Symmetric Jackknife Instrumental Variable Estimation

Bekker, Paul A. and Crudu, Federico

5 April 2012

Symmetric Jackknife Instrumental Variable Estimation

Paul A. Bekker*
Faculty of Economics and Business
University of Groningen

Federico Crudu†
University of Sassari and CRENoS

April, 2012

Abstract

This paper gives a new jackknife estimator for instrumental variable inference with unknown heteroskedasticity. The estimator is derived by using a method of moments approach similar to the one that produces LIML in case of homoskedasticity. The estimator is symmetric in the endogenous variables including the dependent variable. Many instruments and many weak instruments asymptotic distributions are derived using high-level assumptions that allow for the simultaneous presence of weak and strong instruments for different explanatory variables. Standard errors are formulated compactly. We review briefly known estimators and show in particular that the symmetric jackknife estimator performs well when compared to the HLIM and HFUL estimators of Hausman et al. (2011) in Monte Carlo experiments.

Key words: Instrumental Variables, Heteroskedasticity, Many Instruments, Jackknife
JEL classification: C12, C13, C23

*p.a.bekker@rug.nl, P.O. Box 800, 9700 AV Groningen, The Netherlands.
†federico.crudu@gmail.com, Via Torre Tonda 34, 07100, Sassari, Italy. Financial support from Marie Curie Excellence Grant MEXT-CT-2006-042471 is gratefully acknowledged.

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 and LIML is inconsistent under heteroskedasicity. A number of estimators have been considered, including the two step feasible GMM estimator of Hansen (1982), the continuously updated GMM estimator of Hansen Heaton and Yaron (1996), the grouping estimators of Bekker and Van der Ploeg (2005), the jackknife estimators of Angrist, Imbens and Krueger (1999) and the HLIM and HFUL estimators of Hausman et al. (2011). 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, Imbens and Krueger (1999), Donald and Newey (2000), Ackerberg and Deveraux (2003). Our genuine jackknife estimator shares with LIML the property that the endogenous variables are treated symmetrically and that the estimation is not affected by the type of normalization beyond the normalization itself.

Hausman et al. (2011) use a LIML version of the JIVE2 estimator of Angrist, Imbens and Krueger (1999). The JIVE2 estimator is not a genuine jackknife estimator, but in the LIML version it treats endogenous variables symmetrically. 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. Thus it seems the efficiency problems of jackknife estimators noted in Davidson and McKinnon (2006) are overcome. Here we show there is room for improvement. The symmetric jacknife estimator is a genuine jackknife estimator and it has a signal component that is larger than found for HLIM. The Monte Carlo experiments, with the same set up as used in Hausman et al. (2011), show it performs better than HLIM and its Fuller modifications.

The asymptotic theory allows for both many instruments and many weak instruments asymptotics. Influential papers in this area include Donald and Newey (2001), Hahn, Hausman and Kuersteiner (2004), Hahn (2002), Hahn and Inoue (2002), Chamberlain and Imbens (2004), Chao and Swanson (2005), Stock and Yogo (2005), Han and Phillips (2006) and Andrews and Stock (2007). 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 presents the Monte Carlo findings.

2 The Model and some estimators

Consider observations in the n vector $\boldsymbol{y}$ and the $n \times k$ 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. (2011) 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. (2010). The assumption $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{Z\Pi}$ is made for convenience and could be generalized as in Hausman et al. (2011), or as in Bekker (1994). The disturbances in the $n \times (1+g)$ matrix $(\boldsymbol{\varepsilon}, \boldsymbol{V})$ have rows $(\boldsymbol{\varepsilon}_i, \boldsymbol{V}_i)$, 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 rows of $(\boldsymbol{y}_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$, 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 formable unit vectors.

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

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

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

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \{\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X} - \lambda_f \boldsymbol{X}'(\boldsymbol{I}_n - \boldsymbol{P})\boldsymbol{X}\}^{-1} \{\boldsymbol{X}'\boldsymbol{P}\boldsymbol{y} - \lambda_f \boldsymbol{X}'(\boldsymbol{I}_n - \boldsymbol{P})\boldsymbol{y}\}, \\ \lambda_f &= \lambda - \alpha/(n-k), \\ \lambda &= 1/\lambda_{\max}[\{(\boldsymbol{y}, \boldsymbol{X})'\boldsymbol{P}(\boldsymbol{y}, \boldsymbol{X})\}^{-1} \, (\boldsymbol{y}, \boldsymbol{X})'(\boldsymbol{I}_n - \boldsymbol{P})(\boldsymbol{y}, \boldsymbol{X})], \end{aligned}$$

where $\lambda_{\max}$ indicates the largest eigenvalue. For $\alpha = 0$ LIML is found, which has no moments under normality. For $\alpha = 1$ the Fuller estimator is found. Under normality and homoskedasticity, where the matrices $\boldsymbol{\Sigma}_i$ do not vary over $i = 1, \ldots, n$, it 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 inconsistent under many-instruments asymptotics with heteroskedasticity.

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

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

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, Heaton and Yaron (1996), which is found by minimizing the objective function

$$Q_{\text{CUE}}(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{Z}\left\{\sum_{i=1}^{n} \hat{\sigma}_i^2(\boldsymbol{\beta}) \boldsymbol{Z}_i'\boldsymbol{Z}_i\right\}^{-1}\boldsymbol{Z}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{6}$$

where $\hat{\sigma}_i^2(\boldsymbol{\beta}) = (y_i - \boldsymbol{X}_i\boldsymbol{\beta})^2$. Newey and Windmeijer (2009) showed this estimator and other generalized empirical likehood 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. (2011) in Monte Carlo experiments.

To reduce problems related to the consistent estimation of the weighting matrix Bekker and Van der Ploeg (2005) use exogenous clustering of observations. Let $\mathcal{C}_s$ define the sth cluster of size n_s such that the ith observation is in the sth cluster if $i \in \mathcal{C}_s$, $s = 1, \ldots, m$. 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. Let the group means and group sample covariance matrices of the data be denoted by $(\bar{y}_s, \bar{\boldsymbol{X}}_s)$ and $\boldsymbol{S}_s$, respectively, then the continuously updated Group-GMM estimator is found by minimizing

$$Q_{\text{G-CUE}} = \sum_{s=1}^{m} \frac{n_s(\bar{y}_s - \bar{\boldsymbol{X}}_s\boldsymbol{\beta})^2}{(1, -\boldsymbol{\beta}')\boldsymbol{S}_s(1, -\boldsymbol{\beta}')'}. \tag{7}$$

Bekker and Van der Ploeg (2005) give standard errors that are consistent for sequences where the number of groups m 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 goup heteroskedasticity to gain efficiency, yet it loses efficiency as the within goup sample covariance matrices of the instruments are different from zero in general.

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

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\text{JIVE1}} &= (\tilde{\boldsymbol{X}}'\boldsymbol{X})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{y}, \\ \boldsymbol{e}_i'\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{X}}_i &= \frac{\boldsymbol{Z}_i(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X} - h_i\boldsymbol{X}_i}{1 - h_i}, \end{aligned} \tag{8}$$

where $h_i = P_{ii}$, and $i = 1, \ldots, n$. It is robust against heteroskedasticity and many-instruments consistent. The JIVE2 estimator of Angrist, Imbens and Krueger (1999) is not a genuine jackknife estimator but it 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

$$Q_{\text{JIVE2}}(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\{\boldsymbol{P} - \boldsymbol{D}\}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{9}$$

where $\boldsymbol{D} = \text{Diag}(\boldsymbol{h})$ 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 Deveraux (2003). However, Davidson and McKinnon (2006) have shown that the jackknife estimators can have low efficiency relative to LIML under homoskedasticity.

