Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: https://www.tandfonline.com/loi/lecr20

ML and GMM with concentrated instruments in the static panel data model

Paul Bekker & Jelle van Essen

To cite this article: Paul Bekker & Jelle van Essen (2020) ML and GMM with concentrated instruments in the static panel data model, Econometric Reviews, 39:2, 181-195, DOI: 10.1080/07474938.2019.1580946

To link to this article: https://doi.org/10.1080/07474938.2019.1580946

Published online: 23 Mar 2019.

ECONOMETRIC REVIEWS
2020, VOL. 39, NO. 2, 181–195
https://doi.org/10.1080/07474938.2019.1580946

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

ML and GMM with concentrated instruments in the static panel data model

Paul Bekker and Jelle van Essen

Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands

ABSTRACT
We study the asymptotic behavior of instrumental variable estimators in the static panel model under many-instruments asymptotics. We provide new estimators and standard errors based on concentrated instruments as alternatives to an estimator based on maximum likelihood. We prove that the latter estimator is consistent under many-instruments asymptotics only if the starting value in an iterative procedure is root-N consistent. A similar approach for continuous updating GMM shows the derivation is nontrivial. For the standard cross-sectional case ($T = 1$), the simple formulation of standard errors offer an alternative to earlier formulations.

KEYWORDS
Bekker standard errors; LIML; many-instruments asymptotics; panel data; weak instruments

JEL CLASSIFICATION
C23; C26

1. Introduction

In the standard linear cross-sectional model with endogenous regressors, the Limited Information Maximum Likelihood (LIML) estimator of Anderson and Rubin (1949) is known to be consistent under many-instruments asymptotics where the number of instruments increases with the sample size. Bekker (1994) formulated many-instruments consistent standard errors resulting in more accurate cover rates of confidence sets in case of many or weak instruments. Recently, Bekker and Wansbeek (2016) provided a simple formulation of many-instruments consistent standard errors based on so-called concentrated instruments. As 2SLS is inconsistent under many-instruments asymptotics, in particular when instruments are weak, it would be interesting to look for LIML-like alternatives for 2SLS in a wider context.[1]

In panel data models the Generalized Method of Moments (GMM) approach of Arellano and Bond (1991), or more recently, the 2SLS approach of Arellano (2016), the number of instruments increases with the time dimension, resulting in many-instruments inconsistent estimators that suffer from bias (e.g., Bun and Sarafidis, 2015; Kiviet, 1995; Ziliak, 1997). Wansbeek and Prak (2017) observe that LIML estimators in panel data models, as developed by, e.g., Alvarez and Arellano (2003), Akashi and Kunitomo (2012) and Moral-Benito (2013) for the dynamic panel data model, are least-variance ratio estimators just as LIML, but not true ML estimators obtained from maximizing a likelihood function. They aim at filling this gap by deriving the ML estimator for the static linear panel data model and investigating its properties in a framework of many-instruments asymptotics.

Here we compare the approach of Wansbeek and Prak (2017), henceforth WP, with continuous updating GMM and a new approach, which uses concentrated instruments. We find that the

CONTACT Paul Bekker p.a.bekker@rug.nl; Jelle van Essen j.van.essen@rug.nl Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands.

[1]For example, Hausman et al. (2012) and Bekker and Crudu (2015) provide LIML-like many-instruments consistent estimators for the cross-sectional linear model with unknown heteroskedasticity.

ML-based estimator of WP is actually many-instruments inconsistent due to an iterative procedure that starts with an inconsistent estimator. We prove that starting with a root-N consistent estimator indeed produces the claimed asymptotic distribution. However, the result is nontrivial and a similar procedure for continuous updating GMM produces another result. Starting with root-N consistent estimators, we distinguish between one-step estimators, using only one iteration step, and fully iterated estimators. In the ML approach the two estimators have the same asymptotic distribution, but for the continuous updating GMM approach the distributions are different.

The new approach can be interpreted as 3SLS based on M concentrated instruments, where M is the number of regressors. The one-step estimator that we call panel concentrated instrumental variable estimator (P-CIVE) has the same asymptotic distribution as the ML-based estimator and the fully iterated continuous updating GMM estimator. Moreover, the estimator has an appealing form and its simple 3SLS-like standard errors are many-instruments consistent. In particular for the cross-sectional case, where $T = 1$, the standard errors offer a simple alternative to the original formulation in Bekker (1994) and the more recent formulation in Bekker and Wansbeek (2016).

The section structure of this paper is as follows. Section 2 introduces the model with multiple regressors and the approach based on concentrated instruments resulting in P-CIVE and its standard errors. To keep the presentation simple, we provide derivations for the case of a single regressor. We discuss the WP approach and its issues in Section 3. In Section 4, we present the continuous updating GMM estimator as an alternative. We find that it can be improved upon by the P-CIVE estimator, which uses concentrated instruments as discussed in Section 5. The differences between the fully iterated estimators is discussed in Section 6. Section 7 gives the outcomes of Monte Carlo simulations in which the performances of the original and corrected WP estimators and the P-CIVE estimator are assessed and compared.

2. The panel model and the P-CIVE estimator

We first discuss the static panel model as considered by Wansbeek and Prak (2017) with multiple time periods $T \geqslant 1$ and multiple regressors $M \geqslant 1$. We present the P-CIVE, which is a 3SLS estimator with 3SLS standard errors that are both consistent under many-instruments asymptotics. The case $T = 1$, where LIML is the ML estimator, will be discussed separately. The case $M = 1$ will be considered to present the derivations of the many-instruments asymptotic distributions.

2.1. The static panel model, $T \geqslant 1$ and $M \geqslant 1$

Consider a static panel data setting with N entities that are each observed at times $t = 1, \ldots, T$. For multiple regressors the model is given by

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

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

where β is an M vector of parameters of interest, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T$ are observed N vectors and $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T$ are observed $N \times M$ matrices of regressors, which may be endogenous. It is assumed that any fixed effects have been eliminated by an appropriate data transformation. An $N \times K$ matrix $\boldsymbol{Z}$ of instruments is observed as well, where $\text{rank}(\boldsymbol{Z}) = K$. The disturbances in $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T)$ and $\boldsymbol{V} = (\boldsymbol{V}_1, \ldots, \boldsymbol{V}_T)$ are not observed. Conditional on the instruments $\boldsymbol{Z}$, the N rows of the matrix of disturbances $(\boldsymbol{U}, \boldsymbol{V})$ are assumed to be independently normally distributed with zero mean and covariance matrix Σ. Furthermore, we use many-instruments asymptotic theory, where the number of instruments K increases with the number of observations N. It is assumed that $K/N \to \alpha$ and $\Pi_t'\boldsymbol{Z}'\boldsymbol{Z}\Pi_s/N \to \boldsymbol{S}_{ts}$ and $\sum_{t=1}^{T} \boldsymbol{S}_{tt} = \boldsymbol{S} > 0$ as $N \to \infty$ and T is fixed.

We consider GMM with continuous updating and an estimator based on maximum likelihood, as described by WP. In particular, we propose a simple many-instruments consistent estimator based on concentrated instruments. To introduce the latter approach, let $\boldsymbol{P}_H = \boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'$ and $\boldsymbol{M}_H = \boldsymbol{I} - \boldsymbol{P}_H$ for any matrix $\boldsymbol{H}$ of full column rank. Let $\boldsymbol{C}_Z = \boldsymbol{P}_Z - \lambda(\boldsymbol{I}_N - \boldsymbol{P}_Z)$, where λ is a scalar. As a starting point we use the LIML estimator, which ignores the covariance structure of the disturbances. It is given by

$$\hat{\beta}_{\text{LIML}} = \left\{\sum_{t=1}^{T} \boldsymbol{X}_t' \boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\boldsymbol{X}_t\right\}^{-1} \sum_{t=1}^{T} \boldsymbol{X}_t{}' \boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\boldsymbol{y}_t, \tag{3}$$

where $\hat{\lambda}_{\text{LIML}}$ is the smallest eigenvalue of the matrix.

$$\sum_{t=1}^{T} (\boldsymbol{y}_t, \boldsymbol{X}_t)' \boldsymbol{P}_Z (\boldsymbol{y}_t, \boldsymbol{X}_t) \left\{\sum_{t=1}^{T} (\boldsymbol{y}_t, \boldsymbol{X}_t)'(\boldsymbol{I}_N - \boldsymbol{P}_Z)(\boldsymbol{y}_t, \boldsymbol{X}_t)\right\}^{-1}.$$

Let the LIML residuals be given by $\hat{\boldsymbol{u}}_t = \boldsymbol{y}_t - \boldsymbol{X}_t\hat{\beta}_{\text{LIML}}$, which are collected in the $N \times T$ matrix $\hat{\boldsymbol{U}} = (\hat{\boldsymbol{u}}_1, ..., \hat{\boldsymbol{u}}_T)$. The M concentrated instruments are given by $\bar{\boldsymbol{Z}} = (\bar{\boldsymbol{Z}}_1', ..., \bar{\boldsymbol{Z}}_T')'$, where $\bar{\boldsymbol{Z}}_t = \boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\boldsymbol{M}_{\hat{U}}\boldsymbol{X}_t$. Let $\boldsymbol{y} = (\boldsymbol{y}_1', ..., \boldsymbol{y}_T')'$ and $\boldsymbol{X} = (\boldsymbol{X}_1', ..., \boldsymbol{X}_T')'$, then the P-CIVE is given by

$$\hat{\beta}\boldsymbol{P}\text{–CIVE} = \left[\bar{\boldsymbol{Z}}'\left\{(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}})^{-1} \otimes \boldsymbol{I}_N\right\}\boldsymbol{X}\right]^{-1} \bar{\boldsymbol{Z}}'\left\{(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}})^{-1} \otimes \boldsymbol{I}_N\right\}\boldsymbol{y}. \tag{4}$$

The estimated covariance matrix is given by

$$\hat{\boldsymbol{V}}_{\boldsymbol{P}\text{–CIVE}} = \left[\boldsymbol{X}'\left\{(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}})^{-1} \otimes \boldsymbol{I}_N\right\}\bar{\boldsymbol{Z}}\left[\bar{\boldsymbol{Z}}'\left\{(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}})^{-1} \otimes \boldsymbol{I}_N\right\}\bar{\boldsymbol{Z}}\right]^{-1}\bar{\boldsymbol{Z}}'\left\{(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}})^{-1} \otimes \boldsymbol{I}_N\right\}\boldsymbol{X}\right]^{-1}. \tag{5}$$

We show in Section 5 that the many-instruments asymptotic distribution of the P-CIVE estimator is normal, $N^{1/2}(\hat{\beta}\boldsymbol{P}\text{–CIVE} - \beta) \overset{a}{\sim} \mathcal{N}(0, \boldsymbol{V}_{\boldsymbol{P}\text{–CIVE}})$ and the asymptotic covariance matrix $\boldsymbol{V}_{\boldsymbol{P}\text{–CIVE}}$ can be many-instruments consistently estimated by $\hat{\boldsymbol{V}}_{\boldsymbol{P}\text{–CIVE}}$.

