University of Groningen

Simple many-instruments robust standard errors through concentrated instrumental variables

Bekker, Paul; Wansbeek, Tom

Published in:
Economics Letters

DOI:
[10.1016/j.econlet.2016.09.017](https://doi.org/10.1016/j.econlet.2016.09.017)

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:
2016

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

Citation for published version (APA):

Bekker, P., & Wansbeek, T. (2016). Simple many-instruments robust standard errors through concentrated instrumental variables. *Economics Letters*, 149, 52-55. <https://doi.org/10.1016/j.econlet.2016.09.017>

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.

Simple many-instruments robust standard errors through concentrated instrumental variables

Paul Bekker, Tom Wansbeek*

Faculty of Economics and Business, University of Groningen, Netherlands

HIGHLIGHTS

- Concentrated Instrumental Variables (CIV) are proposed as an alternative to LIML.
- Reformulating LIML through CIV leads to simple many-instruments robust standard errors.
- A two-step, 2SLS-based estimator is proposed that is almost as good as LIML.

ARTICLE INFO

Article history:

Received 10 August 2016

Received in revised form

16 September 2016

Accepted 19 September 2016

Available online 11 October 2016

JEL classification:

C2

Keywords:

LIML

Weak instruments

Concentrated instruments

ABSTRACT

In a weak and many instruments setting, 2SLS can be severely biased towards OLS and the standard errors can be way too small. LIML is an attractive alternative, especially when the many-instruments robust (MIR) standard errors are used as proposed by Bekker (1994).

In this note we present an alternative approach to LIML through concentrated instrumental variables (CIV). This is a class of instrumental variables indexed by a single parameter. When this parameter belongs to a certain set, 2SLS estimators using CIV instruments, are many-instruments consistent. We refer to them as CIV estimators. Moreover, the usual expression for the asymptotic variance of IV estimators is many-instruments consistent. One particular choice of CIV parameter, within this set, yields LIML. We thus get an alternative to the many-instruments robust (MIR) standard errors of Bekker (1994) that is of the usual simple IV form.

As a byproduct, we present a new estimator, which we call the CIVE. This is a two-step estimator where the CIV parameter is based on 2SLS in the first step. It avoids LIML but is equally good to a very high degree. This suggests a simple MIR estimation strategy in a weak and many instruments setting: do 2SLS to obtain the CIV parameter, and then obtain the CIVE by another 2SLS step.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

In a weak and many instruments setting, two-stage least squares (2SLS) can be severely biased towards ordinary least squares (OLS) and the standard errors can be way too small. LIML (limited-information maximum likelihood) is an attractive alternative, especially when the many-instruments robust (MIR) standard errors are used as proposed by Bekker (1994). These are sometimes referred to as “Bekker standard errors” and are based on many-instruments asymptotics as introduced by Anderson (1976).

In this note we present an alternative approach to LIML through concentrated instrumental variables (CIV). This is a class of

instrumental variables whose number remains fixed under many-instruments asymptotics. It is indexed by a single parameter. When this parameter belongs to a certain set, the 2SLS estimator based on the CIV instruments is many-instruments consistent, and the usual expression for the asymptotic variance of IV estimators is many-instruments consistent. One particular choice of parameter, within this set, yields LIML. We thus get an alternative to the original MIR standard errors that is of the usual simple IV form.

As a byproduct, we present a new estimator, which we call the CIVE. This is a two-step estimator where the CIV parameter is based on 2SLS in the first step. It avoids LIML but is, in terms of consistency and asymptotic efficiency, equally good to a very high degree. This suggests a simple estimation strategy in a many-instruments setting: first, do 2SLS to obtain the CIV parameter, and then obtain the CIVE by another 2SLS step.

The paper is organized as follows. In Section 2 we present the model and some further notation, and describe many-instruments asymptotics. In Section 3, we introduce the notion of

* Corresponding author.