Therefore, Hausman et al. (2011) consider jackknife versions of LIML and the Fuller (1977) estimator by using the objective function

$$Q_{\text{HLIM}}(\boldsymbol{\beta}) = \frac{(\boldsymbol{y}-\boldsymbol{X\beta})'\{\boldsymbol{P}-\boldsymbol{D}\}(\boldsymbol{y}-\boldsymbol{X\beta})}{(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})}. \tag{10}$$

The estimators are given by

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \{\boldsymbol{X}'(\boldsymbol{P}-\boldsymbol{D})\boldsymbol{X} - \hat{\alpha}\boldsymbol{X}'\boldsymbol{X}\}^{-1}\{\boldsymbol{X}'(\boldsymbol{P}-\boldsymbol{D})\boldsymbol{y} - \hat{\alpha}\boldsymbol{X}'\boldsymbol{y}\}, \qquad (11)\\ \hat{\alpha} &= \frac{(n+c)\widetilde{\alpha}-c}{n+c\widetilde{\alpha}-c},\\ \widetilde{\alpha} &= \lambda_{\min}[\{(\boldsymbol{y},\boldsymbol{X})'(\boldsymbol{y},\boldsymbol{X})\}^{-1}\,(\boldsymbol{y},\boldsymbol{X})'\{\boldsymbol{P}-\boldsymbol{D}\}(\boldsymbol{y},\boldsymbol{X})].\end{aligned}$$

For $c=0$, $\hat{\boldsymbol{\beta}}_{\text{HLIM}}$ is found, and $c=1$ produces $\hat{\boldsymbol{\beta}}_{\text{HFUL}}$. Hausman et al. (2011) 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, but there is room for improvement as will be argued below.

3 A method of moments and jackknife estimator

In order to handle heteroskedasticity the grouping estimator uses 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 $\boldsymbol{\beta}$ contained in the matrix $(\boldsymbol{y},\boldsymbol{X})'\boldsymbol{D}(\boldsymbol{y},\boldsymbol{X})$. As an alternative to the jackknife objective function Q_{HLIM}, we consider a method-of-moments approach that maintains the signal component in the expectation of $(\boldsymbol{y},\boldsymbol{X})'\boldsymbol{P}(\boldsymbol{y},\boldsymbol{X})$ 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. In a second stage we find our method-of-moments estimator can be interpreted as a jackknife estimator.

3.1 A method of moments estimator

To find a method-of moments estimator we need a many-instruments consistent estimator of the noise component $\boldsymbol{\Omega}_{(n)}$ in

$$\mathrm{E}\left\{(\boldsymbol{y},\boldsymbol{X})'\boldsymbol{P}(\boldsymbol{y},\boldsymbol{X})\right\} = \boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi} + \boldsymbol{\Omega}_{(n)}, \tag{12}$$

$$\boldsymbol{\Omega}_{(n)} = \sum_{i=1}^{n} h_i \boldsymbol{\Omega}_i. \tag{13}$$

If $\hat{\boldsymbol{\Omega}}_{(n)} = (\boldsymbol{y},\boldsymbol{X})'\boldsymbol{B}(\boldsymbol{y},\boldsymbol{X})$ were an unbiased estimator,

$$\mathrm{E}(\widehat{\boldsymbol{\Omega}}_{(n)}) = \mathrm{E}\left\{(\boldsymbol{y},\boldsymbol{X})'\boldsymbol{B}(\boldsymbol{y},\boldsymbol{X})\right\} = \sum_{i=1}^{n} \boldsymbol{B}_{ii}\boldsymbol{\Omega}_i = \boldsymbol{\Omega}_{(n)}, \tag{14}$$

then a LIML-like method-of-moments estimator would be given by minimizing

$$Q(\boldsymbol{\beta}) = \frac{(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{P}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})}{(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{B}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})}. \tag{15}$$

As a method-of-moments estimator, the many-instruments consistency would follow easily. Furthermore, as $\boldsymbol{P}$ and $\boldsymbol{B}$ must have the same diagonal elements, third and fourth order moments of the disturbances would not affect the many-instruments asymptotic distribution. Since the signal component $\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi}$ is maintained, the estimator would really have LIML-like features, but now robust against heteroskedasticity.

The problem is to formulate an unbiased estimator $\widehat{\boldsymbol{\Omega}}_{(n)}$. As a starting point we consider an estimator for $\boldsymbol{\Omega}_i$ given by

$$\widehat{\boldsymbol{\Omega}}_i = (\boldsymbol{y},\boldsymbol{X})'\frac{(\boldsymbol{I}_n-\boldsymbol{P})\boldsymbol{e}_i\boldsymbol{e}'(\boldsymbol{I}_n-\boldsymbol{P})}{\boldsymbol{e}_i'(\boldsymbol{I}_n-\boldsymbol{P})\boldsymbol{e}_i}(\boldsymbol{y},\boldsymbol{X}),$$

$$\mathrm{E}(\widehat{\boldsymbol{\Omega}}_i) = \sum_{j=1}^{n}\frac{\boldsymbol{e}_j'(\boldsymbol{I}_n-\boldsymbol{P})\boldsymbol{e}_i\boldsymbol{e}_i'(\boldsymbol{I}_n-\boldsymbol{P})\boldsymbol{e}_j}{\boldsymbol{e}_i'(\boldsymbol{I}_n-\boldsymbol{P})\boldsymbol{e}_i}\boldsymbol{\Omega}_j = (1-h_i)\boldsymbol{\Omega}_i + (1-h_i)^{-1}\sum_{j\neq i}^{n}\boldsymbol{P}_{ij}^2\boldsymbol{\Omega}_j.$$

Obviously, $\widehat{\boldsymbol{\Omega}}_i$ is biased, which also holds for an estimator of $\boldsymbol{\Omega}_{(n)}$ given by

$$\sum_{i=1}^{n} h_i \widehat{\boldsymbol{\Omega}}_i = (\boldsymbol{y}, \boldsymbol{X})' \boldsymbol{B}_{\text{HR}} (\boldsymbol{y}, \boldsymbol{X}), \tag{16}$$

$$\boldsymbol{B}_{\text{HR}} = (\boldsymbol{I}_n - \boldsymbol{P}) \boldsymbol{D} (\boldsymbol{I}_n - \boldsymbol{D})^{-1} (\boldsymbol{I}_n - \boldsymbol{P}). \tag{17}$$

