \mathbf{S} \delta}{\text{tr}(\mathbf{B})} \right\}^{-1} (\mathbf{0}, \mathbf{I}_g) (\mathbf{M} - \mathbf{S} + \mathbf{R}) \begin{pmatrix} \mathbf{0} \\ \mathbf{I}_g \end{pmatrix}, \end{aligned}$$

where

$$\begin{aligned} \mathbf{R} &= -2(\mathbf{M} - \mathbf{S}) \delta \frac{\delta' \mathbf{S}}{\delta' \mathbf{S} \delta} - 2 \frac{\mathbf{S} \delta}{\delta' \mathbf{S} \delta} \delta' (\mathbf{M} - \mathbf{S}) \\ &\quad - \delta' (\mathbf{M} - \mathbf{S}) \delta \left(\frac{\mathbf{S}}{\delta' \mathbf{S} \delta} - 4 \frac{\mathbf{S} \delta \delta' \mathbf{S}}{(\delta' \mathbf{S} \delta)^2} \right). \end{aligned}$$

Due to (39) and (40) we find

$$\mathbf{H}^{-1/2}(\mathbf{0}, \mathbf{I}_g) \mathbf{R} \begin{pmatrix} \mathbf{0} \\ \mathbf{I}_g \end{pmatrix} \mathbf{H}^{-1/2} = o_p(1).$$

Consequently

$$\begin{aligned} &\frac{1}{2} \left\{ \frac{\delta' \mathbf{S} \delta}{\text{tr}(\mathbf{B})} \right\} \mathbf{H}^{-1/2} \frac{\partial^2 Q_{\text{SJIVE}}^*(\beta)}{\partial \beta \partial \beta'} \mathbf{H}^{-1/2} \\ &= \mathbf{H}^{-1/2}(\mathbf{0}, \mathbf{I}_g) \{ \mathbf{M} - \mathbf{S} \} \begin{pmatrix} \mathbf{0} \\ \mathbf{I}_g \end{pmatrix} \mathbf{H}^{-1/2} + o_p(1). \end{aligned}$$

Based on [Assumption 3](#) we thus find

$$\begin{aligned} \mathbf{H}^{-1/2} \frac{\partial^2 Q_{\text{SJIVE}}^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \mathbf{H}^{-1/2} &= 2\sigma^{-2} \left\{ \mathbf{H}^{-1/2} \mathbf{X}' \mathbf{C} \mathbf{X} \mathbf{H}^{-1/2} \right\} + o_p(1) \\ &= 2\sigma^{-2} \left\{ \mathbf{H}^{-1/2} \mathbf{E}(\mathbf{X}' \mathbf{C} \mathbf{X}) \mathbf{H}^{-1/2} \right\} + o_p(1) \\ &= 2\sigma^{-2} \mathbf{I}_g + o_p(1). \end{aligned} \quad (42)$$

So, $\boldsymbol{\Xi} = 2\sigma^{-2} \mathbf{I}_g$. [Assumption 4](#) says $r_{\min} \rightarrow \infty$, so (42) implies that $\lambda_{\min} \left(\frac{\partial^2 Q_{\text{SJIVE}}^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right) \rightarrow \infty$, and by [Assumption 5](#) we find $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$. Finally we find, applying (37), (41) and (42),

$$\begin{aligned} \boldsymbol{\Psi}_s &= \lim_{n \rightarrow \infty} \left[\mathbf{H}^{-1/2} \left\{ \sum_{i=1}^n \sigma_i^2 \boldsymbol{\Pi}' \mathbf{Z}' \mathbf{C} \mathbf{e}_i \mathbf{e}_i' \mathbf{C} \mathbf{Z} \boldsymbol{\Pi} \right. \right. \\ &\quad + \sum_{i=1}^n \sum_{j=1}^n C_{ij}^2 \left(-\frac{\sigma_{21}}{\sigma^2}, \mathbf{I}_g \right) (\sigma_j^2 \boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_i \mathbf{e}_1 \mathbf{e}_1' \boldsymbol{\Sigma}_j) \\ &\quad \left. \left. \times \left(-\frac{\sigma_{21}}{\sigma^2}, \mathbf{I}_g \right)' \right\} \mathbf{H}^{-1/2} \right]. \end{aligned}$$

A.2. Derivation of [Theorem 2](#)