E-mail addresses: p.a.bekker@rug.nl (P. Bekker), t.j.wansbeek@rug.nl (T. Wansbeek).

concentrated instrumental variables. We show that LIML is a case of concentrated instruments, and use this to derive a simple and “natural” alternative to MIR standard errors. Section 4 introduces the CIVE. We conclude with some simulation experiments, reported in Section 5.

2. LIML and many-instruments asymptotics

Consider the structural form of the classical normal linear IV model given by

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i,$$

$$\mathbf{x}_i = \boldsymbol{\Pi}' \mathbf{z}_i + \mathbf{v}_i,$$

where $\mathbf{x}_i$ has m elements and the nonrandom vector of instrumental variables $\mathbf{z}_i$ has $k \geq m + 1$ elements, $i = 1, \dots, n$. So there are instruments for all observed variables in $\mathbf{x}_i$ and y_i . In matrix notation,

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

$$\mathbf{X} = \mathbf{Z}\boldsymbol{\Pi} + \mathbf{V}.$$

The errors ε_i and $\mathbf{v}_i$ are i.i.d. over i .¹ We assume

$$\begin{pmatrix} \varepsilon_i \\ \mathbf{v}_i \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}), \quad \text{with } \boldsymbol{\Omega} = \begin{pmatrix} \sigma^2 & \boldsymbol{\omega}' \\ \boldsymbol{\omega} & \boldsymbol{\Omega}_v \end{pmatrix}.$$

We write $\mathbf{P}_A \equiv \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ and $\mathbf{M}_A \equiv \mathbf{I} - \mathbf{P}_A$ for any matrix $\mathbf{A}$ of full column rank. Also, for any r ,

$$\mathbf{C}(r) \equiv \mathbf{P}_Z - r\mathbf{M}_Z.$$

Following Nagar (1959), let $\tilde{\mathbf{V}} \equiv \mathbf{V} - \boldsymbol{\varepsilon}\boldsymbol{\omega}'/\sigma^2$ and $\tilde{\mathbf{X}} \equiv \mathbf{X} - \boldsymbol{\varepsilon}\boldsymbol{\omega}'/\sigma^2$. With $\tilde{\mathbf{v}}_i$ the i th column of $\tilde{\mathbf{V}}$,

$$\boldsymbol{\Omega}_{v|\varepsilon} \equiv E(\tilde{\mathbf{v}}_i \tilde{\mathbf{v}}_i') = \boldsymbol{\Omega}_v - \boldsymbol{\omega}\boldsymbol{\omega}'/\sigma^2.$$

Notice that $\tilde{\mathbf{X}} = \mathbf{Z}\boldsymbol{\Pi} + \tilde{\mathbf{V}}$ and that $\boldsymbol{\varepsilon}$ and $\tilde{\mathbf{X}}$ are independent.

Anderson and Rubin (1949, 1950) showed that the maximum likelihood estimator of $\boldsymbol{\beta}$ in this model, for historical reasons called the LIML estimator, follows from minimization of $\boldsymbol{\varepsilon}'\mathbf{P}_Z\boldsymbol{\varepsilon}/\boldsymbol{\varepsilon}'\mathbf{M}_Z\boldsymbol{\varepsilon}$. This minimization leads to an eigenequation in $\hat{\boldsymbol{\beta}}$,

$$(\mathbf{y}, \mathbf{X})' \mathbf{C}(\hat{\lambda})(\mathbf{y}, \mathbf{X}) \begin{pmatrix} 1 \\ -\hat{\boldsymbol{\beta}} \end{pmatrix} = \mathbf{0}, \quad (1)$$

where

$$\hat{\lambda} \equiv \frac{\hat{\boldsymbol{\varepsilon}}'\mathbf{P}_Z\hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}'\mathbf{M}_Z\hat{\boldsymbol{\varepsilon}}}, \quad (2)$$

with $\hat{\boldsymbol{\varepsilon}} \equiv \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, is the smallest value for which the matrix $(\mathbf{y}, \mathbf{X})' \mathbf{C}(\hat{\lambda})(\mathbf{y}, \mathbf{X})$ is singular.