We find the trace of $\boldsymbol{B}_{\text{HR}}$ equals the trace of $\boldsymbol{P}$, but the diagonal elements, given by

$$\begin{aligned}
\boldsymbol{B}_{\text{HR}tt} &= \boldsymbol{e}_t'(\boldsymbol{I}_n - \boldsymbol{P}) \boldsymbol{D}_P \{\boldsymbol{I}_n - \boldsymbol{D}_P\}^{-1} (\boldsymbol{I}_n - \boldsymbol{P}) \boldsymbol{e}_t \\
&= \sum_{i=1}^{n} h_i \frac{\boldsymbol{e}_t'(\boldsymbol{I}_n - \boldsymbol{P}) \boldsymbol{e}_i \boldsymbol{e}_i' (\boldsymbol{I}_n - \boldsymbol{P}) \boldsymbol{e}_t}{\boldsymbol{e}_i'(\boldsymbol{I}_n - \boldsymbol{P}) \boldsymbol{e}_i} \\
&= h_t - \sum_{i=1}^{n} h_i \frac{\boldsymbol{e}_t'(\boldsymbol{I}_n - \boldsymbol{P}) \boldsymbol{e}_i \boldsymbol{e}_i' \boldsymbol{P} \boldsymbol{e}_t}{\boldsymbol{e}_i'(I_n - \boldsymbol{P}) \boldsymbol{e}_i} \\
&= h_t - \frac{h_t^2}{1 - h_t} + \sum_{i=1}^{n} \frac{h_i P_{it}^2}{1 - h_i},
\end{aligned} \tag{18}$$

are different from the diagonal elements h_t of $\boldsymbol{P}$. For the second and third terms on the right hand side we find

$$0 \leq \frac{h_t^2}{1 - h_t} \leq \frac{h_t \max_t(h_t)}{1 - \max_t(h_t)}, \tag{19}$$

$$0 \leq \sum_{i=1}^{n} \frac{h_i P_{it}^2}{1 - h_i} \leq \left(\frac{\max_t(h_t)}{1 - \max_t(h_t)} \right) \sum_{i=1}^{n} P_{it}^2 = \frac{h_t \max_t(h_t)}{1 - \max_t(h_t)}. \tag{20}$$

If $\max_t(h_t) \to 0$, then $B_{\text{HR}tt} \to h_t$, but that will not happen with many instruments, if $k/n \to 0$ does not hold. Yet k/n may be small in practice, and under normality the bias of $\sum_{i=1}^{n} h_i \widehat{\boldsymbol{\Omega}}_i$ may have a negligible effect on the distribution of the estimator of $\boldsymbol{\beta}$. However, due to the difference between the diagonals of $\boldsymbol{P}$ and $\boldsymbol{B}_{\text{HR}}$, third and fourth order moments would enter the asymptotic distribution in case of nonnormality. Therefore, instead of minimizing the objective function $Q_{MM}(\boldsymbol{\beta})$ in (15) for $\boldsymbol{B}$ as given by $\boldsymbol{B}_{\text{HR}}$ in (17), we prefer to remove the bias.

The matrix $\boldsymbol{B}_{\text{HR}}$ can be written as $\boldsymbol{B}_1 + \boldsymbol{B}_2$, where

$$\boldsymbol{B}_1 = \boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1} - \frac{1}{2}\left\{\boldsymbol{P}\boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1} + \boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1}\boldsymbol{P}\right\}, \tag{21}$$

$$\boldsymbol{B}_2 = \boldsymbol{P}\boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1}\boldsymbol{P} - \frac{1}{2}\left\{\boldsymbol{P}\boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1} + \boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1}\boldsymbol{P}\right\}. \tag{22}$$

We find $B_{1ii} = h_i$, $|B_{2ii}| \leq h_i \max_i(h_i)/\{1 - \max_i(h_i)\}$, $i = 1, \ldots, n$, and $\boldsymbol{Z}'\boldsymbol{B}_1\boldsymbol{Z} = \boldsymbol{Z}'\boldsymbol{B}_2\boldsymbol{Z} = \mathbf{O}$. Since an unbiased estimator for $\boldsymbol{\Omega}_{(n)}$ is given by $\widehat{\boldsymbol{\Omega}}_{(n)} = (\boldsymbol{y}, \boldsymbol{X})'\boldsymbol{B}_1(\boldsymbol{y}, \boldsymbol{X})$, we could use $\boldsymbol{B}_1$ instead of $\boldsymbol{B}_{\text{HR}}$ and thus find a method-moments estimator.

For practical reasons we choose a slightly different estimator with the same many-instruments asymptotic distribution. The point is that in the presence of exogenous explanatory variables, or a constant term, the covariance matrix $\boldsymbol{\Omega}_{(n)}$ is block-diagonal with one block different from zero. However, the estimator $\widehat{\boldsymbol{\Omega}}_{(n)}$ is not block diagonal. If the off-diagonal blocks are replaced by zeros, we would still have an unbiased estimator, but one that no longer can be written as $(\boldsymbol{y}, \boldsymbol{X})\boldsymbol{B}^*(\boldsymbol{y}, \boldsymbol{X})$ for a suitable matrix $\boldsymbol{B}^*$. A block-diagonal structure easily allows for minimization over coefficients of endogenous variables only. Therefore we choose to maintain $\boldsymbol{B}$. We add $\boldsymbol{B}_2$ to the matrix $\boldsymbol{P}$ in the numerator of the objective function instead. Summarizing we have

$$\boldsymbol{A}_{\text{HR}} = \boldsymbol{P} - \frac{1}{2}\left\{\boldsymbol{P}\boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1}(\boldsymbol{I}_n - \boldsymbol{P}) + (\boldsymbol{I}_n - \boldsymbol{P})\boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1}\boldsymbol{P}\right\},$$

$$\boldsymbol{B}_{\text{HR}} = (\boldsymbol{I}_n - \boldsymbol{P})\boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1}(\boldsymbol{I}_n - \boldsymbol{P}),$$

with objective function

$$Q_{\text{HRMM}}(\boldsymbol{\beta}) = \frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{A}_{\text{HR}}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{B}_{\text{HR}}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}. \tag{23}$$

The diagonals of $\boldsymbol{A}_{\text{HR}}$ and $\boldsymbol{B}_{\text{HR}}$ are equal. In comparison with the jackknife objective function Q_{HLIM} in (10) the signal is larger, since $\boldsymbol{\Pi}'\boldsymbol{Z}\boldsymbol{A}_{\text{HR}}\boldsymbol{Z}\boldsymbol{\Pi} = \boldsymbol{\Pi}\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi}$.

3.2 Interpretation as a symmetric jackknife estimator

Interestingly, and to our surprise, our objective function Q_{HRMM} can be related to jackknife estimation. That is to say, the minimizer of Q_{HRMM} also minimizes $Q_{\text{HRMM}} - 1$, which replaces the matrix $\boldsymbol{A}_{\text{HR}}$ by

$$\begin{aligned}
\boldsymbol{A}_{\text{HR}} - \boldsymbol{B}_{\text{HR}} &= \boldsymbol{P} - \boldsymbol{B}_1 \\
&= \boldsymbol{P} - \boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1} + \left\{\boldsymbol{P}\boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1} + (\boldsymbol{I}_n - \boldsymbol{D})^{-1}\boldsymbol{D}\boldsymbol{P}\right\}/2 \\
&= (\tilde{\boldsymbol{P}} + \tilde{\boldsymbol{P}}')/2, \qquad (24)\\
\tilde{\boldsymbol{P}} &= \boldsymbol{P} - \boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1} + \boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1}\boldsymbol{P} \\
&= (\boldsymbol{I}_n - \boldsymbol{D})^{-1}(\boldsymbol{P} - \boldsymbol{D}). \qquad (25)
\end{aligned}$$

The jackknife estimator $\boldsymbol{\beta}_{\text{JIVE1}} = (\tilde{\boldsymbol{X}}'\boldsymbol{X})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{y}$ is based on $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{P}}\boldsymbol{X}$. So, if we define $\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{P}}\boldsymbol{y}$, then we find the numerator of the objective function is given by

$$\begin{aligned}
(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{A}_{\text{HR}} - \boldsymbol{B}_{\text{HR}})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) &= \frac{1}{2}(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta}) \\
&= (\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}).
\end{aligned}$$

That is to say, genuine 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. Thus HRMM can be interpreted as a symmetric jackknife (SJIVE) procedure.