2.2. The classic model, T = 1

If $T = 1$, the model is the classic limited information instrumental variable model, where 2SLS is the classic IV estimator and LIML is the ML estimator. LIML is many instruments consistent, whereas 2SLS is inconsistent a shown, e.g., in Bekker (1994). The estimators are given by

$$\hat{\beta}_{\text{2SLS}} = (\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{y},$$
$$\hat{\beta}_{\text{LIML}} = \{\boldsymbol{X}'\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\boldsymbol{X}\}^{-1}\boldsymbol{X}'\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\boldsymbol{y}.$$

Let $\hat{\boldsymbol{u}}_{\text{2SLS}} = \boldsymbol{y} - \boldsymbol{X}\hat{\beta}_{\text{2SLS}}$, then $\boldsymbol{X}'\boldsymbol{P}_Z\hat{\boldsymbol{u}}_{\text{2SLS}} = 0$. Similarly, let $\hat{\boldsymbol{u}}_{\text{LIML}} = \boldsymbol{y} - \boldsymbol{X}\hat{\beta}_{\text{LIML}}$, then $\boldsymbol{X}'\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\hat{\boldsymbol{u}}_{\text{LIML}} = 0$. However, whereas $\boldsymbol{y}'\boldsymbol{P}_Z\hat{\boldsymbol{u}}_{\text{2SLS}} \neq 0$, we can show $\boldsymbol{y}'\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\hat{\boldsymbol{u}}_{\text{LIML}} = 0$. That is to say,

$$\hat{\lambda}_{\text{LIML}} = \arg\min_b \frac{(\boldsymbol{y} - \text{Xb})'\boldsymbol{P}_Z(\boldsymbol{y} - \text{Xb})}{(\boldsymbol{y} - \text{Xb})'\boldsymbol{M}_Z(\boldsymbol{y} - \text{Xb})}$$

is the smallest value λ such that $(\boldsymbol{y}, \boldsymbol{X})'(\boldsymbol{P}_Z - \lambda\boldsymbol{M}_Z)(\boldsymbol{y}, \boldsymbol{X})$ is singular. Consequently, $(\boldsymbol{y}, \boldsymbol{X})'\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})(\boldsymbol{y}, \boldsymbol{X}) \geqslant 0$ and $(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}_{\text{LIML}})'\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}_{\text{LIML}}) = 0$, which implies $(\boldsymbol{y}, \boldsymbol{X})'\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\hat{\boldsymbol{u}}_{\text{LIML}} = 0$.

Bekker and Wansbeek (2016) used this to reformulate LIML as a 2SLS-like estimator $\hat{\beta}_{\text{LIML}} = (\boldsymbol{X}'\boldsymbol{P}_{\bar{Z}_{\text{BW}}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}_{\bar{Z}_{\text{BW}}}\boldsymbol{y}$, where $\bar{\boldsymbol{Z}}_{\text{BW}} = \boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})(\boldsymbol{y}, \boldsymbol{X})$ are $M+1$ concentrated instruments. They showed that the 2SLS-like standard errors are many-instruments consistent. That is, let $\hat{\sigma}_u^2 = \hat{\boldsymbol{u}}_{\text{LIML}}'\hat{\boldsymbol{u}}_{\text{LIML}}/N$, then $\hat{V}_{\text{LIML}}^{\text{BW}} = \hat{\sigma}_u^2 N(\boldsymbol{X}'\boldsymbol{P}_{\bar{\boldsymbol{Z}}_{\text{BW}}}\boldsymbol{X})^{-1} \overset{p}{\to} \boldsymbol{V}_{\text{LIML}}$, where $N^{1/2}(\hat{\beta}_{\text{LIML}} - \beta) \overset{a}{\sim} \mathcal{N}(0, \boldsymbol{V}_{\text{LIML}})$

under many-instrument asymptotics. Consequently, these standard errors may serve as a simple alternative to the original standard errors in Bekker (1994).

Here we go one step further and formulate LIML as a 2SLS-like estimator based on M concentrated instruments. The first-order equations for the optimization to find LIML can be formulated as $\boldsymbol{X}'\boldsymbol{M}_{\hat{\boldsymbol{u}}}\boldsymbol{P}_Z\hat{\boldsymbol{u}} = 0$, or equivalently, $\boldsymbol{X}'\boldsymbol{M}_{\hat{\boldsymbol{u}}}\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\hat{\boldsymbol{u}} = 0$, where $\hat{\boldsymbol{u}} = \boldsymbol{y} - \boldsymbol{X}\hat{\beta}_{\text{LIML}}$. This suggests to use M concentrated instruments $\bar{\boldsymbol{Z}} = \boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\boldsymbol{M}_{\hat{\boldsymbol{u}}}\boldsymbol{X}$. Again LIML can be formulated as a 2SLS-like estimator $\hat{\beta}_{\text{LIML}} = (\bar{\boldsymbol{Z}}'\boldsymbol{X})^{-1}\bar{\boldsymbol{Z}}'\boldsymbol{y}$ since $\bar{\boldsymbol{Z}}'\hat{\boldsymbol{u}}_{\text{LIML}} = 0$. We find $\hat{\beta}_{\text{LIML}} = \hat{\beta}_{\boldsymbol{P}-\text{CIVE}}$ and as a result of the derivations in Section 5 and Appendix A.4, $\hat{\boldsymbol{V}}_{\text{LIML}} = \hat{\sigma}_u^2 N(\boldsymbol{X}'\boldsymbol{P}_{\bar{Z}}\boldsymbol{X})^{-1}$ is a many-instruments consistent estimator of $\boldsymbol{V}_{\text{LIML}}$ as well. Appendix A.1 derives $\hat{\boldsymbol{V}}^{\text{BW}}_{\text{LIML}} \leqslant \hat{\boldsymbol{V}}_{\text{LIML}}$. A separate study may show how the various standard errors behave in finite samples.

2.3. The model when M = 1

WP present the model for $M = 1$. They find that their derivation of the maximum-likelihood-based estimator is hardly affected when there are multiple regressors and that the generalization carries on to the many-instruments consistent standard errors. Similarly, this holds for the P-CIVE estimator with multiple regressors (4) and its standard errors based on (5). In order to keep the presentation simple, we follow WP and consider the case of a single regressor.

To emphasize that $\boldsymbol{X}_t$ and Π_t are vectors when $M = 1$, we write $\boldsymbol{x}_t = \boldsymbol{X}_t, \pi_t = \Pi_t$ and $\boldsymbol{x} = \boldsymbol{X}$. Let $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$ and $\bar{\boldsymbol{X}} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$, so that $\boldsymbol{y} = \text{vec}(\boldsymbol{Y})$ and $\boldsymbol{x} = \text{vec}(\bar{\boldsymbol{X}})$. The model Eqs. (1) and (2) can thus be written as $\boldsymbol{Y} = \bar{\boldsymbol{X}}\beta + \boldsymbol{U}$, where β is a scalar, and $\bar{\boldsymbol{X}} = \boldsymbol{Z}\Pi + \boldsymbol{V}$, where $\Pi = (\pi_1, \ldots, \pi_T)$. Similar to WP, we use many-instruments asymptotics and let $\Pi'\boldsymbol{Z}'\boldsymbol{Z}\Pi/N \to \boldsymbol{Q} \geqslant 0$, where $S = \text{tr}(\boldsymbol{Q}) > 0$.

First, we discuss the ML-based estimator of WP in Section 3 and continuous updating GMM estimation in Section 4. In Section 5 we discuss the derivations for P-CIVE.

3. The panel LIML estimator of Wansbeek and Prak

WP introduce the panel LIML estimator, which is the maximum likelihood estimator of β under normality. Using $\boldsymbol{U} = \boldsymbol{Y} - \bar{\boldsymbol{X}}\beta$, it satisfies

$$\hat{\beta}_{\text{ML}} = \arg\min_{\beta} \frac{|\boldsymbol{U}'\boldsymbol{U}|}{|\boldsymbol{U}'\boldsymbol{M}_Z\boldsymbol{U}|}, \tag{6}$$

$$\in \arg\operatorname*{solve}_{\beta} \left[\text{tr}\left\{(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\bar{\boldsymbol{X}} - (\boldsymbol{U}'\boldsymbol{M}_Z\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{M}_Z\bar{\boldsymbol{X}}\right\} = 0\right], \tag{7}$$

where "|.|" indicates the determinant. They show the ratio in the right-hand side of (6) converges in probability to a function of β with a unique minimum in the true value, leading to the many-instruments consistency of $\hat{\beta}_{\text{ML}}$.

In order to find the asymptotic variance, WP consider the infeasible estimator

$$\begin{aligned} \tilde{\beta} &= \text{tr}\{\boldsymbol{A}(\boldsymbol{U})\}/\text{tr}\{\boldsymbol{B}(\boldsymbol{U})\}, \\ \boldsymbol{A}(\boldsymbol{U}) &= (\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{Y}'\bar{\boldsymbol{X}} - (\boldsymbol{U}'\boldsymbol{M}_Z\boldsymbol{U})^{-1}\boldsymbol{Y}'\boldsymbol{M}_Z\bar{\boldsymbol{X}}, \\ \boldsymbol{B}(\boldsymbol{U}) &= (\boldsymbol{U}'\boldsymbol{U})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}} - (\boldsymbol{U}'\boldsymbol{M}_Z\boldsymbol{U})^{-1}\bar{\boldsymbol{X}}'\boldsymbol{M}_Z\bar{\boldsymbol{X}}, \end{aligned} \tag{8}$$

and claim that $\hat{\beta}_{\text{ML}}$ and $\tilde{\beta}$ have the same asymptotic variance.[2] Using Σ_{uu}, Σ_{vv} and Σ_{vu} for submatrices of Σ, with $\Sigma_{vv|u} = \Sigma_{vv} - \Sigma_{vu}\Sigma_{uu}^{-1}\Sigma_{uv}$, they find the asymptotic variance of $N^{1/2}(\tilde{\beta} - \beta)$ equals

[2]The result is true as is shown in Appendix A2. The result and the proof are nontrivial. The same approach applied to continuous updating GMM has a different outcome, as shown in Section 4.

$$\nu_{\mathrm{ML}} = \frac{\mathrm{tr}\left\{\Sigma_{uu}^{-1}\left(\boldsymbol{Q} + \lambda\Sigma_{vv|u}\right)\right\}}{\mathrm{tr}^2\left(\Sigma_{uu}^{-1}\boldsymbol{Q}\right)}, \tag{9}$$

where $\lambda = \alpha/(1-\alpha)$.