With weak and many instruments, asymptotic theory of the usual large n -type can be a poor guide to the exact distribution of the estimator. An improvement is obtained by the use of many-instruments asymptotics, characterized by

$$n \rightarrow \infty, \quad k/n = \alpha + o(n^{-1/2}), \quad \alpha \geq 0, \quad n^{-1} \boldsymbol{\Pi}' \mathbf{Z}' \mathbf{Z} \boldsymbol{\Pi} \rightarrow \mathbf{Q} > \mathbf{0}.$$

Below, α will appear mostly through

$$\lambda \equiv \frac{\alpha}{1 - \alpha} \quad \text{so} \quad \frac{k}{n - k} \rightarrow \lambda.$$

Large- n asymptotics (sometimes called fixed- k asymptotics) arises as the special case with $\alpha = \lambda = 0$. Crucially, the LIML estimator is consistent under many-instruments asymptotics while 2SLS is not. We will use many-instruments asymptotics throughout.

The asymptotic variance of the LIML estimator $\hat{\boldsymbol{\beta}}$ is shown by Bekker (1994) to be

$$\text{avar}(\hat{\boldsymbol{\beta}}) = \sigma^2 \mathbf{Q}^{-1} (\mathbf{Q} + \lambda \boldsymbol{\Omega}_{v|\varepsilon}) \mathbf{Q}^{-1}, \quad (3)$$

albeit in a different form and notation.² To make (3) operational, he proposes

$$\widehat{\text{avar}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 \mathbf{B}_{22}^{-1} [\mathbf{C}_{22} + \hat{\lambda} (\mathbf{C}_{22} - \mathbf{B}_{22})] \mathbf{B}_{22}^{-1},$$

which is MIR consistent, with $\hat{\sigma}^2 = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}/(n - m)$ and

$$\mathbf{B} \equiv \bar{\mathbf{S}} - \hat{\lambda} \mathbf{S}^\perp,$$

$$\mathbf{C} \equiv \bar{\mathbf{S}} - \hat{\lambda} \frac{\mathbf{S}^\perp (1, -\hat{\boldsymbol{\beta}}')' (1, -\hat{\boldsymbol{\beta}}') \mathbf{S}^\perp}{(1, -\hat{\boldsymbol{\beta}}') \mathbf{S}^\perp (1, -\hat{\boldsymbol{\beta}}')'},$$

$$\bar{\mathbf{S}} \equiv \frac{1}{n - m} (\mathbf{y}, \mathbf{X})' \mathbf{P}_Z (\mathbf{y}, \mathbf{X}),$$

$$\mathbf{S}^\perp \equiv \frac{1}{n - m} (\mathbf{y}, \mathbf{X})' \mathbf{M}_Z (\mathbf{y}, \mathbf{X});$$

the subscript “22” to a matrix denotes its lower-right $m \times m$ submatrix.

3. Concentrated instrumental variables

The theory underlying MIR standard errors is complicated and the shape of MIR standard errors is not intuitively appealing. A major simplification is possible when we consider estimators of the form

$$\hat{\boldsymbol{\beta}}(r) \equiv (\mathbf{X}' \mathbf{P}_{Z(r)} \mathbf{X})^{-1} \mathbf{X}' \mathbf{P}_{Z(r)} \mathbf{y}, \quad (4)$$

for any (possibly stochastic) scalar r , where

$$\mathbf{Z}(r) \equiv \mathbf{C}(r)(\mathbf{y}, \mathbf{X}).$$

So $\hat{\boldsymbol{\beta}}(r)$ is the 2SLS estimator with instruments $\mathbf{Z}(r)$. We call $\mathbf{Z}(r)$ concentrated instrumental variables (CIV).³ The number of instruments in $\mathbf{Z}(r)$, $m+1$, is fixed when the number of instruments $\mathbf{Z}$ increases. As a result, the projection matrix $\mathbf{P}_{Z(r)}$ has constant rank even in a many-instruments setting.