Equivalent to minimizing Q_{HRMM} we use

$$Q_{\text{SJIVE}} = \frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{C}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{B}_{\text{HR}}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}, \qquad (26)$$

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

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\boldsymbol{X}'\boldsymbol{C}\boldsymbol{X} - \hat{\lambda}\boldsymbol{X}\boldsymbol{B}_{\text{HR}}\boldsymbol{X}\right)^{-1} \left(\boldsymbol{X}'\boldsymbol{C}\boldsymbol{y} - \hat{\lambda}\boldsymbol{X}\boldsymbol{B}_{\text{HR}}\boldsymbol{y}\right), \qquad (27)\\ &\qquad \hat{\lambda} = \lambda - \alpha/\operatorname{tr}(\boldsymbol{B}_{\text{HR}}),\\ &\qquad \lambda = \lambda_{\min}\left[\{(\boldsymbol{y},\boldsymbol{X}_1)'\boldsymbol{B}_{\text{HR}}(\boldsymbol{y},\boldsymbol{X}_1)\}^{-1}\ (\boldsymbol{y},\boldsymbol{X}_1)'\boldsymbol{C}^*(\boldsymbol{y},\boldsymbol{X}_1)\right].\end{aligned}$$

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 below we compute standard errors as the square root of the diagonal elements of the estimated covariance matrix, which is formulated concisely as

$$\widehat{\operatorname{Var}}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\widehat{\boldsymbol{C}}\boldsymbol{X})^{-1}(\boldsymbol{X} - \hat{\sigma}^{-2}\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\sigma}}_{12})'\left(\widehat{\boldsymbol{C}}\boldsymbol{D}_{\hat{\varepsilon}}^2\widehat{\boldsymbol{C}} + \boldsymbol{D}_{\hat{\varepsilon}}\widehat{\boldsymbol{C}}^{(2)}\boldsymbol{D}_{\hat{\varepsilon}}\right)(\boldsymbol{X} - \hat{\sigma}^{-2}\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\sigma}}_{12})(\boldsymbol{X}'\widehat{\boldsymbol{C}}\boldsymbol{X})^{-1}, \qquad (28)$$

where $\widehat{\boldsymbol{C}} = \boldsymbol{C} - \hat{\lambda}\boldsymbol{B}_{\text{HR}}$ and $\widehat{\boldsymbol{C}}^{(2)}$ is the element wise or Hadamard product $\widehat{\boldsymbol{C}} * \widehat{\boldsymbol{C}}$. The diagonal matrix $\boldsymbol{D}_{\hat{\varepsilon}}$ has the residuals $\hat{\boldsymbol{\varepsilon}} = y - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ on the diagonal. Finally, $\hat{\sigma}^2$ and $\hat{\boldsymbol{\sigma}}_{21}$ are found based on $\widehat{\boldsymbol{\Omega}} = (\boldsymbol{y},\boldsymbol{X})'\boldsymbol{B}_{\text{HR}}(\boldsymbol{y},\boldsymbol{X})/\operatorname{tr}(\boldsymbol{B}_{\text{HR}})$, which is transformed to $\widehat{\boldsymbol{\Sigma}}$ similar to (30) below.

4 Asymptotic distributions

We consider many instruments and many weak instruments parameter sequences to describe the asymptotic distributions of the heteroskedasticity robust estimator SJIVE as given in (27). 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. (2011) could be used, although our results hold more generally.

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

Assumption 2. *The covariance matrices of the disturbances are bounded,* $0 \leq \boldsymbol{\Omega}_i \leq c_u \boldsymbol{I}_{g+1}$ *and satisfy* $\operatorname{tr}(\boldsymbol{B}_{HR})^{-1} \sum_{i=1}^n \boldsymbol{e}_i' \boldsymbol{B}_{HR} \boldsymbol{e}_i \boldsymbol{\Omega}_i \to \boldsymbol{\Omega}$.

Assumption 3. *Let* $\boldsymbol{S} = (\boldsymbol{y}, \boldsymbol{X})' \boldsymbol{B}_{HR} (\boldsymbol{y}, \boldsymbol{X})$ *and* $\boldsymbol{C} = \boldsymbol{A}_{HR} - \boldsymbol{B}_{HR}$*, then*

$$
\begin{aligned}
&\operatorname{plim}_{n\to\infty} \operatorname{tr}(\boldsymbol{B}_{HR})^{-1} \boldsymbol{S} = \lim_{n\to\infty} \operatorname{tr}(\boldsymbol{B}_{HR})^{-1} \operatorname{E} \boldsymbol{S} = \boldsymbol{\Omega},\\
&\operatorname{plim}_{n\to\infty} (\boldsymbol{\Pi} \boldsymbol{Z}' \boldsymbol{Z} \boldsymbol{\Pi})^{-1} \boldsymbol{X}' \boldsymbol{C} \boldsymbol{X} = \lim_{n\to\infty} (\boldsymbol{\Pi} \boldsymbol{Z}' \boldsymbol{Z} \boldsymbol{\Pi})^{-1} \operatorname{E}(\boldsymbol{X}' \boldsymbol{C} \boldsymbol{X}) = \boldsymbol{I}_g.
\end{aligned}
$$

Let $r_{\min} = \lambda_{\min}(\boldsymbol{\Pi}' \boldsymbol{Z}' \boldsymbol{Z} \boldsymbol{\Pi})$ be the smallest eigenvalue of the signal matrix.

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

Let $Q^*_{\text{SJIVE}}(\boldsymbol{\beta}) = \operatorname{tr}(\boldsymbol{B}_{\text{HR}}) Q_{\text{SJIVE}}(\boldsymbol{\beta})$, then the many-instruments asymptotic approximations are based on the following assumption.

Assumption 5. Many instruments*:* $k/r_{\min} \to \gamma$*, and*

$$
\left\{ \frac{\partial^2 Q^*_{SJIVE}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right\}^{1/2} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left\{ \frac{\partial^2 Q^*_{SJIVE}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right\}^{-1/2} \frac{\partial Q^*_{SJIVE}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + o_p(1) \overset{a}{\sim} \mathcal{N}\left(\mathbf{0},\ \boldsymbol{\Phi}\right),
$$

where $\boldsymbol{\Phi} = \boldsymbol{H}^{-1/2} \boldsymbol{J} \boldsymbol{H}^{-1/2}$ *and*

$$
\begin{aligned}
\boldsymbol{H} &= \operatorname{plim}_{n\to\infty} (\boldsymbol{\Pi}' \boldsymbol{Z}' \boldsymbol{Z} \boldsymbol{\Pi})^{-1/2} \frac{\partial^2 Q^*_{SJIVE}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} (\boldsymbol{\Pi}' \boldsymbol{Z}' \boldsymbol{Z} \boldsymbol{\Pi})^{-1/2},\\
&(\boldsymbol{\Pi}' \boldsymbol{Z}' \boldsymbol{Z} \boldsymbol{\Pi})^{-1/2} \frac{\partial Q^*_{SJIVE}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \overset{a}{\sim} \mathcal{N}(\mathbf{0},\ \boldsymbol{J}),\\
\boldsymbol{J} &= \lim_{n\to\infty} \operatorname{Var}\left\{ (\boldsymbol{\Pi}' \boldsymbol{Z}' \boldsymbol{Z} \boldsymbol{\Pi})^{-1/2} \frac{\partial Q^*_{SJIVE}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + o_p(1) \right\}.
\end{aligned}
\tag{29}
$$

The $o_p(1)$ term in (29) is defined explicitly in (39) in the Appendix.