To compute standard errors, we estimate $\text{Var}\{(\mathbf{M} - \mathbf{S})\delta\}$ using (36). As $\hat{\varepsilon}_i = \varepsilon_i - \mathbf{Z}_i \boldsymbol{\Pi}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \mathbf{V}_i(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and $\mathbf{H}^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = O_p(1)$, $r_{\min} \rightarrow \infty$ and $\boldsymbol{\Omega}_i = O(1)$, we may use $\hat{\varepsilon}_i$ instead of ε_i if $\max_{1 \leq i \leq n} \|\mathbf{Z}_i \boldsymbol{\Pi} \mathbf{H}^{-1/2}\| = o(1)$, which is guaranteed by [Assumption 6](#). A consistent estimator for $\boldsymbol{\Psi}_s$ is then given by

$$\begin{aligned} \widehat{\boldsymbol{\Psi}}_s &= (\mathbf{X}' \widehat{\mathbf{C}} \mathbf{X})^{-1/2} \left\{ (\mathbf{0}, \mathbf{I}_g) - \hat{\sigma}^{-2} \hat{\sigma}_{21} \hat{\boldsymbol{\delta}}' \right\} \widehat{\text{Var}}[(\mathbf{M} - \mathbf{S})\delta] \\ &\quad \times \left\{ (\mathbf{0}, \mathbf{I}_g)' - \hat{\sigma}^{-2} \hat{\boldsymbol{\delta}} \hat{\sigma}_{12} \right\} (\mathbf{X}' \widehat{\mathbf{C}} \mathbf{X})^{-1/2}, \end{aligned}$$

where

$$\begin{aligned} \widehat{\text{Var}}[(\mathbf{M} - \mathbf{S})\delta] &= \sum_i \hat{\varepsilon}_i^2(\mathbf{y}, \mathbf{X}) \widehat{\mathbf{C}} \mathbf{e}_i \mathbf{e}_i' \widehat{\mathbf{C}}(\mathbf{y}, \mathbf{X}) \\ &\quad + \sum_{i=1}^n \sum_{j=1}^n \hat{\varepsilon}_i \hat{\varepsilon}_j (\widehat{\mathbf{C}}_{ij})^2(\mathbf{y}, \mathbf{X})' \mathbf{e}_j \mathbf{e}_j'(\mathbf{y}, \mathbf{X}). \end{aligned}$$

The estimated covariance matrix for $\hat{\boldsymbol{\beta}}$ is given by

$$\begin{aligned} \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}' \widehat{\mathbf{C}} \mathbf{X})^{-1} \left\{ (\mathbf{0}, \mathbf{I}_g) - \hat{\sigma}^{-2} \hat{\sigma}_{21} \hat{\boldsymbol{\delta}}' \right\} \widehat{\text{Var}}[(\mathbf{M} - \mathbf{S})\delta] \\ &\quad \times \left\{ (\mathbf{0}, \mathbf{I}_g)' - \hat{\sigma}^{-2} \hat{\boldsymbol{\delta}} \hat{\sigma}_{12} \right\} (\mathbf{X}' \widehat{\mathbf{C}} \mathbf{X})^{-1} \\ &= (\mathbf{X}' \widehat{\mathbf{C}} \mathbf{X})^{-1} (\mathbf{X} - \hat{\sigma}^{-2} \hat{\boldsymbol{\delta}} \hat{\sigma}_{12})' \left\{ \widehat{\mathbf{C}} \mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}^2 \widehat{\mathbf{C}} + \mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}(\widehat{\mathbf{C}} * \widehat{\mathbf{C}}) \mathbf{D}_{\hat{\boldsymbol{\varepsilon}}} \right\} \\ &\quad \times (\mathbf{X} - \hat{\sigma}^{-2} \hat{\boldsymbol{\delta}} \hat{\sigma}_{12}) (\mathbf{X}' \widehat{\mathbf{C}} \mathbf{X})^{-1}, \end{aligned}$$

which is (22).