We first show that 2SLS and LIML are CIV estimators. For 2SLS, note that

$$\begin{aligned} \mathbf{P}_{Z(0)} \mathbf{X} &= \mathbf{P}_Z (\mathbf{y}, \mathbf{X}) [(\mathbf{y}, \mathbf{X})' \mathbf{P}_Z (\mathbf{y}, \mathbf{X})]^{-1} (\mathbf{y}, \mathbf{X})' \mathbf{P}_Z (\mathbf{y}, \mathbf{X}) (\mathbf{0}, \mathbf{I}_m)' \\ &= \mathbf{P}_Z \mathbf{X}, \end{aligned}$$

so 2SLS is the CIV estimator with $r = 0$. The LIML estimator $\hat{\boldsymbol{\beta}}$ follows from (1), which we can write as $\mathbf{Z}(\hat{\lambda})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0$ or again $\mathbf{P}_{Z(\hat{\lambda})} \mathbf{y} = \mathbf{P}_{Z(\hat{\lambda})} \mathbf{X}\hat{\boldsymbol{\beta}}$, and

$$\hat{\boldsymbol{\beta}}(\hat{\lambda}) = (\mathbf{X}' \mathbf{P}_{Z(\hat{\lambda})} \mathbf{X})^{-1} \mathbf{X}' \mathbf{P}_{Z(\hat{\lambda})} \mathbf{y} = \hat{\boldsymbol{\beta}}.$$

So the LIML estimator is the CIV estimator with $r = \hat{\lambda}$.

We now discuss the consistency of $\hat{\boldsymbol{\beta}}(r)$. It is 2SLS with instruments $\mathbf{Z}(r) \equiv \mathbf{C}(r)(\mathbf{y}, \mathbf{X})$. Since $(\mathbf{y}, \mathbf{X})$ spans the same space as $\mathbf{R} \equiv (\boldsymbol{\varepsilon}, \tilde{\mathbf{X}})$, $\hat{\boldsymbol{\beta}}(r)$ can equally be seen as 2SLS with instruments $\tilde{\mathbf{Z}}(r) \equiv \mathbf{C}(r)(\boldsymbol{\varepsilon}, \tilde{\mathbf{X}})$; using the latter simplifies the analysis of the

¹ In this paper we cover the case of homoskedasticity. Whether or not the approach to be presented in the sequel can be extended to the case of unknown heteroskedasticity as discussed by Chao et al. (2012), Hausman et al. (2012), and Bekker and Cruje (2015) is left for future research.

² See Newey (2004) for the form (3).

³ The CIV estimator $\hat{\boldsymbol{\beta}}(r)$ should not be confused with $\hat{\boldsymbol{\beta}}(k) \equiv [\mathbf{X}' \mathbf{C}(k) \mathbf{X}]^{-1} \mathbf{X}' \mathbf{C}(k) \mathbf{y}$, the k -class estimator of Theil (1953).

properties of $\hat{\beta}(r)$ as ε and $\tilde{\mathbf{X}}$ are independent. Notice that $\mathbf{X} = \mathbf{R}(\omega/\sigma^2, \mathbf{I}_m)'$ and $\varepsilon = \mathbf{R}\mathbf{e}_1$, with $\mathbf{e}_1$ the first unit vector. The essential factor in assessing consistency is

$$\begin{aligned}\mathbf{X}'\mathbf{P}_{\mathbf{Z}(r)}\varepsilon &= (\omega/\sigma^2, \mathbf{I}_m)'\mathbf{R}'\mathbf{P}_{\tilde{\mathbf{Z}}(r)}\mathbf{R}\mathbf{e}_1 \\ &= (\omega/\sigma^2, \mathbf{I}_m)'\mathbf{R}'\mathbf{C}(r)\mathbf{R}[\mathbf{C}'(r)^2\mathbf{R}]^{-1}\mathbf{R}'\mathbf{C}(r)\mathbf{R}\mathbf{e}_1.\end{aligned}$$