To formulate the main theorem we use

$$
\boldsymbol{\Sigma} = \begin{pmatrix} 1 & -\boldsymbol{\beta}' \\ \mathbf{0} & \boldsymbol{I}_g \end{pmatrix} \boldsymbol{\Omega} \begin{pmatrix} 1 & \mathbf{0} \\ -\boldsymbol{\beta} & \boldsymbol{I}_g \end{pmatrix} = \begin{pmatrix} \sigma^2 & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \tag{30}
$$

where $\boldsymbol{\Omega}$ is defined in Assumption 2.

Theorem 1. Many instruments *If Assumptions 1-5 are satisfied, then* $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{SJIVE}$ *is consistent and* $(\boldsymbol{X}\widehat{\boldsymbol{C}}\boldsymbol{X})^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \overset{a}{\sim} \mathcal{N}(\boldsymbol{0},\ \boldsymbol{\Psi})$*, where* $\widehat{\boldsymbol{C}} = \boldsymbol{C} - \hat{\lambda}\boldsymbol{B}_{HR}$ *and*

$$\boldsymbol{\Psi} = \lim_{n\to\infty}\left[(\boldsymbol{\Pi Z'Z\Pi})^{-1/2}\left\{\sum_{i=1}^{n}\sigma_i^2\boldsymbol{\Pi Z'Pe}_i\boldsymbol{e}_i'\boldsymbol{PZ\Pi}\right.\right. \tag{31}$$
$$\left.\left. + \sum_{i=1}^{n}\sum_{j=1}^{n}C_{ij}^2\left(-\frac{\boldsymbol{\sigma}_{21}}{\sigma^2}, \boldsymbol{I}_g\right)\left(\sigma_j^2\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_i\boldsymbol{e}_1\boldsymbol{e}_1'\boldsymbol{\Sigma}_j\right)\left(-\frac{\boldsymbol{\sigma}_{21}}{\sigma^2}, \boldsymbol{I}_g\right)'\right\}(\boldsymbol{\Pi Z'Z\Pi})^{-1/2}\right].$$

A consistent estimator for $\boldsymbol{\Psi}$ *is given by*

$$\widehat{\boldsymbol{\Psi}} = (\boldsymbol{X}'\widehat{\boldsymbol{C}}\boldsymbol{X})^{-1/2}(\boldsymbol{X} - \hat{\sigma}^{-2}\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\sigma}}_{12})'\left(\widehat{\boldsymbol{C}}\boldsymbol{D}_{\hat{\varepsilon}}^2\widehat{\boldsymbol{C}} + \boldsymbol{D}_{\hat{\varepsilon}}\widehat{\boldsymbol{C}}^{(2)}\boldsymbol{D}_{\hat{\varepsilon}}\right)(\boldsymbol{X} - \hat{\sigma}^{-2}\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\sigma}}_{12})(\boldsymbol{X}'\widehat{\boldsymbol{C}}\boldsymbol{X})^{-1/2},$$

where $\boldsymbol{D}_{\hat{\varepsilon}}$ *is diagonal with* $\hat{\boldsymbol{\varepsilon}} = y - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ *on the diagonal, and* $\widehat{\boldsymbol{C}}^{(2)}$ *is the element wise or Hadamard product* $\widehat{\boldsymbol{C}} * \widehat{\boldsymbol{C}}$*. Finally,* $\hat{\sigma}^2$ *and* $\hat{\boldsymbol{\sigma}}_{21}$ *are found based on* $\widehat{\boldsymbol{\Omega}} = \boldsymbol{S}/\operatorname{tr}(\boldsymbol{B}_{HR})$*, which is transformed to* $\widehat{\boldsymbol{\Sigma}}$ *similar to* (30)*.*

The proof is given in the Appendix. The asymptotic covariance matrix $\boldsymbol{\Psi}$ in (31) has two terms. Under large-sample asymptotics, when $k/r_{\min} \to 0$ 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). 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 and showed the first term actually vanishes, while estimators such as LIML under homoskedasticity are still consistent. Hausman et al. (2011) derived the many-weak instruments asymptotic distribution of HLIM and HFUL as given in (11). We have a similar result.

Let $r_{\max} = \lambda_{\max}(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z\Pi})$ be the largest eigenvalue of the signal matrix.

Assumption 6. Many weak instruments: $k/r_{\max} \to \infty$, $k^{1/2}/r_{\min} \to 0$ *and*

$$k^{-1/2}\frac{\partial^2 Q^*_{SJIVE}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = k^{-1/2}\frac{\partial Q^*_{SJIVE}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} + o_p(1) \overset{a}{\sim} \mathcal{N}\left(\mathbf{0},\ \boldsymbol{\Phi}_w\right),$$

where

$$\boldsymbol{\Phi}_w = \lim_{n\to\infty} \operatorname{Var}\left\{k^{-1/2}\frac{\partial Q^*_{SJIVE}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} + o_p(1)\right\}. \tag{32}$$

The $o_p(1)$ term in (32) is defined explicitly in (41) in the Appendix.

Theorem 2. Many weak instruments *If Assumptions 1-4 and 6 are satisfied, then* $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{SJIVE}$ *is consistent and* $k^{-1/2}\boldsymbol{X}\widehat{\boldsymbol{C}}\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \overset{a}{\sim} \mathcal{N}(\mathbf{0},\ \boldsymbol{\Psi}_w)$, *where*

$$\boldsymbol{\Psi}_w = \lim_{n\to\infty} k^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}^2\left(-\frac{\sigma_{21}}{\sigma^2}, \boldsymbol{I}_g\right)\left(\sigma_j^2\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_i\boldsymbol{e}_1\boldsymbol{e}_1'\boldsymbol{\Sigma}_j\right)\left(-\frac{\sigma_{21}}{\sigma^2}, \boldsymbol{I}_g\right)'. \tag{33}$$

For the actual computation of standard errors the many weak instruments asymptotic distribution is not needed, since the many-instruments standard errors of Theorem 1 remain consistent.

5 Monte Carlo simulations

We compare the finite sample properties of the HLIM and SJIVE and their Fuller modifications given by (11) and (27), respectively. We use the same Monte Carlo set up as Hausman et al. (2011).

The data generating process is given by $\boldsymbol{y} = \boldsymbol{\iota}\gamma + \boldsymbol{x\beta} + \boldsymbol{\varepsilon}$ and $x = \boldsymbol{z}\pi + \boldsymbol{v}$, 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, $\boldsymbol{z} \sim \mathcal{N}\left(\mathbf{0},\ \boldsymbol{I}_n\right)$ and independently $\boldsymbol{v} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{I}_n\right)$. The disturbances $\boldsymbol{\varepsilon}$ are generated by

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

where $\rho = 0.3$, $\psi = 0.86$ and conditional on $\boldsymbol{z}$, independent of $\boldsymbol{v}$, $\boldsymbol{w}_1 \sim \mathcal{N}(\boldsymbol{0},\ \mathrm{Diag}(\boldsymbol{z})^2)$ and $\boldsymbol{w}_2 \sim \mathcal{N}(\boldsymbol{0},\ \psi^2 \boldsymbol{I}_n)$. The values $\phi = 0$ and $\phi = 1.38072$ are chosen such that the R-squared between ε_i^2 and the instruments equals 0 and 0.2, respectively.[1] The instruments $\boldsymbol{Z}$ are given for $k = 2$, $k = 10$ and $k = 30$ by matrices with rows $(1, z_i)$, $(1, z_i, z_i^2, z_i^3, z_i^4, z_i b_{1i}, \ldots, z_i b_{5i})$ and $(1, z_i, z_i^2, z_i^3, z_i^4, z_i b_{1i}, \ldots, z_i b_{25i})$, respectively, where independent of other random variables, the elements b_{ji} are i.i.d. Bernoulli distributed with $p = 1/2$. We used 20,000 simulations.