As a feasible estimator WP use $\hat{\beta} = \mathrm{tr}\{\boldsymbol{A}(\tilde{\boldsymbol{U}})\}/\mathrm{tr}\{\boldsymbol{B}(\tilde{\boldsymbol{U}})\}$, where $\tilde{\boldsymbol{U}} = \boldsymbol{Y}-\bar{\boldsymbol{X}}\hat{\beta}$, which is solved iteratively by using $\hat{\beta}_{2\mathrm{SLS}} = \left(\sum_{t=1}^{T}\boldsymbol{x}_t{'}\boldsymbol{P}_Z\boldsymbol{x}_t\right)^{-1}\sum_{t=1}^{T}\boldsymbol{x}_t{'}\boldsymbol{P}_Z\boldsymbol{y}_t$ as a starting value. WP claim that the resulting estimator, $\hat{\beta}_{\wp}$, has the same asymptotic variance as $\tilde{\beta}$, or $\hat{\beta}_{\mathrm{ML}}$. However, this claim is incorrect since $\hat{\beta}_{\mathrm{ML}}$ is not a unique solution to the first-order condition (7).[3] As a result, $\hat{\beta}_{\wp}$ is many-instruments inconsistent, due to the many-instruments inconsistency of $\hat{\beta}_{2\mathrm{SLS}}$.

As an alternative, we suggest to use the LIML estimator defined in (3) as a starting value, where $M=1$ and $\boldsymbol{X}_t = \boldsymbol{x}_t$. It can be formulated as an extremum estimator, $\hat{\beta}_{\mathrm{LIML}} = \arg\min_b \lambda(b)$, where

$$\lambda(b) = \frac{\sum_{t=1}^{T}\left(\boldsymbol{y}_t - \boldsymbol{x}_t b\right)'\boldsymbol{P}_Z\left(\boldsymbol{y}_t-\boldsymbol{x}_t b\right)}{\sum_{t=1}^{T}\left(\boldsymbol{y}_t - \boldsymbol{x}_t b\right)'(\boldsymbol{I}_N - \boldsymbol{P}_Z)\left(\boldsymbol{y}_t - \boldsymbol{x}_t b\right)}.$$

As $\boldsymbol{y}_t-\boldsymbol{x}_t b = \boldsymbol{Z}\pi_t(\beta-b) + \boldsymbol{u}_t + \boldsymbol{v}_t(\beta-b)$, and using $\lambda = \alpha/(1-\alpha)$, we find similar to the steps taken by WP to prove the many-instruments consistency of the maximum likelihood estimator (6) that

$$\lambda(b) = \lambda + \frac{(b-\beta)^2(1+\lambda)\mathrm{tr}(\boldsymbol{Q})}{\mathrm{tr}(\Sigma_{uu}) + 2(\beta-b)\mathrm{tr}(\Sigma_{uv}) + (\beta-b)^2\mathrm{tr}(\Sigma_{vv})} + o_p(1).$$

Consequently, $\hat{\beta}_{\mathrm{LIML}}$ is a many-instruments consistent root. In particular, we find it to be many-instruments root-N consistent: $\hat{\beta}_{\mathrm{LIML}} = \beta + O_p(N^{-1/2})$. As $\hat{\lambda}_{\mathrm{LIML}} = \lambda(\hat{\beta}_{\mathrm{LIML}})$, we also have $\hat{\lambda}_{\mathrm{LIML}} = \lambda + O_p(N^{-1})$.

If we use $\hat{\beta}_{\mathrm{LIML}}$ as a starting value, then the iterative procedure would produce $\hat{\beta}_{\mathrm{ML}}$, at least if the sample size is not too small. In Appendix A.2 we prove the following result. Let the initial estimator $\hat{\beta}_0$ be root-N consistent and let the one-step estimator be given by $\hat{\beta}_1 = \mathrm{tr}\{\boldsymbol{A}(\hat{\boldsymbol{U}}_0)\}/\mathrm{tr}\{\boldsymbol{B}(\hat{\boldsymbol{U}}_0)\}$, where $\hat{\boldsymbol{U}}_0 = \boldsymbol{Y}-\bar{\boldsymbol{X}}\hat{\beta}_0$, then $N^{1/2}(\hat{\beta}_1-\beta)\overset{a}{\sim}\mathcal{N}(0,\nu_{\mathrm{ML}})$, where ν_{ML} is given by (9).

We find the asymptotic distribution is not affected by the choice of $\hat{\beta}_0$ as long as $(\hat{\beta}_0-\beta)^2 = o_p(N^{-1/2})$. In particular, for $\hat{\beta}_0 = \hat{\beta}_{\mathrm{ML}}$, we find the asymptotic distribution of the ML estimator. For $\hat{\beta}_0 = \beta$ we find the asymptotic distribution of the infeasible $\tilde{\beta}$ estimator. For $\hat{\beta}_0 = \hat{\beta}_{\mathrm{LIML}}$ we find the asymptotic distribution of the feasible one-step estimator $\hat{\beta}_{\mathrm{ML},1}$, where further iterations do not change the many-instruments asymptotic distribution.

4. Continuous updating GMM estimation

Consider the case of a single regressor. Let $\hat{\Sigma} = \hat{\Sigma}(\beta,\Pi) = N^{-1}(\boldsymbol{U},\boldsymbol{V})'(\boldsymbol{U},\boldsymbol{V})$. The GMM objective function is given by

$$\begin{aligned} Q_{\mathrm{GMM}}(\beta,\Pi) &= \mathrm{vec}'\{(\boldsymbol{U},\boldsymbol{V})\}\left(\hat{\Sigma}^{-1}\otimes\boldsymbol{P}_Z\right)\mathrm{vec}\{(\boldsymbol{U},\boldsymbol{V})\} \\ &= \mathrm{tr}\left\{\hat{\Sigma}^{-1}(\boldsymbol{U},\boldsymbol{V})'\boldsymbol{P}_Z(\boldsymbol{U},\boldsymbol{V})\right\}. \end{aligned}$$

[3]For example, consider the case $T=1$, where $\hat{\beta}_{\mathrm{ML}}$ is the standard LIML estimator, which is the solution to the first-order condition that produces the smallest eigenvalue $\hat{\boldsymbol{u}}_1\boldsymbol{P}_Z\hat{\boldsymbol{u}}_1/\hat{\boldsymbol{u}}_1(\boldsymbol{I}_n-\boldsymbol{P}_Z)\hat{\boldsymbol{u}}_1$ of the matrix $(\boldsymbol{y}_1,\boldsymbol{x}_1)'\boldsymbol{P}_Z(\boldsymbol{y}_1,\boldsymbol{x}_1)\{(\boldsymbol{y}_1,\boldsymbol{x}_1)'(\boldsymbol{I}_N-\boldsymbol{P}_Z)(\boldsymbol{y}_1,\boldsymbol{x}_1)\}^{-1}$. Another solution is found for the largest eigenvalue. The outcome of an iterative procedure would depend on the starting value and it need not converge to $\hat{\beta}_{\mathrm{ML}}$.

In order to find the continuously updated GMM (CUGMM) estimator, the objective function is minimized while $\boldsymbol{U}, \boldsymbol{V}$ and $\hat{\Sigma}$ depend on β and Π. Appendix A.3 shows that given a value for β, the objective function $Q_{\text{GMM}}(\beta, \Pi)$ is minimized by $\hat{\Pi}(\beta) = (\boldsymbol{Z}'\boldsymbol{M}_U\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{M}_U\bar{\boldsymbol{X}}$, which is the same matrix function as found by WP for maximizing the normal likelihood. The CUGMM estimator $\hat{\beta}_{\text{CUGMM}}$ is given by

$$\hat{\beta}_{\text{CUGMM}} = \arg\min_{\beta} \text{tr}\left(\hat{\Sigma}_{uu}^{-1}\boldsymbol{U}'\boldsymbol{P}_Z\boldsymbol{U}\right), \tag{10}$$

$$\in \arg\operatorname*{solve}_{\beta}\left[\text{tr}\left\{(\boldsymbol{U}'\boldsymbol{U})^{-1}\bar{\boldsymbol{X}}'\boldsymbol{M}_U\boldsymbol{P}_Z\boldsymbol{U}\right\} = 0\right]. \tag{11}$$

Notice that for $T=1$, where $\boldsymbol{U}'\boldsymbol{U}$ is a scalar, $\hat{\beta}_{\text{ML}} = \hat{\beta}_{\text{CUGMM}}$. However, for $T>1$ the functions are different. Again, similar to (7), the first-order condition (11) may allow for more than one solution.

Let an initial estimator $\hat{\beta}_0$ be root-N consistent and let $\hat{\beta}_1$ be the one-step estimator

$$\hat{\beta}_1 = \frac{\text{tr}\left\{\left(\hat{\boldsymbol{U}}_0'\hat{\boldsymbol{U}}_0\right)^{-1}\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}_0}\boldsymbol{P}_Z\boldsymbol{Y}\right\}}{\text{tr}\left\{\left(\hat{\boldsymbol{U}}_0'\hat{\boldsymbol{U}}_0\right)^{-1}\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}_0}\boldsymbol{P}_Z\bar{\boldsymbol{X}}\right\}}, \tag{12}$$

where $\hat{\boldsymbol{U}}_0 = \boldsymbol{Y} - \bar{\boldsymbol{X}}\hat{\beta}_0$. Appendix A.3 shows, for the many-instruments asymptotic sequence, that

$$\begin{aligned} N^{1/2}\left(\hat{\beta}_1 - \beta\right) &= N^{-1/2}\frac{\text{tr}\left\{\Sigma_{uu}^{-1}(\boldsymbol{Z}\Pi + \boldsymbol{V}^{\perp})'(\boldsymbol{P}_Z - \alpha\boldsymbol{I}_N)\boldsymbol{U}\right\}}{\text{tr}\left\{\Sigma_{uu}^{-1}(\boldsymbol{Q} + \alpha\Sigma_{vv|u})\right\}} \\ &\quad + \alpha N^{1/2}\left(\hat{\beta}_0 - \beta\right)\frac{\text{tr}\left\{\Sigma_{uu}^{-1}(\boldsymbol{Q} + \Sigma_{vv|u})\right\}}{\text{tr}\left\{\Sigma_{uu}^{-1}(\boldsymbol{Q} + \alpha\Sigma_{vv|u})\right\}} + o_p(1), \end{aligned} \tag{13}$$