As ε and $\tilde{\mathbf{X}}$ are independent and $E(\varepsilon) = \mathbf{0}$, $n^{-1}\mathbf{R}'\mathbf{P}_{\mathbf{Z}}\mathbf{R}$ and $n^{-1}\mathbf{R}'\mathbf{M}_{\mathbf{Z}}\mathbf{R}$ converge in probability to block-diagonal matrices. Therefore, with $r \xrightarrow{p} \rho$, say, the upper-left element of the probability limit of $n^{-1}\mathbf{R}'\mathbf{C}(r)\mathbf{R}$ equals $[\alpha - \rho(1 - \alpha)]\sigma^2$ for $\tau = 1$ and $[\alpha + \rho^2(1 - \alpha)]\sigma^2$ for $\tau = 2$. Consequently,

$$n^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{Z}(r)}\varepsilon \xrightarrow{p} (1 - \alpha)\omega \frac{(\lambda - \rho)^2}{\lambda + \rho^2}. \quad (5)$$

So consistency requires that $\rho = \lambda$.

We next derive the asymptotic distribution of consistent CIV estimators. First let $r = \lambda$ and notice

$$\begin{aligned}\text{tr}[\mathbf{C}(\lambda)] &= k - \lambda(n - k) \quad \text{so } \text{tr}[\mathbf{C}(\lambda)] = o(n^{1/2}) \\ \text{tr}[\mathbf{C}(\lambda)^2] &= k + \lambda^2(n - k) \quad \text{so } \text{tr}[\mathbf{C}(\lambda)^2] = n\lambda + o(n^{1/2}).\end{aligned}$$

Proceeding again with $\tilde{\mathbf{Z}}(\cdot)$ rather than $\mathbf{Z}(\cdot)$, there holds

$$\begin{aligned}n^{-1}\tilde{\mathbf{Z}}(\lambda)'\mathbf{X} &= n^{-1}(\varepsilon, \tilde{\mathbf{X}})'\mathbf{C}(\lambda)(\tilde{\mathbf{X}} + \varepsilon\omega'/\sigma^2) \xrightarrow{p} \begin{pmatrix} \mathbf{0}' \\ \mathbf{Q} \end{pmatrix}, \\ n^{-1}\tilde{\mathbf{Z}}(\lambda)'\tilde{\mathbf{Z}}(\lambda) &= n^{-1}(\varepsilon, \tilde{\mathbf{X}})'\mathbf{C}(\lambda)^2(\varepsilon, \tilde{\mathbf{X}}) \\ &\xrightarrow{p} \begin{pmatrix} \lambda\sigma^2 & \mathbf{0}' \\ \mathbf{0} & \mathbf{Q} + \lambda\Omega_{\mathbf{V}|\varepsilon} \end{pmatrix}.\end{aligned}$$

Since $n^{-1/2}\tilde{\mathbf{Z}}(\lambda)'\varepsilon$ is bounded in probability and

$$n^{-1/2}\tilde{\mathbf{X}}'\mathbf{C}(\lambda)\varepsilon \stackrel{a}{\sim} \mathcal{N}(\mathbf{0}, \sigma^2(\mathbf{Q} + \lambda\Omega_{\mathbf{V}|\varepsilon})),$$

we obtain

$$\begin{aligned}n^{-1/2}\mathbf{X}'\mathbf{P}_{\tilde{\mathbf{Z}}(\lambda)}\varepsilon &= n^{-1/2}[\mathbf{0}, \mathbf{Q}(\mathbf{Q} + \lambda\Omega_{\mathbf{V}|\varepsilon})^{-1}](\varepsilon, \tilde{\mathbf{X}})'\mathbf{C}(\lambda)\varepsilon + o_p(1) \\ &= \mathbf{Q}(\mathbf{Q} + \lambda\Omega_{\mathbf{V}|\varepsilon})^{-1}(n^{-1/2}\tilde{\mathbf{X}}'\mathbf{C}(\lambda)\varepsilon) + o_p(1) \\ &\sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{Q}(\mathbf{Q} + \lambda\Omega_{\mathbf{V}|\varepsilon})^{-1}\mathbf{Q}).\end{aligned}$$