A.3. Derivation of [Theorem 3](#)

Instead of (41) we now have, using [Assumptions 3](#) and [7](#),

$$\begin{aligned} & - \frac{k^{-1/2}}{2} \left\{ \frac{\boldsymbol{\delta}' \mathbf{S} \boldsymbol{\delta}}{\text{tr}(\mathbf{B})} \right\} \frac{\partial Q_{\text{SJIVE}}^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\ &= k^{-1/2} \left\{ (\mathbf{0}, \mathbf{I}_g) - \sigma^{-2} \sigma_{21} \boldsymbol{\delta}' \right\} (\mathbf{M} - \mathbf{S}) \boldsymbol{\delta} \\ &\quad - \left\{ \frac{(\mathbf{0}, \mathbf{I}_g) \mathbf{S} \boldsymbol{\delta}}{\boldsymbol{\delta}' \mathbf{S} \boldsymbol{\delta}} - \frac{\sigma_{21}}{\sigma^2} \right\} k^{-1/2} \boldsymbol{\delta}' (\mathbf{M} - \mathbf{S}) \boldsymbol{\delta} \\ &= k^{-1/2} \left\{ (\mathbf{0}, \mathbf{I}_g) - \sigma^{-2} \sigma_{21} \boldsymbol{\delta}' \right\} (\mathbf{M} - \mathbf{S}) \boldsymbol{\delta} + o_p(1). \end{aligned} \quad (43)$$

Using (37) we thus find

$$\begin{aligned} \boldsymbol{\Psi}_w &= \lim_{n \rightarrow \infty} k^{-1} \left\{ \sum_{i=1}^n \sigma_i^2 \boldsymbol{\Pi}' \mathbf{Z}' \mathbf{C} \mathbf{e}_i \mathbf{e}_i' \mathbf{C} \mathbf{Z} \boldsymbol{\Pi} \right. \\ &\quad \left. + \sum_{i=1}^n \sum_{j=1}^n C_{ij}^2 \left(-\frac{\sigma_{21}}{\sigma^2}, \mathbf{I}_g \right) (\sigma_j^2 \boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_i \mathbf{e}_1 \mathbf{e}_1' \boldsymbol{\Sigma}_j) \left(-\frac{\sigma_{21}}{\sigma^2}, \mathbf{I}_g \right)' \right\}. \end{aligned}$$

As (42) remains valid under [Assumption 7](#) we find the result of [Theorem 3](#), where $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$ since $r_{\min}/k^{1/2} \rightarrow \infty$.

A.4. Primitive assumptions

In order to arrive at asymptotic distributions for their estimators [Hausman et al. \(2012\)](#), [Kunitomo \(2012\)](#) considered quite primitive assumptions. They applied martingale central limit theorems for quadratic forms as used by [Chao et al. \(2012\)](#) and [Anderson et al. \(2010\)](#), respectively. Here we consider the conditions of specification (27), which are very similar to the ones used by [Kunitomo \(2012\)](#). The only differences are that their matrix $\mathbf{P} - \mathbf{D}$ is replaced by our matrix $\mathbf{C}$, and their vector $\sum_i \sigma_{21i} / \sum_i \sigma_i^2$ is replaced by our σ_{21} / σ^2 .

To verify this, observe that $r_{\min}$ and $\mu_n^2 = \min_{1 \leq i \leq g} \mu_{in}^2$ have the same growth rate. Furthermore, if $\mathbf{z}_i' = \mathbf{e}_i' \mathbf{Z} \tilde{\boldsymbol{\Pi}}$, then the existence of the positive definite limit $\tilde{\mathbf{H}}$ implies the existence of c_u such that $\|\sum_i \mathbf{z}_i \mathbf{z}_i' / n\| \leq c_u$ and $\lambda_{\min}(\sum_i \mathbf{z}_i \mathbf{z}_i' / n) \geq 1/c_u$ for n sufficiently large, and $\sum_i \|\mathbf{z}_i\|^4 / n^2 \rightarrow 0$, as required.¹³

The asymptotic normality of [Assumption 5](#) can now be verified by following the steps made in [Hausman et al. \(2012\)](#) to derive the asymptotic normality of HLIM. Care should be taken when the martingale central limit theorem is applied. That is to say, Lemma A6 in [Hausman et al. \(2012\)](#), which is Lemma A2 of [Chao et al. \(2012\)](#), is formulated explicitly in terms of the off-diagonal elements of $\mathbf{P}$. For our purposes the result should remain valid when the off-diagonal elements of $\mathbf{C}$ are used instead of $\mathbf{P}$. As $C_{ij} = P_{ij}\{1/(1-h_i) + 1/(1-h_j)\}/2$, we find as a result of [Assumption 1](#), $C_{ij}^2 \leq P_{ij}^2 c_u^2$. Using this bound we can check the conditions of the martingale central limit theorem when P_{ij} is replaced by C_{ij} in the same way as is done in the proof of Lemma A2 of [Chao et al. \(2012\)](#). Thus [Assumptions 1–4](#) and $k/r_{\min} \rightarrow \gamma$, together with the conditions of specification (27) are sufficient for [Assumption 5](#).