where $\boldsymbol{V}^{\perp} = \boldsymbol{V} - \boldsymbol{U}\Sigma_{uu}^{-1}\Sigma_{uv}$. We find, different from maximum likelihood, that the asymptotic distribution of the one-step CUGMM estimator depends on the initial estimator. As

$$\begin{aligned} N^{-1/2}\text{tr}\left\{\Sigma_{uu}^{-1}(\boldsymbol{Z}\Pi + \boldsymbol{V}^{\perp})'(\boldsymbol{P}_Z - \alpha\boldsymbol{I}_N)\boldsymbol{U}\right\} &= N^{-1/2}\text{vec}(\boldsymbol{Z}\Pi + \boldsymbol{V}^{\perp})'\left\{\Sigma_{uu}^{-1} \otimes (\boldsymbol{P}_Z - \alpha\boldsymbol{I}_N)\right\}\text{vec}(\boldsymbol{U}) \\ &\overset{a}{\sim} \mathcal{N}\left(0, (1-\alpha^2)\text{tr}\left\{\Sigma_{uu}^{-1}(\boldsymbol{Q} + \lambda\Sigma_{vv|u})\right\}\right), \end{aligned}$$

we find in particular, when $\hat{\beta}_0 = \hat{\beta}_1 = \hat{\beta}_{\text{CUGMM}}$, that

$$N^{1/2}\left(\hat{\beta}_{\text{CUGMM}} - \beta\right) = \frac{\text{tr}\left\{\Sigma_{uu}^{-1}(\boldsymbol{Z}\Pi + \boldsymbol{V}^{\perp})'(\boldsymbol{P}_Z - \alpha\boldsymbol{I}_N)\boldsymbol{U}\right\}}{(1-\alpha)\text{tr}(\Sigma_{uu}^{-1}\boldsymbol{Q})} + o_p(1) \overset{a}{\sim} \mathcal{N}(0, \nu_{\text{ML}}),$$

as in (9). This shows that the fully iterated estimator, $\hat{\beta}_{\text{CUGMM}}$, has the same many-instruments asymptotic distribution as the one-step ML estimator. However, when $\hat{\beta}_0 = \beta$, the infeasible estimator $\hat{\beta}_1$ has a different asymptotic distribution. It shows that the claim of WP that $\tilde{\beta}$ as defined in (8) and $\hat{\beta}_{\text{ML}}$ have the same asymptotic distribution is nontrivial. A similar claim about CUGMM would be wrong.

5. P-CIVE

Consider a single regressor. Let $\tilde{\boldsymbol{z}} = \text{vec}\{\boldsymbol{C}_Z(\lambda)(\boldsymbol{Z}\Pi + \boldsymbol{V}^{\perp})\}$ be a single infeasible instrument. As $\tilde{\boldsymbol{z}}$ is independent of $\boldsymbol{u}$, we find $N^{-1/2}\tilde{\boldsymbol{z}}'(\Sigma_{uu}^{-1} \otimes \boldsymbol{I}_N)\boldsymbol{u} \overset{a}{\sim} \mathcal{N}(0, \nu)$, where $\boldsymbol{u} = \text{vec}(\boldsymbol{U})$ and

$$\nu = \operatorname*{plim}_{n\to\infty}\left(\frac{\tilde{\boldsymbol{z}}'\text{Var}^{-1}(\boldsymbol{u})\tilde{\boldsymbol{z}}}{N}\right) = \operatorname*{plim}_{n\to\infty}\left(\frac{\text{tr}\left\{\Sigma_{uu}^{-1}(\boldsymbol{Z}\Pi + \boldsymbol{V}^{\perp})'\boldsymbol{C}_Z^2(\lambda)(\boldsymbol{Z}\Pi + \boldsymbol{V}^{\perp})\right\}}{N}\right)$$
$$= \text{tr}\left\{\Sigma_{uu}^{-1}\left(\boldsymbol{Q} + \lambda\Sigma_{vv|u}\right)\right\}.$$

Furthermore, $\tilde{\boldsymbol{z}}'(\Sigma_{uu}^{-1} \otimes \boldsymbol{I}_N)\boldsymbol{x}/N = \text{tr}(\Sigma_{uu}^{-1}\boldsymbol{Q}) + o_p(1)$. Consequently, we find that the infeasible estimator $\tilde{\beta} = (\tilde{\boldsymbol{z}}'(\Sigma_{uu}^{-1} \otimes \boldsymbol{I}_N)\boldsymbol{x})^{-1}\tilde{\boldsymbol{z}}'(\Sigma_{uu}^{-1} \otimes \boldsymbol{I}_N)\boldsymbol{y}$ satisfies $N^{1/2}(\tilde{\beta}-\beta) \overset{a}{\sim} \mathcal{N}(0, \nu_{\text{ML}})$, where ν_{ML} is given in (9) and

$$\nu_{\text{ML}} = \operatorname*{plim}_{n\to\infty}\left(\frac{\boldsymbol{x}'(\Sigma_{uu}^{-1} \otimes \boldsymbol{I}_N)\tilde{\boldsymbol{z}}\{\tilde{\boldsymbol{z}}'(\Sigma_{uu}^{-1} \otimes \boldsymbol{I}_N)\tilde{\boldsymbol{z}}\}^{-1}\tilde{\boldsymbol{z}}\prime(\Sigma_{uu}^{-1} \otimes \boldsymbol{I}_N)\boldsymbol{x}}{N}\right)^{-1}.$$

It has the same asymptotic distribution as the one-step ML estimator or the fully iterated CUGMM estimator.

To make this approach feasible, consider the P-CIVE, (4), which for $M=1$ amounts to

$$\hat{\beta}\boldsymbol{P}\text{−CIVE} = \left[\bar{\boldsymbol{z}}'\left\{\left(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}}\right)^{-1} \otimes \boldsymbol{I}_N\right\}\boldsymbol{x}\right]^{-1}\bar{\boldsymbol{z}}'\left\{\left(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}}\right)^{-1} \otimes \boldsymbol{I}_N\right\}\boldsymbol{y},$$
$$\bar{\boldsymbol{z}} = \text{vec}\left\{\boldsymbol{C}_Z(\hat{\lambda})\boldsymbol{M}_{\hat{U}}\bar{\boldsymbol{X}}\right\},$$

where $\hat{\boldsymbol{U}} = \boldsymbol{Y} - \bar{\boldsymbol{X}}\hat{\beta}_{\text{LIML}}$ and $\hat{\lambda} = \hat{\lambda}_{\text{LIML}}$, as described in (3). The difference with the one-step CUGMM estimator is that $\boldsymbol{P}_Z$ in (12) is replaced by $\boldsymbol{C}_Z(\hat{\lambda})$:

$$\hat{\beta}\boldsymbol{P}\text{−CIVE} = \frac{\text{tr}\left\{\left(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}}\right)^{-1}\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}}\boldsymbol{C}_Z(\hat{\lambda})\boldsymbol{Y}\right\}}{\text{tr}\left\{\left(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}}\right)^{-1}\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}}\boldsymbol{C}_Z(\hat{\lambda})\bar{\boldsymbol{X}}\right\}}. \tag{14}$$

In Appendix A.4 we show that $N^{1/2}(\hat{\beta}\boldsymbol{P}\text{−CIVE}-\beta) \overset{a}{\sim} \mathcal{N}(0, \nu_{\boldsymbol{P}\text{−CIVE}})$, where $\nu_{\boldsymbol{P}\text{−CIVE}} = \nu_{\text{ML}}$ as given in (9). It can be consistently estimated by (5), which amounts to

$$\hat{\nu}_{\boldsymbol{P}\text{−CIVE}} = \left[\boldsymbol{x}'\left\{\left(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}}\right)^{-1} \otimes \boldsymbol{I}_N\right\}\bar{\boldsymbol{z}}\left[\bar{\boldsymbol{z}}'\left\{\left(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}}\right)^{-1} \otimes \boldsymbol{I}_N\right\}\bar{\boldsymbol{z}}\right]^{-1}\bar{\boldsymbol{z}}'\left\{\left(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}}\right)^{-1} \otimes \boldsymbol{I}_N\right\}\boldsymbol{x}\right]^{-1}$$
$$= \frac{\text{tr}\left\{\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}}\boldsymbol{C}_Z^2(\hat{\lambda})\boldsymbol{M}_{\hat{U}}\bar{\boldsymbol{X}}\left(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}}\right)^{-1}\right\}}{\text{tr}^2\left\{\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}}\boldsymbol{C}_Z(\hat{\lambda})\bar{\boldsymbol{X}}\left(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}}\right)^{-1}\right\}}. \tag{15}$$

6. The iterated estimators after convergence

We already found that for $T=1$ the three one-step estimators ML, CUGMM and P-CIVE coincide if the starting value is given by LIML. Would this also hold true for the iterated estimators if $T>1$? To answer this question we reconsider the first-order conditions. For ML we reformulate (7), given by

$$\text{tr}\left\{\left(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}}\right)^{-1}\hat{\boldsymbol{U}}'\bar{\boldsymbol{X}} - \left(\hat{\boldsymbol{U}}'\boldsymbol{M}_Z\hat{\boldsymbol{U}}\right)^{-1}\hat{\boldsymbol{U}}'\boldsymbol{M}_Z\bar{\boldsymbol{X}}\right\} = 0,$$

by using

$$\begin{aligned}
&(\hat{U}'\hat{U})^{-1}\hat{U}'\bar{X}-(\hat{U}'M_Z\hat{U})^{-1}\hat{U}'M_Z\bar{X}\\
&\quad=(\hat{U}'\hat{U})^{-1}\left\{\hat{U}'\bar{X}-\hat{U}'P_Z\hat{U}(\hat{U}'M_Z\hat{U})^{-1}\hat{U}'M_Z\bar{X}-\hat{U}'M_Z\bar{X}\right\}\\
&\quad=(\hat{U}'\hat{U})^{-1}\left\{\hat{U}'P_Z\bar{X}-\hat{U}'P_Z\hat{U}(\hat{U}'M_Z\hat{U})^{-1}\hat{U}'M_Z\bar{X}\right\}\\
&\quad=(\hat{U}'\hat{U})^{-1}\left\{\hat{U}'P_ZM_{\hat{U}}\bar{X}-\hat{U}'P_Z\hat{U}(\hat{U}'M_Z\hat{U})^{-1}\hat{U}'M_ZM\bar{X}\right\}\\
&\quad=(\hat{U}'\hat{U})^{-1}\left\{\hat{U}'P_ZM_{\hat{U}}\bar{X}+\hat{U}'P_Z\hat{U}(\hat{U}'M_Z\hat{U})^{-1}\hat{U}'P_ZM_{\hat{U}}\bar{X}\right\}\\
&\quad=(\hat{U}'\hat{U})^{-1}\left\{I_T+\hat{U}'P_Z\hat{U}(\hat{U}'M_Z\hat{U})^{-1}\right\}\hat{U}'P_ZM_{\hat{U}}\bar{X}\\
&\quad=(\hat{U}'M_Z\hat{U})^{-1}\hat{U}'P_ZM_{\hat{U}}\bar{X},
\end{aligned}$$

as

$$\operatorname{tr}\left[(\hat{U}'M_Z\hat{U})^{-1}\hat{U}'P_ZM_{\hat{U}}\bar{X}\right]=0. \tag{16}$$