Figure 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. (2011), 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 also very similar 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. Figure 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.

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

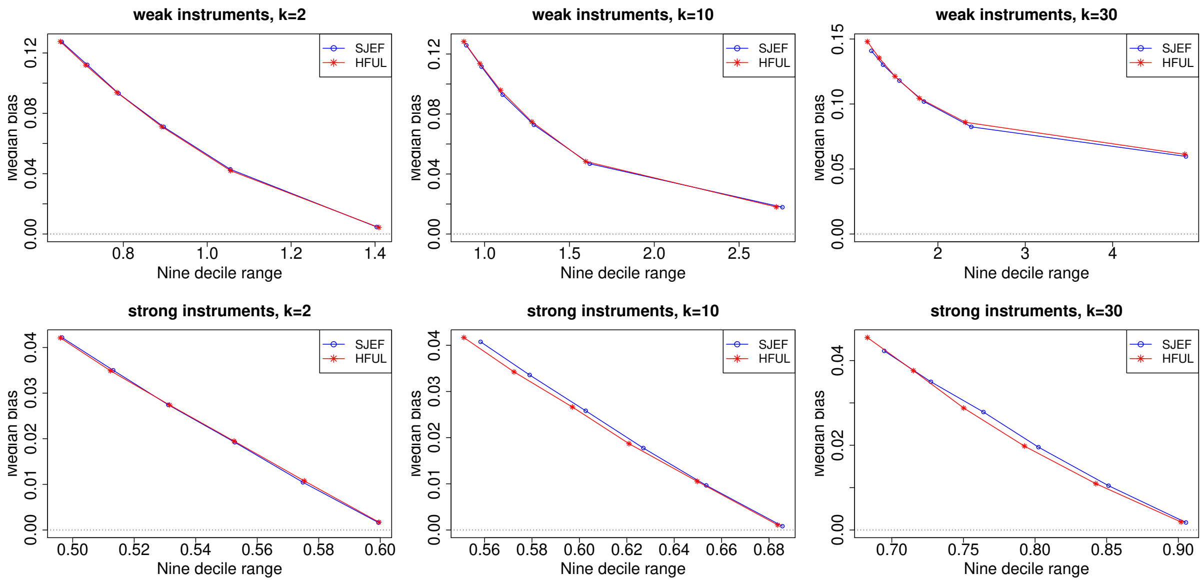

Figure 1: $R^2_{\varepsilon^2|z} = 0$: Median bias against the Nine decimal range of HFUL with $c = 0, 1, 2, 3, 4, 5$ from right to left, and SJEF for $\alpha = 0, 1, 2, 3, 4, 5$ from right to left, based on 20,000 replications.

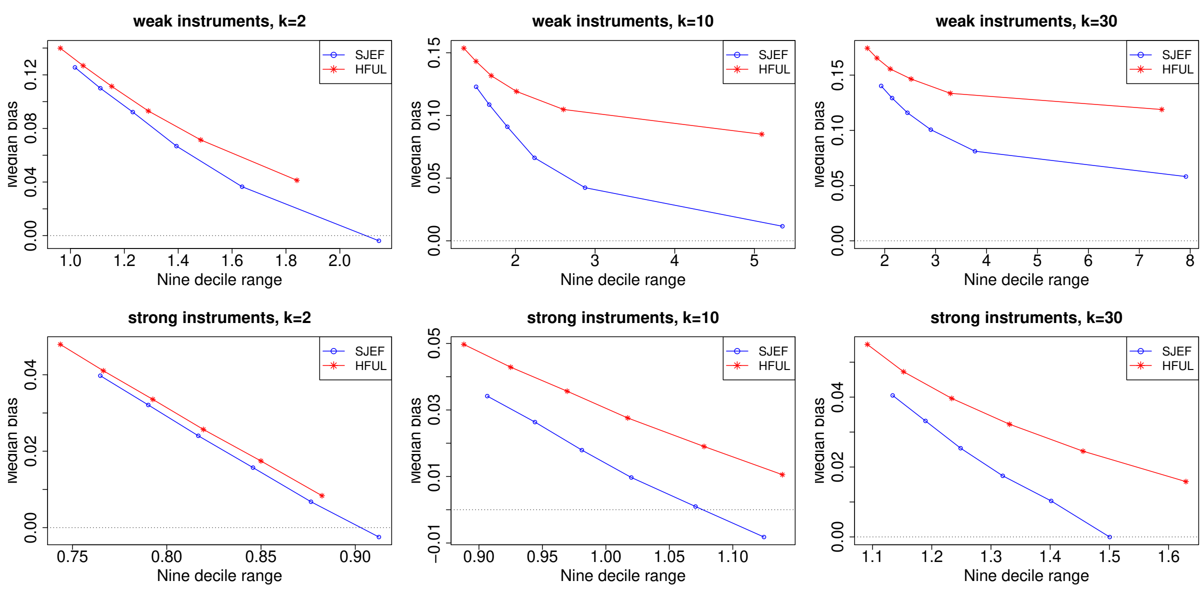

Figure 2: $R^2_{\varepsilon^2|z} = 0.2$: Median bias against the Nine decimal range of HFUL with $c = 0, 1, 2, 3, 4, 5$ from right to left, and SJEF for $\alpha = 0, 1, 2, 3, 4, 5$ from right to left, based on 20,000 replications.

		Median bias		*Nine decile range*		*Rejection rates*	
μ^2	k	HFUL	SJEF	HFUL	SJEF	HFUL	SJEF
8	2	0.071	0.067	1.484	1.393	0.033	0.026
8	10	0.105	0.066	2.603	2.240	0.044	0.025
8	30	0.134	0.101	3.291	2.906	0.043	0.036
32	2	0.017	0.016	0.850	0.846	0.048	0.043
32	10	0.019	0.010	1.077	1.020	0.047	0.038
32	30	0.025	0.017	1.455	1.320	0.045	0.041

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.

6 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 jackknife estimator. Asymptotic theory based on high level assumptions, which allow for both many instruments and many weak instruments, resulted in a concise formulation of asymptotic distributions and standard errors. A Monte Carlo comparison with the HFUL estimator of Hausman et al. (2011) showed the symmetric jackknife estimator SJIVE performs better.