A.5. Derivation of [Theorem 4](#)

To prove [Theorem 4](#) let

$$\begin{aligned} \mathbf{H}_1 &= (\boldsymbol{\Pi}' \mathbf{Z}' \mathbf{C} \mathbf{Z} \boldsymbol{\Pi})^{-1} \boldsymbol{\Pi}' \mathbf{Z}' \mathbf{C}^2 \mathbf{Z} \boldsymbol{\Pi} (\boldsymbol{\Pi}' \mathbf{Z}' \mathbf{C} \mathbf{Z} \boldsymbol{\Pi})^{-1}, \\ \mathbf{H}_2 &= \left\{ \boldsymbol{\Pi}' \mathbf{Z}' (\mathbf{P} - \mathbf{D}) \mathbf{Z} \boldsymbol{\Pi} \right\}^{-1} \boldsymbol{\Pi}' \mathbf{Z}' (\mathbf{P} - \mathbf{D})^2 \\ &\quad \times \mathbf{Z} \boldsymbol{\Pi} \left\{ \boldsymbol{\Pi}' \mathbf{Z}' (\mathbf{P} - \mathbf{D}) \mathbf{Z} \boldsymbol{\Pi} \right\}^{-1}, \end{aligned}$$

Using (34) we find

$$\begin{aligned} \mathbf{H}_1 &= \mathbf{H}^{-1} \boldsymbol{\Pi}' \mathbf{Z}' \left\{ \mathbf{I}_n + (1/4)(\mathbf{I}_n - \mathbf{D})^{-1}(\mathbf{I}_n - \mathbf{P})(\mathbf{I}_n - \mathbf{D})^{-1} \right\} \\ &\quad \times \mathbf{Z} \boldsymbol{\Pi} \mathbf{H}^{-1}, \\ \mathbf{H}_2 &= \left\{ \boldsymbol{\Pi}' \mathbf{Z}' (\mathbf{I}_n - \mathbf{D}) \mathbf{Z} \boldsymbol{\Pi} \right\}^{-1} \boldsymbol{\Pi}' \mathbf{Z}' (\mathbf{I}_n - \mathbf{D})^2 \mathbf{Z} \boldsymbol{\Pi} \\ &\quad \times \left\{ \boldsymbol{\Pi}' \mathbf{Z}' (\mathbf{I}_n - \mathbf{D}) \mathbf{Z} \boldsymbol{\Pi} \right\}^{-1}. \end{aligned}$$

¹³ The convergence $\sum_i \|\mathbf{z}_i\|^4 / n^2 \rightarrow 0$ is implied by the convergence $\sum_i \mathbf{z}_i' (\sum_i \mathbf{z}_i \mathbf{z}_i' / n)^{-1} \mathbf{z}_i / n^2 \rightarrow 0$, since $\sum_i \mathbf{z}_i \mathbf{z}_i' / n \rightarrow \tilde{\mathbf{H}} > 0$. Furthermore, $\sum_i \mathbf{z}_i' (\sum_i \mathbf{z}_i \mathbf{z}_i' / n)^{-1} \mathbf{z}_i / n^2 = \sum_i d_i^2$, where d_i is the i th diagonal element of the projection matrix $\mathbf{P}_{\tilde{\mathbf{H}}}$. As $\sum_i d_i^2 \leq g \max_{1 \leq i \leq n} d_i$ and $\max_{1 \leq i \leq n} d_i \rightarrow 0$ since $\sum_i \mathbf{z}_i \mathbf{z}_i' / n \rightarrow \tilde{\mathbf{H}}$, we find $\sum_i \|\mathbf{z}_i\|^4 / n^2 \rightarrow 0$.