For CUGMM the first-order condition (11) is

$$\operatorname{tr}\left\{(\hat{U}'\hat{U})^{-1}\hat{U}'P_ZM_{\hat{U}}\bar{X}\right\}=0. \tag{17}$$

Using (14) we find iterated P-CIVE satisfies after convergence

$$\operatorname{tr}\left\{(\hat{U}'\hat{U})^{-1}\hat{U}'\left\{P_Z-\left(\frac{\operatorname{tr}\hat{U}'P_Z\hat{U}}{\operatorname{tr}\hat{U}'M_Z\hat{U}}\right)M_Z\right\}M_{\hat{U}}\bar{X}\right\}=0.$$

As

$$\begin{aligned}
&\hat{U}'\left\{P_Z-\left(\frac{\operatorname{tr}\hat{U}'P_Z\hat{U}}{\operatorname{tr}\hat{U}'M_Z\hat{U}}\right)M_Z\right\}M_{\hat{U}}\bar{X}=\\
&\hat{U}'\left\{P_Z+\left(\frac{\operatorname{tr}\hat{U}'P_Z\hat{U}}{\operatorname{tr}\hat{U}'M_Z\hat{U}}\right)P_Z\right\}M_{\hat{U}}\bar{X}=\left(\frac{\operatorname{tr}\hat{U}'\hat{U}}{\operatorname{tr}\hat{U}'M_Z\hat{U}}\right)\hat{U}'P_ZM_{\hat{U}}\bar{X}.
\end{aligned}$$

iterated P-CIVE satisfies

$$\operatorname{tr}\left\{(\hat{U}'\hat{U})^{-1}\hat{U}'P_ZM_{\hat{U}}\bar{X}\right\}=0,$$

which amounts to the same first-order condition (17) as CUGMM. Therefore, iterated P-CIVE is equivalent to CUGMM. Furthermore, comparing (16) and (17), ML is equivalent to iterated P-CIVE if $\hat{U}'\hat{U}/N$ is a scalar multiple of $\hat{U}'M_Z\hat{U}/N$, which occurs asymptotically, or when $T=1$. In general, when $T>1$, the first-order conditions are not equivalent, so that ML is different from P-CIVE.

7. Simulations

The performance of the WP estimator, the one-step ML estimator and the P-CIVE estimator is assessed by means of Monte Carlo simulations. We consider the simulation setup of WP, which uses the model described in Section 2 with $T=2$, $N=500$ and $\beta=1$. They vary an endogeneity parameter ω, an instrument strength parameter F and the number of instruments K to create different simulation settings. However, their instrument strength parameter F is too low to properly reflect the central location of the F-values as they hold in the simulated samples. The difference is approximately one, so that the parameter value $F=3$ produces a median of the F statistics that on average is close to 4, where the bias is less severe when compared to the case where the median of the F statistics is actually close to 3. Therefore, we follow the approach of Bekker and

Table 1. Results of Monte Carlo simulation.

	$K=10$						$K=30$					
	$\omega=2$			$\omega=0.5$			$\omega=2$			$\omega=0.5$		
	WP	1ML	P-CIVE	WP	1ML	P-CIVE	WP	1ML	P-CIVE	WP	1ML	P-CIVE
						$F^*=2$						
Median bias (×1000)	59.71	−0.06	0.02	30.78	4.59	4.77	22.14	0.20	0.23	1.85	−0.32	−0.03
90% range (×10)	5.88	4.30	4.29	11.00	11.34	11.31	5.57	2.21	2.21	5.61	5.66	5.66
5% rejection rate	0.258*	0.070	0.070	0.051*	0.049	0.049	0.191	0.052	0.052	0.049	0.046	0.046
Median F		1.96			1.96			1.99			1.98	
						$F^*=3$						
Median bias (×1000)	4.22	−0.44	−0.39	0.86	−0.59	−0.36	0.45	0.37	0.38	0.34	0.32	0.41
90% range (×10)	2.55	2.62	2.62	5.94	5.98	5.97	1.46	1.46	1.46	3.31	3.31	3.31
5% rejection rate	0.078*	0.056	0.056	0.043	0.045	0.045	0.052	0.052	0.052	0.048	0.049	0.048
Median F		2.95			2.95			2.99			2.98	
						$F^*=5$						
Median bias (×1000)	0.12	0.11	0.09	0.33	0.29	0.39	0.04	0.04	0.01	−0.17	−0.19	−0.18
90% range (×10)	1.76	1.76	1.76	3.71	3.71	3.71	0.99	0.99	0.99	2.11	2.11	2.11
5% rejection rate	0.051	0.052	0.052	0.045	0.046	0.046	0.051	0.051	0.051	0.048	0.050	0.050
Median F		4.95			4.96			4.98			4.98	
						$F^*=10$						
Median bias (×1000)	−0.06	−0.06	−0.06	0.49	0.47	0.48	0.02	0.02	0.02	0.23	0.23	0.25
90% range (×10)	1.13	1.13	1.13	2.30	2.30	2.30	0.64	0.64	0.64	1.33	1.33	1.33
5% rejection rate	0.049	0.049	0.049	0.047	0.048	0.048	0.050	0.050	0.050	0.050	0.050	0.050
Median F		9.93			9.95			9.98			9.98	

Note: Number of replications: 50,000. Fixed parameters $N=500$, $T=2$, $\beta=1$. Other parameters (F^*, ω, K) are varied as indicated in the table. "WP," "1ML" and "P − CIVE" refer to the Wansbeek–Prak, one-step ML and panel CIVE estimators, respectively. Estimated variances for the WP estimator were negative in some replications. These variances were counted as zero variances, resulting in rejections. Parameter settings for which one or more variances were negative are indicated by a star (*).

Wansbeek (2016), so that the strength parameter F^*, as we use it, is more in agreement with the actual median value of the F statistics.

The elements of the matrices $\boldsymbol{Z}, \boldsymbol{U}$ and an $N \times T$ matrix $\boldsymbol{E}$ are drawn independently from a standard normal distribution. Let $\boldsymbol{V} = \omega \boldsymbol{U} + \boldsymbol{E}$ so that

$$\Sigma = \left(\begin{bmatrix} 1 & \omega \\ \omega & 1+\omega^2 \end{bmatrix} \otimes \boldsymbol{I}_2 \right).$$

The matrix Π is defined as $\Pi = (\pi \imath_2, 0)'$, where $\imath_2$ is a 2×1-vector of ones. The value of the scalar π is determined by F^*. We use $\pi = \sqrt{\frac{K}{N}(\omega^2+1)(F^*-1)}$, which solves

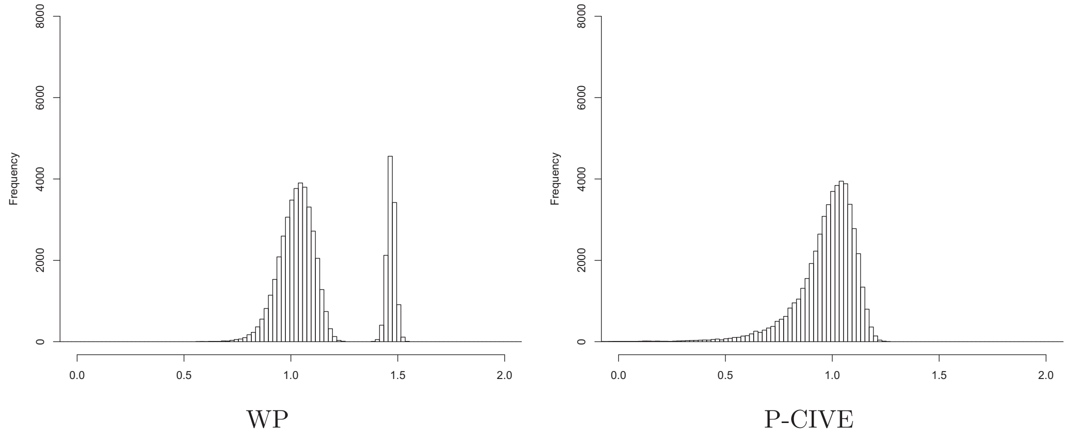
$$F^* = \frac{\boldsymbol{E}\{\boldsymbol{x}'(\boldsymbol{I}_2 \otimes \boldsymbol{P}_Z)\boldsymbol{x}\}/(2K)}{\boldsymbol{E}\left[\boldsymbol{x}'\{\boldsymbol{I}_2 \otimes (\boldsymbol{I}_2 - \boldsymbol{P}_Z)\}\boldsymbol{x}\right]/(2N-2K)} = \frac{N\pi^2 + K(\omega^2+1)}{K(\omega^2+1)}.$$

for π.[4] In order to illustrate how the empirical F-statistics correspond to the specified level F^*, we report the median of the F-statistics for each parameter setting.

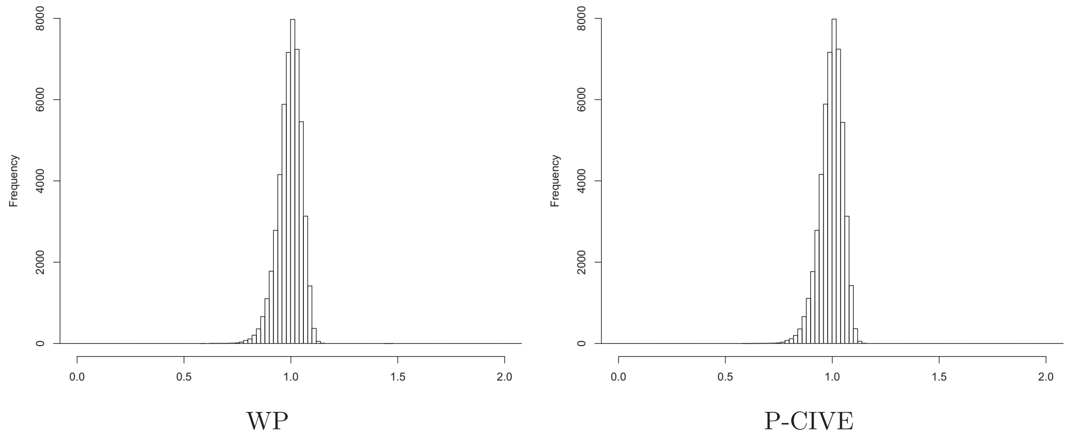
We use 16 specifications found for $\omega \in \{1/2, 2\}, F^* \in \{2, 3, 5, 10\}, K \in \{10, 30\}$.[5] For each setting, we simulate 50,000 replications.