Hence,

$$n^{1/2}(\hat{\beta}(\lambda) - \beta) \stackrel{a}{\sim} \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{Q}^{-1}(\mathbf{Q} + \lambda\Omega_{\mathbf{V}|\varepsilon})\mathbf{Q}^{-1}), \quad (6)$$

cf. (3). To make the estimator of the variance feasible we need to estimate λ . The asymptotic results apply as well when λ is replaced by $\tilde{\lambda} = \lambda + o_p(n^{-1/2})$. In that case $n^{-1/2}\tilde{\mathbf{Z}}(\tilde{\lambda})'\varepsilon = n^{-1/2}\tilde{\mathbf{Z}}(\lambda)'\varepsilon + o_p(1)$ and

$$n^{1/2}(\hat{\beta}(\tilde{\lambda}) - \beta) \stackrel{a}{\sim} \mathcal{N}(\mathbf{0}, \sigma^2 \text{plim}_{n \rightarrow \infty} n(\mathbf{X}'\mathbf{P}_{\tilde{\mathbf{Z}}(\tilde{\lambda})}\mathbf{X})^{-1}).$$

So any estimator with instruments $\mathbf{Z}(\tilde{\lambda}) = (\mathbf{P}_{\mathbf{Z}} - \tilde{\lambda}\mathbf{M}_{\mathbf{Z}})(\mathbf{y}, \mathbf{X})$ is many-instruments consistent, when $\tilde{\lambda} = \lambda + o_p(n^{-1/2})$. LIML is the leading case, and the standard errors that result from taking the square roots of the diagonal elements of

$$\hat{\sigma}^2(\mathbf{X}'\mathbf{P}_{\tilde{\mathbf{Z}}(\tilde{\lambda})}\mathbf{X})^{-1},$$

with $\hat{\lambda}$ from (2), offer an alternative to the original MIR standard errors of Bekker (1994). They might be called “natural” as they are just of the form that is familiar from basic IV theory. This is our main result.

4. CIVE

Among all CIV estimators there is one estimator that we will call “the” CIV estimator. We will use the acronym CIVE solely for

this estimator. The starting point is 2SLS. As we saw, 2SLS is the CIV estimator with $r = 0$. So 2SLS is inconsistent. Its inconsistency is proportional to λ and hence is $O(\alpha)$. In particular,

$$\hat{\beta}_{2\text{SLS}} \xrightarrow{p} \beta + \alpha(\mathbf{Q} + \alpha\Omega_{\mathbf{V}})^{-1}\omega. \quad (7)$$

Yet we can use the 2SLS estimator and compute its residual vector $\hat{\varepsilon}_{2\text{SLS}}$, to obtain

$$\hat{\lambda}_{2\text{SLS}} = \frac{\hat{\varepsilon}_{2\text{SLS}}'\mathbf{P}_{\mathbf{Z}}\hat{\varepsilon}_{2\text{SLS}}}{\hat{\varepsilon}_{2\text{SLS}}'\mathbf{M}_{\mathbf{Z}}\hat{\varepsilon}_{2\text{SLS}}}.$$

We call the CIV estimator based on this choice of r “the” CIV estimator, CIVE:

$$\hat{\beta}_{\text{CIVE}} \equiv \hat{\beta}(\hat{\lambda}_{2\text{SLS}}).$$

We investigate its inconsistency. In general, let $\hat{\beta}$ be such that $\hat{\beta} - \beta \xrightarrow{p} O(\alpha^i)$. Then

$$\begin{aligned}n^{-1}\hat{\varepsilon}'\mathbf{P}_{\mathbf{Z}}\hat{\varepsilon} &= n^{-1}[\varepsilon + \mathbf{X}(\beta - \hat{\beta})]'\mathbf{P}_{\mathbf{Z}}[\varepsilon + \mathbf{X}(\beta - \hat{\beta})] \\ &\xrightarrow{p} \alpha\sigma^2 + O(\alpha^{i+1}), \\ n^{-1}\hat{\varepsilon}'\mathbf{M}_{\mathbf{Z}}\hat{\varepsilon} &= n^{-1}[\varepsilon + \mathbf{X}(\beta - \hat{\beta})]'\mathbf{M}_{\mathbf{Z}}[\varepsilon + \mathbf{X}(\beta - \hat{\beta})] \\ &\xrightarrow{p} (1 - \alpha)\sigma^2 + O(\alpha^i),\end{aligned}$$