Let $\mathbf{L} = \mathbf{Z}\Pi\mathbf{H}^{-1/2}$, so $\mathbf{L}'\mathbf{L} = \mathbf{I}_g$ and $\mathbf{L}\mathbf{L}' = \mathbf{P}_{Z\Pi}$ is a projection matrix and

$$\mathbf{H}_1 = \mathbf{H}^{-1/2}\mathbf{L}'\{\mathbf{I}_n + (1/4)(\mathbf{I}_n - \mathbf{D})^{-1}(\mathbf{I}_n - \mathbf{P})(\mathbf{I}_n - \mathbf{D})^{-1}\}\mathbf{L}\mathbf{H}^{-1/2},$$

$$\mathbf{H}_2 = \mathbf{H}^{-1/2}\{\mathbf{L}'(\mathbf{I}_n - \mathbf{D})\mathbf{L}\}^{-1}\mathbf{L}'(\mathbf{I}_n - \mathbf{D})^2\mathbf{L}\{\mathbf{L}'(\mathbf{I}_n - \mathbf{D})\mathbf{L}\}^{-1}\mathbf{H}^{-1/2}.$$

After premultiplying by $\mathbf{L}'(\mathbf{I}_n - \mathbf{D})\mathbf{L}\mathbf{H}^{1/2}$, and postmultiplying by its transpose, we find $\mathbf{H}_1 \leq \mathbf{H}_2$ if and only if

$$0 \leq \mathbf{L}'(\mathbf{I}_n - \mathbf{D})\{\mathbf{I}_n - \mathbf{P}_{Z\Pi} - (1/4)\mathbf{P}_{Z\Pi}(\mathbf{I}_n - \mathbf{D})^{-1} \\ \times (\mathbf{I}_n - \mathbf{P})(\mathbf{I}_n - \mathbf{D})^{-1}\mathbf{P}_{Z\Pi}\}(\mathbf{I}_n - \mathbf{D})\mathbf{L} \\ = \mathbf{L}'(\mathbf{I}_n - \mathbf{D})(\mathbf{I}_n - \mathbf{P}_{Z\Pi})\{\mathbf{I}_n - (1/4)(\mathbf{I}_n - \mathbf{D})^{-1}(\mathbf{I}_n - \mathbf{P}) \\ \times (\mathbf{I}_n - \mathbf{D})^{-1}\}(\mathbf{I}_n - \mathbf{P}_{Z\Pi})(\mathbf{I}_n - \mathbf{D})\mathbf{L},$$

where the equality follows from $(\mathbf{I}_n - \mathbf{P})\mathbf{L} = \mathbf{O}$. This inequality is implied by $\mathbf{D} < (1/2)\mathbf{I}_n$ since, if it holds,

$$\mathbf{I}_n - (1/4)(\mathbf{I}_n - \mathbf{D})^{-1}(\mathbf{I}_n - \mathbf{P})(\mathbf{I}_n - \mathbf{D})^{-1} \\ \geq \mathbf{I}_n - (1/4)(\mathbf{I}_n - \mathbf{D})^{-2} > \mathbf{O},$$

which establishes $\mathbf{H}_1 \leq \mathbf{H}_2$ as required. Furthermore, the rank of the difference is given by $\text{rank}\{(\mathbf{I}_n - \mathbf{P}_{Z\Pi})(\mathbf{I}_n - \mathbf{D})\mathbf{L}\} = \text{rank}\{(\mathbf{I}_n - \mathbf{P}_{Z\Pi})\mathbf{D}\Pi\}$.