7 Appendix

7.1 Derivation of Theorem 1

To derive Theorem 1 we use the notation $\boldsymbol{\delta} = (1, -\boldsymbol{\beta}')'$ and $\boldsymbol{M} = (\boldsymbol{y}, \boldsymbol{X})'\boldsymbol{A}_{\mathrm{HR}}(\boldsymbol{y}, \boldsymbol{X})$ in addition to the definitions of $\boldsymbol{S}$ and $\boldsymbol{C}$ in Assumption 3. We find

$$\mathrm{E}\{\boldsymbol{M} - \boldsymbol{S}\}\boldsymbol{\delta} = \boldsymbol{0}, \tag{34}$$

$$\begin{aligned}
&\mathrm{Var}\left[\{\boldsymbol{M} - \boldsymbol{S}\}\boldsymbol{\delta}\right] = \mathrm{E}\left\{(\boldsymbol{y}, \boldsymbol{X})'\boldsymbol{C}(\boldsymbol{y}, \boldsymbol{X})\boldsymbol{\delta}\boldsymbol{\delta}'(\boldsymbol{y}, \boldsymbol{X})'\boldsymbol{C}(\boldsymbol{y}, \boldsymbol{X})\right\} \\
&= \mathrm{E}\sum_{i=1}^{n}\sum_{s=1}^{n}(\boldsymbol{y}, \boldsymbol{X})'\boldsymbol{C}\boldsymbol{e}_i\boldsymbol{e}_i'(\boldsymbol{y}, \boldsymbol{X})\boldsymbol{\delta}\boldsymbol{\delta}'(\boldsymbol{y}, \boldsymbol{X})\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{C}(\boldsymbol{y}, \boldsymbol{X}) \\
&= \mathrm{E}\sum_{i}\varepsilon_i^2(\boldsymbol{y}, \boldsymbol{X})'\boldsymbol{C}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{C}(\boldsymbol{y}, \boldsymbol{X}) + \mathrm{E}\sum_{i=1}^{n}\sum_{s=1}^{n}\varepsilon_i\varepsilon_j(C_{ij})^2(\boldsymbol{y}, \boldsymbol{X})'\boldsymbol{e}_j\boldsymbol{e}_i'(\boldsymbol{y}, \boldsymbol{X}), \qquad (35) \\
&= \sum_{i=1}^{n}\sigma_i^2\begin{pmatrix}\boldsymbol{\beta}' \\ \boldsymbol{I}_g\end{pmatrix}\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{C}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{C}\boldsymbol{Z}\boldsymbol{\Pi}(\boldsymbol{\beta}, \boldsymbol{I}_g) + \sum_{i=1}^{n}\sum_{s=1}^{n}C_{ij}^2\left(\sigma_j^2\boldsymbol{\Omega}_i + \boldsymbol{\Omega}_i\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\Omega}_j\right). \qquad (36)
\end{aligned}$$

Using Assumption 2 we find

$$\sum_{i=1}^{n}\sum_{s=1}^{n}C_{ij}^2\left(\sigma_j^2\boldsymbol{\Omega}_i + \boldsymbol{\Omega}_i\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\Omega}_j\right) \le c_u^2(1 + \boldsymbol{\delta}'\boldsymbol{\delta})\sum_{i=1}^{n}\sum_{s=1}^{n}C_{ij}^2\boldsymbol{I}_{g+1} = c_u^2(1 + \boldsymbol{\delta}'\boldsymbol{\delta})\,\mathrm{tr}(\boldsymbol{C}^2)\boldsymbol{I}_{g+1},$$

so by Assumption 1 the second term in (36) is of order $O(k)$ just as $\mathrm{tr}(C^2)$ is. Consequently,

$$\boldsymbol{\delta}'(\boldsymbol{M} - \boldsymbol{S})\boldsymbol{\delta} = O_p(k^{1/2}). \tag{37}$$

For the first term we find

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

so by Assumption 5, where $k/r_n \to \gamma$ we find

$$(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}(\mathbf{0}, \boldsymbol{I}_g)(\boldsymbol{M} - \boldsymbol{S})\boldsymbol{\delta} = O_p(1). \tag{38}$$

The first derivative of the objective function is given by

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

Using Assumptions 3 and 5 we find,

$$\begin{aligned}
&-\frac{1}{2}\left\{\frac{\boldsymbol{\delta}'\boldsymbol{S}\boldsymbol{\delta}}{\operatorname{tr}(\boldsymbol{B}_{\text{HR}})}\right\}(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}\frac{\partial Q^*_{\text{SJIVE}}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \\
&\quad(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}\left\{(\mathbf{0}, \boldsymbol{I}_g) - \sigma^{-2}\boldsymbol{\sigma}_{21}\boldsymbol{\delta}'\right\}\{\boldsymbol{M} - \boldsymbol{S}\}\boldsymbol{\delta} \quad - \\
&\quad\left(\frac{k}{r_{\min}}\right)^{1/2}\left(\frac{\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi}}{r_{\min}}\right)^{-1/2}\left\{\frac{(\mathbf{0}, \boldsymbol{I}_g)\boldsymbol{S}\boldsymbol{\delta}}{\boldsymbol{\delta}'\boldsymbol{S}\boldsymbol{\delta}} - \frac{\boldsymbol{\sigma}_{21}}{\sigma^2}\right\}k^{-1/2}\boldsymbol{\delta}'\{\boldsymbol{M} - \boldsymbol{S}\}\boldsymbol{\delta} \quad = \\
&\quad(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}\left\{(\mathbf{0}, \boldsymbol{I}_g) - \sigma^{-2}\boldsymbol{\sigma}_{21}\boldsymbol{\delta}'\right\}\{\boldsymbol{M} - \boldsymbol{S}\}\boldsymbol{\delta} + o_p(1).
\end{aligned} \tag{39}$$

The second derivative of the objective function is given by

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

where

$$\boldsymbol{F} = -2(\boldsymbol{M} - \boldsymbol{S})\boldsymbol{\delta}\frac{\boldsymbol{\delta}'\boldsymbol{S}}{\boldsymbol{\delta}'\boldsymbol{S}\boldsymbol{\delta}} - 2\frac{\boldsymbol{S}\boldsymbol{\delta}}{\boldsymbol{\delta}'\boldsymbol{S}\boldsymbol{\delta}}\boldsymbol{\delta}'(\boldsymbol{M} - \boldsymbol{S}) - \boldsymbol{\delta}'(\boldsymbol{M} - \boldsymbol{S})\boldsymbol{\delta}\left(\frac{\boldsymbol{S}}{\boldsymbol{\delta}'\boldsymbol{S}\boldsymbol{\delta}} - 4\frac{\boldsymbol{S}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{S}}{(\boldsymbol{\delta}'\boldsymbol{S}\boldsymbol{\delta})^2}\right).$$

Due to (37) and (38) we find

$$(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}(\mathbf{0},\boldsymbol{I}_g)\boldsymbol{F}\begin{pmatrix}\mathbf{0}\\ \boldsymbol{I}_g\end{pmatrix}(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}=o_p(1).$$

Consequently

$$\frac{1}{2}\left\{\frac{\boldsymbol{\delta}'\boldsymbol{S}\boldsymbol{\delta}}{\operatorname{tr}(\boldsymbol{B}_{\text{HR}})}\right\}(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}\frac{\partial^2 Q^*_{\text{SJIVE}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}=$$
$$(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}(\mathbf{0},\boldsymbol{I}_g)\left\{\boldsymbol{M}-\boldsymbol{S}\right\}\begin{pmatrix}\mathbf{0}\\ \boldsymbol{I}_g\end{pmatrix}(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}+o_p(1).$$

Based on Assumption 3 we thus find

$$\begin{aligned}
\boldsymbol{H}&=(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}\frac{\partial^2 Q^*_{\text{SJIVE}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}\\
&=2\sigma^{-2}\left\{(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}\boldsymbol{X}'\boldsymbol{C}\boldsymbol{X}(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}\right\}+o_p(1)\\
&=2\sigma^{-2}\left\{(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}\operatorname{E}(\boldsymbol{X}'\boldsymbol{C}\boldsymbol{X})(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}\right\}+o_p(1)\\
&=2\sigma^{-2}\boldsymbol{I}_g+o_p(1). \qquad (40)
\end{aligned}$$