and

$$\hat{\lambda}_{\hat{\beta}} \equiv \frac{\hat{\varepsilon}'\mathbf{P}_{\mathbf{Z}}\hat{\varepsilon}}{\hat{\varepsilon}'\mathbf{M}_{\mathbf{Z}}\hat{\varepsilon}} \xrightarrow{p} \frac{\alpha\sigma^2 + O(\alpha^{i+1})}{(1 - \alpha)\sigma^2 + O(\alpha^i)} \equiv \rho_*.$$

Using (5) we find

$$n^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{Z}(\hat{\lambda}_{\hat{\beta}})}\varepsilon \xrightarrow{p} \omega \frac{(\lambda - \rho_*)^2}{\lambda + \rho_*^2} = \omega O(\alpha^{2i+1}).$$

So, since $\lambda(1 - \alpha) = \alpha$, $\hat{\beta}(\hat{\lambda}_{\hat{\beta}}) \xrightarrow{p} \beta + O(\alpha^{2i+1})$. In particular, because of (7), $\hat{\beta}_{\text{CIVE}} \xrightarrow{p} \beta + O(\alpha^3)$. We could iterate as well. However, the inconsistency of the CIVE is $O(\alpha^3)$ only, which we can safely neglect in most if not all applications. Then its variance is estimated consistently by

$$\widehat{\text{Var}}(\hat{\beta}_{\text{CIVE}}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{P}_{\mathbf{Z}(\hat{\lambda}_{2\text{SLS}})}\mathbf{X})^{-1}.$$

This suggests, in a setting with weak and many instruments, to replace the instruments $\mathbf{Z}$ by the concentrated instruments $\mathbf{Z}(\hat{\lambda}_{2\text{SLS}})$ and then proceed in the usual way with 2SLS. This produces (almost) the LIML estimator and the standard errors are many-instruments consistent.

5. Finite-sample comparisons

We end with a small simulation study to compare 2SLS, LIML and CIVE. The data are generated by

$$y_i = \alpha_1 + x_i\beta + \varepsilon_i,$$

$$x_i = \alpha_2 + z_i\pi + \mathbf{z}_{2i}'\pi_2 + u_i + \varepsilon_{i\omega},$$

$i = 1, \dots, 500$. We set $\alpha_1, \alpha_2, \beta$ and the $k-2$ elements of π_2 equal to zero; the ε_i, u_i, z_i and the elements of $\mathbf{z}_{2i}$ are taken i.i.d. $N(0, 1)$. The parameters of interest are π (instrument strength), ω (degree of endogeneity), and k (number of instruments). For chosen values of ω and k , π is based on a chosen F^* value of the regression of x on the instruments. The F -statistic of this regression is

$$F = \frac{\mathbf{x}'(\mathbf{P}_{\mathbf{Z}} - \mathbf{P}_{\mathbf{I}})\mathbf{x}/(k-1)}{\mathbf{x}'\mathbf{M}_{\mathbf{Z}}\mathbf{x}/(n-k)}.$$

We replace the quadratic terms by their expectation,

$$E[\mathbf{x}'(\mathbf{P}_{\mathbf{Z}} - \mathbf{P}_{\mathbf{I}})\mathbf{x}] = (n-1)\pi^2 + (k-1)(1 + \omega^2)$$

$$E[\mathbf{x}'\mathbf{M}_{\mathbf{Z}}\mathbf{x}] = (n-k)(1 + \omega^2),$$

Table 1

Median bias, range (the distance between the 5th and 95th percentile), and rejection rate ($\alpha = 5\%$) for 2SLS, CIVE and LIML. All figures $\times 1000$.