For each estimator, we compute the median bias, the length of the range between the 5th and 95th quantile, and the empirical rejection rate of the hypothesis that $\beta = 1$ with a theoretical significance level of 5%. The latter measure requires computation of standard errors. For the WP and one-step ML estimators, these are based on the formula for the estimated variance of their estimator as described in their paper,

[4]WP use $\pi = \sqrt{\frac{K}{N-K}(1+\omega^2)F}$, where $F = \frac{(N-K)R^2}{K(1-R^2)}$ and $R^2 = \pi^2/(\pi^2+\omega^2+1)$.
[5]WP use $\omega \in \{1/2, 2\}, F \in \{3, 5, 10\}$ and $K \in \{10, 30\}$.



(a) $F^* = 2$



(b) $F^* = 5$

Figure 1. Histograms of estimates of β from Monte Carlo simulation on the interval $[0, 2]$. Number of replications: 50,000. Parameter settings: $N = 500$, $T = 2$, $\beta = 1$, $K = 10$, $\omega = 2$. Parameter F^* is varied as indicated. "WP" and "P-CIVE" refer to the Wansbeek–Prak and panel CIVE estimator, respectively.

$$\hat{\nu}_{\wp} = \frac{\mathrm{tr}\left\{ \left(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}}\right)^{-1} \bar{\boldsymbol{X}}' \left(\boldsymbol{C}_Z^2(\hat{\lambda}) - \hat{\lambda} \boldsymbol{P}_{\hat{U}} \right) \bar{\boldsymbol{X}} \right\}}{\mathrm{tr}^2\left\{ \left(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}}\right)^{-1} \bar{\boldsymbol{X}}' \boldsymbol{C}_Z(\hat{\lambda}) \bar{\boldsymbol{X}} \right\}},$$

where $\hat{\lambda} = K/(N-K)$. This estimator can take on negative values, implying that the standard error cannot be computed. When this happens, these variances are counted as zero variances, resulting in rejections. Parameter settings for which one or more variances were negative are indicated by a star (*). For the P-CIVE estimator, the simple standard errors (15) are used.

Results are given in Table 1. We observe that the one-step ML and P-CIVE estimators perform very similarly. One might conjecture, based on the empirical distributions, that the two estimators are identical and the differences are due to numerical imprecision. However, the observed differences between the estimators in 50,000 replications are not only due to numerical imprecision. The estimators are different, as has been shown in Section 6. They perform well in terms of bias and inference, with a slight degree of overrejection for $\omega = 2$. For higher instrument strength,

$F^* \in \{5, 10\}$, the WP estimator performs similarly to the other two estimators, but it is increasingly biased for lower settings of F^*.

Figure 1 gives insight into the bias of the WP procedure. It contains histograms of the WP and P-CIVE estimators; we have omitted the one-step ML estimator, because its distribution is rather similar to the distribution of P-CIVE. However, all estimators are numerically different. In particular, we have chosen $K = 10$ and $\omega = 2$ as an example, and $F^* = 2, 5$. We observe that the WP estimates follow a bimodal distribution when the instruments are weak. This is visual evidence of the issue discussed in Section 3 about the multiplicity of solutions to the first-order condition (7). When instruments are weak, the 2SLS estimator is median biased. This results in the WP procedure, which uses 2SLS to obtain a starting value, converging to the wrong root with substantial positive probability.

Interestingly, the inclusion of additional instruments improves the performance of the WP estimator in most settings. Although the 2SLS estimator is known to be biased for larger numbers of instruments, the present settings control for F^*. As a result the inconsistency of 2SLS does not increase with K if F^* is fixed. That is to say, WP derive the inconsistency as

$$\text{plim}\left(\hat{\beta}_{2\text{SLS}} - \beta\right) = \frac{\alpha \text{tr}(\Sigma_{vu})}{\text{tr}(\boldsymbol{Q}) + \alpha \text{tr}(\Sigma_{vv})},$$

which, in the present setting amounts to

$$\hat{\beta}_{2\text{SLS}} = \beta + \frac{\frac{K}{N}\omega}{\pi^2 + \frac{K}{N}(1 + \omega^2)} + o_p(1) = \beta + \frac{\omega}{(\omega^2 + 1)F^*} + o_p(1).$$

So, indeed, the inconsistency does not vary with K when F^* is fixed. Apparently, this fixed inconsistency of the starting value of the WP procedure causes the biggest problems when there are few weak instruments.

Appendix

A.1. Standard errors when T = 1

How do the standard errors of Section 2.2 based on $\hat{\boldsymbol{V}}_{\text{LIML}}^{\text{BW}}$ and $\hat{\boldsymbol{V}}_{\text{LIML}}$ compare? To answer this question, first observe $(\boldsymbol{y}, \boldsymbol{X})\boldsymbol{A} = (\hat{\boldsymbol{u}}, \boldsymbol{M}_{\hat{u}}\boldsymbol{X})$, where $\boldsymbol{A}$ is nonsingular,

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0' \\ -\hat{\beta}_{\text{LIML}} & \boldsymbol{I}_M \end{pmatrix} \begin{pmatrix} 1 & -\frac{\hat{\boldsymbol{u}}'\boldsymbol{X}}{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}} \\ 0 & \boldsymbol{I}_M \end{pmatrix}.$$

As $\boldsymbol{X}'\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\hat{\boldsymbol{u}} = 0$, we find

$$\begin{aligned}
\boldsymbol{X}'\boldsymbol{P}_{\bar{\boldsymbol{Z}}_{\text{BW}}}\boldsymbol{X} &= \boldsymbol{X}'\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})(\boldsymbol{y}, \boldsymbol{X})\left\{(\boldsymbol{y}, \boldsymbol{X})'\boldsymbol{C}_Z^2(\hat{\lambda}_{\text{LIML}})(\boldsymbol{y}, \boldsymbol{X})\right\}^{-1}(\boldsymbol{y}, \boldsymbol{X})'\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\boldsymbol{X}, \\
&= \boldsymbol{X}'\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})(\hat{\boldsymbol{u}}, \boldsymbol{M}_{\hat{u}}\boldsymbol{X})\left\{(\hat{\boldsymbol{u}}, \boldsymbol{M}_{\hat{u}}\boldsymbol{X})'\boldsymbol{C}_Z^2(\hat{\lambda}_{\text{LIML}})(\hat{\boldsymbol{u}}, \boldsymbol{M}_{\hat{u}}\boldsymbol{X})\right\}^{-1}(\hat{\boldsymbol{u}}, \boldsymbol{M}_{\hat{u}}\boldsymbol{X})'\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\boldsymbol{X}, \\
&= \boldsymbol{X}'\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\boldsymbol{M}_{\hat{u}}\boldsymbol{X}(0, \boldsymbol{I}_M)\left\{(\hat{\boldsymbol{u}}, \boldsymbol{M}_{\hat{u}}\boldsymbol{X})'\boldsymbol{C}_Z^2(\hat{\lambda}_{\text{LIML}})(\hat{\boldsymbol{u}}, \boldsymbol{M}_{\hat{u}}\boldsymbol{X})\right\}^{-1}(0, \boldsymbol{I}_M)'\boldsymbol{X}'\boldsymbol{M}_{\hat{u}}\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\boldsymbol{X}, \\
&\geqslant \boldsymbol{X}'\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\boldsymbol{M}_{\hat{u}}\boldsymbol{X}\left\{\boldsymbol{X}'\boldsymbol{M}_{\hat{u}}\boldsymbol{C}_Z^2(\hat{\lambda}_{\text{LIML}})\boldsymbol{M}_{\hat{u}}\boldsymbol{X}\right\}^{-1}\boldsymbol{X}'\boldsymbol{M}_{\hat{u}}\boldsymbol{C}_Z(\hat{\lambda}_{\text{LIML}})\boldsymbol{X} \\
&= \boldsymbol{X}'\boldsymbol{P}_{\bar{\boldsymbol{Z}}}\boldsymbol{X}.
\end{aligned}$$

Consequently, $\hat{\boldsymbol{V}}_{\text{LIML}}^{\text{BW}} \leqslant \hat{\boldsymbol{V}}_{\text{LIML}}$.

A.2. The asymptotic distribution of the one-step ML estimator

The one-step estimator $\hat{\beta}_1 = \mathrm{tr}\{\boldsymbol{A}(\hat{\boldsymbol{U}}_0)\}/\mathrm{tr}\{\boldsymbol{B}(\hat{\boldsymbol{U}}_0)\}$, where $\hat{\boldsymbol{U}}_0 = \boldsymbol{Y} - \bar{\boldsymbol{X}}\hat{\beta}_0$ satisfies

$$N^{1/2}(\hat{\beta}_1 - \beta) = N^{1/2}\frac{\mathrm{tr}\{(\hat{\boldsymbol{U}}_0'\hat{\boldsymbol{U}}_0)^{-1}\boldsymbol{U}'\bar{\boldsymbol{X}} - (\hat{\boldsymbol{U}}_0'\boldsymbol{M}_Z\hat{\boldsymbol{U}}_0)^{-1}\boldsymbol{U}\boldsymbol{M}_Z\bar{\boldsymbol{X}}\}}{\mathrm{tr}\{(\hat{\boldsymbol{U}}_0'\hat{\boldsymbol{U}}_0)^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}} - (\hat{\boldsymbol{U}}_0'\boldsymbol{M}_Z\hat{\boldsymbol{U}}_0)^{-1}\bar{\boldsymbol{X}}\boldsymbol{M}_Z\bar{\boldsymbol{X}}\}}, \tag{A.1}$$

where $\hat{\beta}_0 = \beta + O_p(N^{-1/2})$ and so $(\hat{\beta}_0 - \beta)^2 = o_p(N^{-1/2})$. We have $\boldsymbol{U}'\boldsymbol{U}/N \xrightarrow{p} \Sigma_{uu}$, $\boldsymbol{U}'\boldsymbol{M}_Z\boldsymbol{U}/(N-K) \xrightarrow{p} \Sigma_{uu}$, $\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}/N \xrightarrow{p} \boldsymbol{Q} + \Sigma_{vv}$ and $\bar{\boldsymbol{X}}'\boldsymbol{M}_Z\bar{\boldsymbol{X}}/(N-K) \xrightarrow{p} \Sigma_{vv}$. As $\hat{\boldsymbol{U}}_0 = \boldsymbol{U} - \boldsymbol{Z}\Pi(\hat{\beta}_0 - \beta) - \boldsymbol{V}(\hat{\beta}_0 - \beta)$, we also find