Assumption 4 says $r_n\to\infty$, so (40) implies that $\lambda_{\min}\left(\frac{\partial^2 Q^*_{\text{SJIVE}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right)\to\infty$, and by Assumption 5 we find $\hat{\boldsymbol{\beta}}\xrightarrow{p}\boldsymbol{\beta}$. Finally we find, applying (36), (39) and 40,

$$\begin{aligned}
\boldsymbol{\Psi}=\lim_{n\to\infty}\Bigg[(\boldsymbol{\Pi}\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}\Bigg\{&\sum_{i=1}^n\sigma_i^2\boldsymbol{\Pi}\boldsymbol{Z}'\boldsymbol{F}\boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{P}\boldsymbol{F}\boldsymbol{Z}\boldsymbol{\Pi}\\
&+\sum_{i=1}^n\sum_{s=1}^n C_{ij}^2\left(-\frac{\boldsymbol{\sigma}_{21}}{\sigma^2},\boldsymbol{I}_g\right)\left(\sigma_j^2\boldsymbol{\Sigma}_i+\boldsymbol{\Sigma}_i\boldsymbol{e}_1\boldsymbol{e}_1'\boldsymbol{\Sigma}_j\right)\left(-\frac{\boldsymbol{\sigma}_{21}}{\sigma^2},\boldsymbol{I}_g\right)'\Bigg\}(\boldsymbol{\Pi}\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{-1/2}\Bigg].
\end{aligned}$$

To compute standard errors, we estimate $\operatorname{Var}\left[\{\boldsymbol{M}-\operatorname{tr}(\boldsymbol{P}\boldsymbol{F}\boldsymbol{P})\boldsymbol{S}\}\boldsymbol{\delta}\right]$ using (35). A consistent estimator for $\boldsymbol{\Psi}$ is then given by

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

where

$$\widehat{\mathrm{Var}}\left[\{\boldsymbol{M}-\mathrm{tr}(\boldsymbol{PFP})\boldsymbol{S}\}\boldsymbol{\delta}\right] = \sum_i \hat{\varepsilon}_i^2(\boldsymbol{y},\boldsymbol{X})'\widehat{\boldsymbol{C}}\boldsymbol{e}_i\boldsymbol{e}_i'\widehat{\boldsymbol{C}}(\boldsymbol{y},\boldsymbol{X}) \quad + \\ \sum_{i=1}^n\sum_{s=1}^n \hat{\varepsilon}_i\hat{\varepsilon}_j(\widehat{C}_{ij})^2(\boldsymbol{y},\boldsymbol{X})'\boldsymbol{e}_j\boldsymbol{e}_i'(\boldsymbol{y},\boldsymbol{X}).$$

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

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

which is (28).

7.2 Derivation of Theorem 2

Instead of (39) we now have, using Assumptions 3 and 6,

$$\begin{aligned}-\frac{k^{-1/2}}{2}\left\{\frac{\boldsymbol{\delta}'\boldsymbol{S}\boldsymbol{\delta}}{\mathrm{tr}(\boldsymbol{B}_{\mathrm{HR}})}\right\}\frac{\partial Q^*_{\mathrm{SJIVE}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} &= \\ & k^{-1/2}\left\{(\boldsymbol{0},\boldsymbol{I}_g)-\sigma^{-2}\boldsymbol{\sigma}_{21}\boldsymbol{\delta}'\right\}(\boldsymbol{M}-\boldsymbol{S})\boldsymbol{\delta} \quad - \\ & \left\{\frac{(\boldsymbol{0},\boldsymbol{I}_g)\boldsymbol{S}\boldsymbol{\delta}}{\boldsymbol{\delta}'\boldsymbol{S}\boldsymbol{\delta}}-\frac{\boldsymbol{\sigma}_{21}}{\sigma^2}\right\}k^{-1/2}\boldsymbol{\delta}'(\boldsymbol{M}-\boldsymbol{S})\boldsymbol{\delta} \\ &= k^{-1/2}\left\{(\boldsymbol{0},\boldsymbol{I}_g)-\sigma^{-2}\boldsymbol{\sigma}_{21}\boldsymbol{\delta}'\right\}(\boldsymbol{M}-\boldsymbol{S})\boldsymbol{\delta}+o_p(1). \qquad (41)\end{aligned}$$

Using (36) we thus find

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

As (40) remains valid under Assumption 6 we find the result of Theorem 2, where $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$ since $r_{\min}/k^{1/2} \to \infty$.

References

Ackerberg, D. A., P. J. Devereux (2003): 'Improved JIVE estimators for overidentified linear models with and without heteroskedasticity', *Working paper.*

Andrews, D. W. K., J. H. Stock (2007): 'Testing with many weak instruments', *Journal of Econometrics*, **138**, 24-46.

Angrist, J. D., G. W. Imbens, A. Krueger (1999): 'Jackknife instrumental variable estimators', *Journal of Applied Econometrics*, **14**, 57-67.

Bekker, P. A., (1994): 'Alternative approximations to the distributions of instrumental variable estimators', *Econometrica*, **54**, 657-682.

Bekker, P. A., J. van der Ploeg (2005): 'Instrumental variable estimation based on grouped data', *Statistica Neerlandica*, **59**, 239-267.

Blomquist, S., M. Dahlberg (1999): 'Small sample properties of LIML and Jackknife IV estimators: experiments with weak instruments', *Journal of Applied Econometrics*, **14**, 69-88.

Chamberlain, G., G. Imbens (2004): 'Random effects estimators with many instrumental variables', *Econometrica*, **72**, 295-306.

Chao, J., N. Swanson (2005): 'Consistent estimation with a large number of weak instruments', *Econometrica*, **73**, 1673-1692.

Davidson, R., J. G. MacKinnon (2006): 'The case against JIVE', *Journal of Applied Econometrics*, **21**, 827-833.

Donald, S. J., W. K. Newey (2000): 'A jackknife interpretation of the continuous updating estimator', *Economics Letters*, **67**, 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., J. A. Hausman, G. M. Kuersteiner (2004): 'Estimation with weak instruments:

accuracy of higher-order bias and MSE approximations', *Econometrics Journal*, **7**, 272-306.

Hahn, J., A. Inoue (2002): 'A Monte Carlo comparison of various asymptotic approximations to the distribution of instrumental variables estimators', *Econometric Reviews*, **21**, 309-336.

Han, C., P. C. B. Phillips (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., J. Heaton, A. Yaron (1996): 'Finite-sample properties of some alternative GMM estimators', *Journal of Business and Economic Statistics*, **14**, 398-422.

Hausman, J. A., W. K. Newey, T. Woutersen, J. Chao, N. Swanson (2011): 'Instrumental variable estimation with heteroskedasticity and many instruments', *Working paper.*

Kunitomo, N., (1980): 'Asymptotic expansions of distributions of estimators in a linear functional relationship and simultaneous equations', *Journal of the American Statistical Association*, **75**, 693-700.

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., F. Windmeijer (2009): 'Generalized method of moments with many weak moment condition', *Econometrica*, **77**, 687-719.

Phillips, G. D. A., C. Hale (1977): 'The bias of instrumental variable estimators of simultaneous equation systems', *International Economic Review*, **18**, 219-228.

Stock, J., M. Yogo (2005): 'Asymptotic distributions of instrumental variables statistics with many instruments', in *Identification and inference for econometric models: essays in honor of Thomas Rothenberg*, Ch. 6, D. W. K Andrews and J. H. Stock eds., Cambridge University Press.