F^*		$\omega = 1/2$			$\omega = 2$		
		2SLS	CIVE	LIML	2SLS	CIVE	LIML
$k = 3$							
10	Median bias	25	4	3	21	1	−0
	Range	728	784	798	358	392	330
	Rejection rate	51	42	42	96	68	67
20	Median bias	11	0	0	10	−0	−0
	Range	482	500	501	230	240	240
	Rejection rate	49	44	44	70	55	55
30	Median bias	6	−0	−0	7	−0	−0
	Range	377	385	385	205	213	213
	Rejection rate	48	45	45	64	51	51
$k = 10$							
5	Median bias	75	1	−0	75	7	0
	Range	440	563	575	186	238	272
	Rejection rate	106	42	43	307	59	56
10	Median bias	34	−2	−2	45	1	0
	Range	306	345	345	157	190	194
	Rejection rate	74	45	46	205	53	53
20	Median bias	19	1	1	16	0	0
	Range	221	235	235	101	108	108
	Rejection rate	66	50	50	104	49	49
$k = 30$							
3	Median bias	129	0	−1	132	32	0
	Range	300	455	477	114	152	214
	Rejection rate	314	41	45	893	78	52
4	Median bias	83	0	0	78	3	−0
	Range	244	320	322	98	133	143
	Rejection rate	220	45	45	658	47	49
10	Median bias	37	−1	−1	39	0	−0
	Range	168	189	189	78	93	93
	Rejection rate	127	49	49	382	48	49

consider the result as the target value F^* , and solve for π to obtain

$$\pi^2 = \frac{k-1}{n-1}(1+\omega^2)(F^*-1).$$

We let $k = 3, 10, 30$ and $\omega = 1/2, 2$ and use values of F^* equal to $F^* = 10, 20, 30$ when $k = 3$, $F^* = 5, 10, 20$ when $k = 10$, and $F^* = 3, 4, 10$ when $k = 30$. We use $N = 50,000$ replications. The table reports the median bias, the range (the distance between the 5th and 95th percentile), and the rejection rate ($\alpha = 5\%$) for 2SLS, CIVE and LIML. The median F statistics were close to the target values and are not reported (see Table 1).

The table shows excellent performance of LIML across the board in terms of median bias and rejection rate, closely followed by CIVE, even with many ($k = 30$) and weak ($F^* = 3$) instruments.

Acknowledgments

We would like to thank Artūras Juodis, Erik Meijer, Laura Spierdijk and Jelle van Essen and an anonymous referee for their comments and suggestions.

References

- Anderson, T.W., 1976. Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 38, 1–36.
- Anderson, T.W., Rubin, H., 1949. Estimator of the parameters of a single equation in a complete system of stochastic equations. *Ann. Math. Statist.* 20, 46–63.
- Anderson, T.W., Rubin, H., 1950. The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations. *Ann. Math. Statist.* 21, 570–582.
- Bekker, P.A., 1994. Alternative approximations to the distributions of instrumental variable estimators. *Econometrica* 62, 657–681.
- Bekker, P.A., Cruje, F., 2015. Jackknife instrumental variable estimation with heteroskedasticity. *J. Econometrics* 185, 332–342.
- Chao, J.C., Swanson, N.R., Hausman, J.A., Newey, T., Woutersen, W.K., 2012. Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments. *Econometric Theory* 28, 42–86.
- Hausman, J.A., Newey, W.K., Woutersen, T., Chao, J., Swanson, N.R., 2012. Instrumental variable estimation with heteroskedasticity and many instruments. *Quant. Econ.* 3, 211–255.
- Nagar, A.L., 1959. The bias and moment matrix of the general k -class estimators of the parameters in simultaneous equations. *Econometrica* 27, 575–595.
- Newey, W.K., 2004. Many instrument asymptotics, Working paper, MIT.
- Theil, H., 1953. Estimation and simultaneous correlation in complete equation systems, Central Planning Bureau, The Hague.