$$\frac{\hat{\boldsymbol{U}}_0'\hat{\boldsymbol{U}}_0}{N} = \frac{\boldsymbol{U}'\boldsymbol{U}}{N} - \left(\frac{\boldsymbol{U}'\boldsymbol{V}}{N} + \frac{\boldsymbol{V}'\boldsymbol{U}}{N}\right)\left(\hat{\beta}_0 - \beta\right) + o_p(N^{-1/2}),$$
$$\frac{\hat{\boldsymbol{U}}_0'\boldsymbol{M}_Z\hat{\boldsymbol{U}}_0}{N-K} = \frac{\boldsymbol{U}'\boldsymbol{M}_Z\boldsymbol{U}}{N-K} - \left(\frac{\boldsymbol{U}'\boldsymbol{M}_Z\boldsymbol{V}}{N-K} + \frac{\boldsymbol{V}'\boldsymbol{M}_Z\boldsymbol{U}}{N-K}\right)\left(\hat{\beta}_0 - \beta\right) + o_p(N^{-1/2}),$$

so, $\hat{\boldsymbol{U}}_0'\hat{\boldsymbol{U}}_0/N \xrightarrow{p} \Sigma_{uu}$, $\hat{\boldsymbol{U}}_0'\boldsymbol{M}_Z\hat{\boldsymbol{U}}_0/(N-K) \xrightarrow{p} \Sigma_{uu}$ and

$$\left(\frac{\hat{\boldsymbol{U}}_0'\hat{\boldsymbol{U}}_0}{N}\right)^{-1} = \left(\frac{\boldsymbol{U}'\boldsymbol{U}}{N}\right)^{-1} + \left(\frac{\boldsymbol{U}'\boldsymbol{U}}{N}\right)^{-1}\left(\frac{\boldsymbol{U}'\boldsymbol{V}}{N} + \frac{\boldsymbol{V}'\boldsymbol{U}}{N}\right)\left(\frac{\boldsymbol{U}'\boldsymbol{U}}{N}\right)^{-1}\left(\hat{\beta}_0 - \beta\right) + o_p(N^{-1/2}),$$

with a similar expression for $(\frac{\hat{\boldsymbol{U}}_0'\boldsymbol{M}_Z\hat{\boldsymbol{U}}_0}{N-K})^{-1}$. We thus find

$$\begin{aligned} N^{1/2}\left\{\left(\frac{\hat{\boldsymbol{U}}_0'\hat{\boldsymbol{U}}_0}{N}\right)^{-1} - \left(\frac{\boldsymbol{U}'\boldsymbol{U}}{N}\right)^{-1}\right\} &= \Sigma_{uu}^{-1}(\Sigma_{uv} + \Sigma_{vu})\Sigma_{uu}^{-1}N^{1/2}\left(\hat{\beta}_0 - \beta\right) + o_p(1), \\ N^{1/2}\left\{\left(\frac{\hat{\boldsymbol{U}}_0'\boldsymbol{M}_Z\hat{\boldsymbol{U}}_0}{N-K}\right)^{-1} - \left(\frac{\boldsymbol{U}'\boldsymbol{M}_Z\boldsymbol{U}}{N-K}\right)^{-1}\right\} &= \Sigma_{uu}^{-1}(\Sigma_{uv} + \Sigma_{vu})\Sigma_{uu}^{-1}N^{1/2}\left(\hat{\beta}_0 - \beta\right) + o_p(1). \end{aligned} \tag{A.2}$$

Consequently, for the denominator of (A.1) we find

$$\mathrm{tr}\left\{\left(\hat{\boldsymbol{U}}_0'\hat{\boldsymbol{U}}_0\right)^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}} - \left(\hat{\boldsymbol{U}}_0'\boldsymbol{M}_Z\hat{\boldsymbol{U}}_0\right)^{-1}\bar{\boldsymbol{X}}'\boldsymbol{M}_Z\bar{\boldsymbol{X}}\right\} \xrightarrow{p} \mathrm{tr}\left(\Sigma_{uu}^{-1}\boldsymbol{Q}\right). \tag{A.3}$$

For the numerator n, say, we have

$$\begin{aligned} n &= N^{1/2}\mathrm{tr}\left\{\left(\hat{\boldsymbol{U}}_0'\hat{\boldsymbol{U}}_0\right)^{-1}\boldsymbol{U}'\bar{\boldsymbol{X}} - \left(\hat{\boldsymbol{U}}_0'\boldsymbol{M}_Z\hat{\boldsymbol{U}}_0\right)^{-1}\boldsymbol{U}\boldsymbol{M}_Z\bar{\boldsymbol{X}}\right\} \\ &= N^{1/2}\mathrm{tr}\left\{(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\bar{\boldsymbol{X}} - (\boldsymbol{U}'\boldsymbol{M}_Z\boldsymbol{U})^{-1}\boldsymbol{U}\boldsymbol{M}_Z\bar{\boldsymbol{X}}\right\} + o_p(1). \end{aligned}$$

Under normality we have $\boldsymbol{V} = \boldsymbol{U}\Sigma_{uu}^{-1}\Sigma_{uv} + \boldsymbol{V}^{\perp}$, where $\boldsymbol{V}^{\perp}$ is independent of $\boldsymbol{U}$. We find

$$n = N^{1/2}\mathrm{tr}\left\{(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{Z}\Pi\right\} + N^{1/2}\mathrm{tr}\left[\left\{(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}' - (\boldsymbol{U}'\boldsymbol{M}_Z\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{M}_Z\right\}\boldsymbol{V}^{\perp}\right] + o_p(1).$$

For the first term, we have

$$\begin{aligned} N^{1/2}\mathrm{tr}\left\{(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{Z}\Pi\right\} &= \mathrm{tr}\left\{\Sigma_{uu}^{-1}(N^{-1/2}\boldsymbol{U}'\boldsymbol{Z}\Pi)\right\} + o_p(1) \\ &= \mathrm{vec}'\left(\Sigma_{uu}^{-1}\right)\left(N^{-1/2}\Pi'\boldsymbol{Z}' \otimes \boldsymbol{I}_T\right)\mathrm{vec}(\boldsymbol{U}') + o_p(1) \\ &\overset{a}{\sim} \mathcal{N}(0, v_1), \end{aligned} \tag{A.4}$$

where $v_1 = \mathrm{vec}'(\Sigma_{uu}^{-1})(\boldsymbol{Q} \otimes \Sigma_{uu})\mathrm{vec}(\Sigma_{uu}^{-1}) = \mathrm{tr}(\Sigma_{uu}^{-1}\boldsymbol{Q})$. For the second term, we find

$$\begin{aligned} &N^{1/2}\mathrm{tr}\left[\left\{(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}' - (\boldsymbol{U}'\boldsymbol{M}_Z\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{M}_Z\right\}\boldsymbol{V}^{\perp}\right] \\ &= \mathrm{vec}\left(\Sigma_{uu}^{-1}\right)\left[\left\{N^{-1/2}\boldsymbol{V}^{\perp\prime} - N^{1/2}(N-K)^{-1}\boldsymbol{V}^{\perp\prime}\boldsymbol{M}_Z\right\} \otimes \boldsymbol{I}_T\right]\mathrm{vec}(\boldsymbol{U}') + o_p(1) \\ &\overset{a}{\sim} \mathcal{N}(0, v_2), \end{aligned} \tag{A.5}$$

where $v_2 = \lambda\mathrm{tr}\{\Sigma_{uu}^{-1}(\Sigma_{vv} - \Sigma_{vu}\Sigma_{uu}^{-1}\Sigma_{uv})\}$. As the relevant terms in (A.4) and (A.5) are uncorrelated, we find the numerator satisfies $n \overset{a}{\sim} \mathcal{N}(0, v_1 + v_2)$. Together with the result for the denominator in (A.3), this gives the desired result (9).

A.3. The asymptotic distribution of the CUGMM estimator

As $\hat{\Sigma} = \hat{\Sigma}(\beta, \Pi) = N^{-1}(\boldsymbol{U}, \boldsymbol{V})'(\boldsymbol{U}, \boldsymbol{V})$ and

$$\hat{\Sigma}^{-1} = \begin{pmatrix} \hat{\Sigma}_{uu}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} + \begin{pmatrix} -\hat{\Sigma}_{uu}^{-1}\hat{\Sigma}_{uv} \\ \boldsymbol{I}_T \end{pmatrix}\left(\hat{\Sigma}_{vv} - \hat{\Sigma}_{vu}\hat{\Sigma}_{uu}^{-1}\hat{\Sigma}_{uv}\right)^{-1}\left(-\hat{\Sigma}_{vu}\hat{\Sigma}_{uu}^{-1}, \boldsymbol{I}_T\right),$$

the objective function $Q_{\text{GMM}}(\beta, \Pi) = \text{tr}\{\hat{\Sigma}^{-1}(\boldsymbol{U}, \boldsymbol{V})'\boldsymbol{P}_Z(\boldsymbol{U}, \boldsymbol{V})\}$ can be rewritten as

$$\begin{aligned} Q_{\text{GMM}}(\beta, \Pi) &= \text{tr}\left(\hat{\Sigma}_{uu}^{-1}\boldsymbol{U}'\boldsymbol{P}_Z\boldsymbol{U}\right) \\ &\quad + \text{tr}\left\{\left(\hat{\Sigma}_{vv} - \hat{\Sigma}_{vu}\hat{\Sigma}_{uu}^{-1}\hat{\Sigma}_{uv}\right)^{-1}\left(\boldsymbol{V} - \boldsymbol{U}\hat{\Sigma}_{uu}^{-1}\hat{\Sigma}_{uv}\right)'\boldsymbol{P}_Z\left(\boldsymbol{V} - \boldsymbol{U}\hat{\Sigma}_{uu}^{-1}\hat{\Sigma}_{uv}\right)\right\}. \end{aligned}$$

The second term on the right-hand side is nonnegative, and for $\Pi = \hat{\Pi}(\beta) = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\bar{\boldsymbol{X}} - (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{U}\hat{\Sigma}_{uu}^{-1}\hat{\Sigma}_{uv}$, where $\hat{\Sigma}_{uv} = \boldsymbol{U}'\{\boldsymbol{X} - \boldsymbol{Z}\hat{\Pi}(\beta)\}/N$, it is zero. This amounts to $\hat{\Pi}(\beta) = (\boldsymbol{Z}'\boldsymbol{M}_U\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{M}_U\bar{\boldsymbol{X}}$. The resulting concentrated objective function is given by

$$Q(\beta) = \text{tr}\left(\hat{\Sigma}_{uu}^{-1}\boldsymbol{U}'\boldsymbol{P}_Z\boldsymbol{U}\right).$$

The first-order condition for minimizing $Q(\beta)$ is given by $\text{tr}\{(\boldsymbol{U}'\boldsymbol{U})^{-1}\bar{\boldsymbol{X}}'\boldsymbol{M}_U\boldsymbol{P}_Z\boldsymbol{U}\} = 0$.