References

- Ackerberg, D.A., Devereux, P.J., 2003. Improved JIVE estimators for overidentified linear models with and without heteroskedasticity, Working Paper.
- Anderson, T.W., 2005. Origins of the limited information maximum likelihood and two-stage least squares estimators. *J. Econometrics* 127, 1–16.
- Anderson, T.W., Kunitomo, N., Matsushita, Y., 2010. On the asymptotic optimality of the LIML estimator with possibly many instruments. *J. Econometrics* 157, 191–204.
- Andrews, D.W.K., Stock, J.H., 2007. Testing with many weak instruments. *J. Econometrics* 138, 24–46.
- Angrist, J.D., Imbens, G.W., Krueger, A., 1999. Jackknife instrumental variable estimators. *J. Appl. Econometrics* 14, 57–67.
- Bekker, P.A., 1994. Alternative approximations to the distributions of instrumental variable estimators. *Econometrica* 54, 657–682.
- Bekker, P.A., Van der Ploeg, J., 2005. Instrumental variable estimation based on grouped data. *Stat. Neerl.* 59, 239–267.
- Blomquist, S., Dahlberg, M., 1999. Small sample properties of LIML and Jackknife IV estimators: experiments with weak instruments. *J. Appl. Econometrics* 14, 96–88.
- Chamberlain, G., Imbens, G., 2004. Random effects estimators with many instrumental variables. *Econometrica* 72, 295–306.
- Chao, J.C., Hausman, J.A., Newey, W.K., Swanson, N.R., Woutersen, T., 2012. Combining two consistent estimators. In: Baltagi, B.H., Carter Hill, R., Newey, W.K., White, H.L. (Eds.), *Advances in Econometrics: Essays in Honor of Jerry Hausman*, Vol. 29. Emerald Group Publishing, UK.
- Chao, J.C., Hausman, J.A., Newey, W.K., Swanson, N.R., Woutersen, T., 2014. Testing overidentifying restrictions with many instruments and heteroskedasticity. *J. Econometrics* 178, 15–21.
- Chao, J.C., Swanson, N.R., 2005. Consistent estimation with a large number of weak instruments. *Econometrica* 73, 1673–1692.
- Chao, J.C., Swanson, N.R., Hausman, J.A., Newey, W.K., Woutersen, T., 2012. Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments. *Econometric Theory* 28, 42–86.
- Chioda, L., Jansson, M., 2009. Optimal invariant inference when the number of instruments is large. *Econometric Theory* 25, 793–805.
- Davidson, R., MacKinnon, J.G., 2006. The case against JIVE. *J. Appl. Econometrics* 21, 827–833.
- Donald, S.J., Newey, W.K., 2000. A Jackknife interpretation of the continuous updating estimator. *Econometric Theory* 16, 239–243.
- Fuller, W.A., 1977. Some properties of a modification of the limited information estimator. *Econometrica* 45, 939–954.
- Hahn, J., 2002. Optimal inference with many instruments. *Econometric Theory* 18, 140–168.
- Hahn, J., Hausman, J.A., Kuersteiner, G.M., 2004. Estimation with weak instruments: accuracy of higher-order bias and MSE approximations. *Econometric Theory* 20, 272–306.
- Hahn, J., Inoue, A., 2002. A Monte Carlo comparison of various asymptotic approximations to the distribution of instrumental variables estimators. *Econometric Theory* 18, 309–336.
- Han, C., Phillips, P.C.B., 2006. GMM with many moment conditions. *Econometrica* 74, 147–192.
- Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.
- Hansen, L.P., Heaton, J., Yaron, A., 1996. Finite-sample properties of some alternative GMM estimators. *J. Bus. Econ. Statist.* 14, 398–422.
- Hausman, J.A., Newey, W.K., Woutersen, T., Chao, J.C., Swanson, N.R., 2012. Instrumental variable estimation with heteroskedasticity and many instruments. *Quant. Econ.* 3, 211–255.
- Kunitomo, N., 1980. Asymptotic expansions of distributions of estimators in a linear functional relationship and simultaneous equations. *J. Amer. Statist. Assoc.* 75, 693–700.
- Kunitomo, N., 2012. An optimal modification of the liml estimation for many instruments and persistent heteroskedasticity. *Ann. Inst. Statist. Math.* 64, 881–910.
- Morimune, K., 1983. Approximate distributions of k-class estimators when the degree of overidentifiability is large compared with the sample size. *Econometrica* 51, 821–842.
- Newey, W.K., Windmeijer, F., 2009. Generalized method of moments with many weak moment condition. *Econometrica* 77, 687–719.
- Phillips, G.D.A., Hale, C., 1977. The bias of instrumental variable estimators of simultaneous equation systems. *Internat. Econ. Rev.* 18, 219–228.
- Stock, J., Yogo, M., 2005. Asymptotic distributions of instrumental variables statistics with many instruments. In: Andrews, D.W.K., Stock, J.H. (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge University Press, (Chapter 6).
- Van der Ploeg, J., Bekker, P.A., 1995. Efficiency bounds for instrumental variable estimators under group asymptotics, Working Paper, University of Groningen.
- Van Hasselt, M., 2010. Many instruments asymptotic approximations under nonnormal error distributions. *Econometric Theory* 26, 633–645.