The one-step estimator in (12) satisfies

$$N^{1/2}\left(\hat{\beta}_1 - \beta\right) = N^{1/2}\frac{\text{tr}\left\{\left(\hat{\boldsymbol{U}}_0'\hat{\boldsymbol{U}}_0\right)^{-1}\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}_0}\boldsymbol{P}_Z\boldsymbol{U}\right\}}{\text{tr}\left\{\left(\hat{\boldsymbol{U}}_0'\hat{\boldsymbol{U}}_0\right)^{-1}\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}_0}\boldsymbol{P}_Z\bar{\boldsymbol{X}}\right\}}, \tag{A.6}$$

where $\hat{\boldsymbol{U}}_0 = \boldsymbol{Y} - \bar{\boldsymbol{X}}\hat{\beta}_0$ and $\hat{\beta}_0 = \beta + O_p(N^{-1/2})$, hence $(\hat{\beta}_0 - \beta)^2 = o_p(N^{-1/2})$. Let again $\boldsymbol{V} = \boldsymbol{U}\Sigma_{uu}^{-1}\Sigma_{uv} + \boldsymbol{V}^{\perp}$. As $\hat{\boldsymbol{U}}_0 = \boldsymbol{U} - \boldsymbol{Z}\Pi(\hat{\beta}_0 - \beta) - \boldsymbol{V}(\hat{\beta}_0 - \beta)$ we find the following results: $N(\hat{\boldsymbol{U}}_0'\hat{\boldsymbol{U}}_0)^{-1} = \Sigma_{uu}^{-1} + o_p(1)$, $N^{-1}\bar{\boldsymbol{X}}\boldsymbol{P}_Z\bar{\boldsymbol{X}} = \boldsymbol{Q} + \alpha\Sigma_{vv} + o_p(1)$ and $N^{-1}\bar{\boldsymbol{X}}'\boldsymbol{P}_{\hat{U}_0}\boldsymbol{P}_Z\bar{\boldsymbol{X}} = \alpha\Sigma_{vu}\Sigma_{uu}^{-1}\Sigma_{uv} + o_p(1)$. Consequently, the denominator of (A.6) satisfies

$$\text{tr}\left\{\left(\hat{\boldsymbol{U}}_0'\hat{\boldsymbol{U}}_0\right)^{-1}\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}_0}\boldsymbol{P}_Z\bar{\boldsymbol{X}}\right\} = \text{tr}\left\{\Sigma_{uu}^{-1}\left(\boldsymbol{Q} + \alpha\Sigma_{vv|u}\right)\right\} + o_p(1). \tag{A.7}$$

For the numerator of (A.6), n_{GMM}, we find

$$n_{\text{GMM}} = N^{1/2}\text{tr}\left\{\left(\hat{\boldsymbol{U}}_0'\hat{\boldsymbol{U}}_0\right)^{-1}\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}_0}\boldsymbol{P}_Z\boldsymbol{U}\right\} = \text{tr}\left\{\Sigma_{uu}^{-1}N^{-1/2}\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}_0}\boldsymbol{P}_Z\boldsymbol{U}\right\} + o_p(1).$$

Furthermore, using (A.2),

$$\begin{aligned} N^{-1/2}\bar{\boldsymbol{X}}'\boldsymbol{P}_Z\boldsymbol{U} &= N^{-1/2}(\boldsymbol{Z}\Pi + \boldsymbol{V}^{\perp})'\boldsymbol{P}_Z\boldsymbol{U} + N^{-1/2}\Sigma_{vu}\Sigma_{uu}^{-1}\boldsymbol{U}'\boldsymbol{P}_Z\boldsymbol{U}, \\ N^{-1/2}\bar{\boldsymbol{X}}'\boldsymbol{P}_U\boldsymbol{P}_Z\boldsymbol{U} &= N^{-1/2}(\boldsymbol{Z}\Pi + \boldsymbol{V}^{\perp})'\boldsymbol{P}_U\boldsymbol{P}_Z\boldsymbol{U} + N^{-1/2}\Sigma_{vu}\Sigma_{uu}^{-1}\boldsymbol{U}'\boldsymbol{P}_Z\boldsymbol{U} \\ &= \alpha N^{-1/2}(\boldsymbol{Z}\Pi + \boldsymbol{V}^{\perp})'\boldsymbol{U} + N^{-1/2}\Sigma_{vu}\Sigma_{uu}^{-1}\boldsymbol{U}'\boldsymbol{P}_Z\boldsymbol{U} + o_p(1), \\ N^{-1/2}\bar{\boldsymbol{X}}'\boldsymbol{P}_{\hat{U}_0}\boldsymbol{P}_Z\boldsymbol{U} &= N^{-1/2}\bar{\boldsymbol{X}}'\hat{\boldsymbol{U}}_0(\boldsymbol{U}'\boldsymbol{U})^{-1}\hat{\boldsymbol{U}}_0'\boldsymbol{P}_Z\boldsymbol{U} \\ &\quad + \left(\frac{\bar{\boldsymbol{X}}'\hat{\boldsymbol{U}}_0}{N}\right)\Sigma_{uu}^{-1}(\Sigma_{uv} + \Sigma_{vu})\Sigma_{uu}^{-1}\left(\frac{\hat{\boldsymbol{U}}_0'\boldsymbol{P}_Z\boldsymbol{U}}{N}\right)N^{1/2}\left(\hat{\beta}_0 - \beta\right) + o_p(1) \\ &= N^{-1/2}\bar{\boldsymbol{X}}'\boldsymbol{P}_U\boldsymbol{P}_Z\boldsymbol{U} - \alpha\left(\boldsymbol{Q} + \Sigma_{vv} - \Sigma_{vu}\Sigma_{uu}^{-1}\Sigma_{vu}\right)N^{1/2}\left(\hat{\beta}_0 - \beta\right) \\ &\quad + \alpha\Sigma_{vu}\Sigma_{uu}^{-1}(\Sigma_{uv} + \Sigma_{vu})N^{1/2}\left(\hat{\beta}_0 - \beta\right) + o_p(1) \\ &= N^{-1/2}\bar{\boldsymbol{X}}'\boldsymbol{P}_U\boldsymbol{P}_Z\boldsymbol{U} - \alpha\left(\boldsymbol{Q} + \Sigma_{vv} - \Sigma_{vu}\Sigma_{uu}^{-1}\Sigma_{uv}\right)N^{1/2}\left(\hat{\beta}_0 - \beta\right). \end{aligned}$$

For the numerator we thus find

$$n_{\text{GMM}} = N^{-1/2}\text{tr}\left\{\Sigma_{uu}^{-1}(\boldsymbol{Z}\Pi + \boldsymbol{V}^{\perp})'(\boldsymbol{P}_Z - \alpha\boldsymbol{I}_N)\boldsymbol{U}\right\} + \alpha\text{tr}\left\{\Sigma_{uu}^{-1}\left(\boldsymbol{Q} + \Sigma_{vv|u}\right)\right\}N^{1/2}\left(\hat{\beta}_0 - \beta\right) + o_p(1).$$

Combining this with (A.7) gives the desired result in (13).

A.4. The asymptotic distribution of the P-CIVE estimator

Concerning the P-CIVE estimator (14)

$$\hat{\beta}\boldsymbol{P}\text{−CIVE} = \frac{\text{tr}\left\{\left(\hat{\boldsymbol{U}}\prime\hat{\boldsymbol{U}}\right)^{-1}\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}}\boldsymbol{C}_Z(\hat{\lambda})\boldsymbol{Y}\right\}}{\text{tr}\left\{\left(\hat{\boldsymbol{U}}\prime\hat{\boldsymbol{U}}\right)^{-1}\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}}\boldsymbol{C}_Z(\hat{\lambda})\bar{\boldsymbol{X}}\right\}}, \tag{14}$$

where $\hat{\lambda} = \lambda + o_p(N^{-1/2})$, we find similar to the steps made following (A.6), that $N^{-1}\bar{\boldsymbol{X}}\boldsymbol{C}_Z(\hat{\lambda})\bar{\boldsymbol{X}} = \boldsymbol{Q} + o_p(1)$ and $N^{-1}\bar{\boldsymbol{X}}'\boldsymbol{P}_{\hat{U}}\boldsymbol{C}_Z(\hat{\lambda})\bar{\boldsymbol{X}} = o_p(1)$. Consequently, the denominator of (14) satisfies

$$\text{tr}\left\{\left(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}}\right)^{-1}\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}}\boldsymbol{C}_Z(\hat{\lambda})\bar{\boldsymbol{X}}\right\} = \text{tr}\left(\Sigma_{uu}^{-1}\boldsymbol{Q}\right) + o_p(1). \tag{A.8}$$

For the numerator of (14), $n_{\boldsymbol{P}\text{−CIVE}}$, we find

$$n_{\boldsymbol{P}\text{−CIVE}} = N^{1/2}\text{tr}\left\{\left(\hat{\boldsymbol{U}}'\hat{\boldsymbol{U}}\right)^{-1}\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}}\boldsymbol{C}_Z(\hat{\lambda})\boldsymbol{U}\right\} = \text{tr}\left\{\Sigma_{uu}^{-1}N^{-1/2}\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}}\boldsymbol{C}_Z(\lambda)\boldsymbol{U}\right\} + o_p(1).$$

Furthermore, using (A.2),

$$\begin{aligned}
N^{-1/2}\bar{\boldsymbol{X}}'\boldsymbol{C}_Z(\lambda)\boldsymbol{U} &= N^{-1/2}(\boldsymbol{Z}\Pi + \boldsymbol{V}^{\perp})'\boldsymbol{C}_Z(\lambda)\boldsymbol{U} + N^{-1/2}\Sigma_{vu}\Sigma_{uu}^{-1}\boldsymbol{U}'\boldsymbol{C}_Z(\lambda)\boldsymbol{U},\\
N^{-1/2}\bar{\boldsymbol{X}}'\boldsymbol{P}_U\boldsymbol{C}_Z(\lambda)\boldsymbol{U} &= N^{-1/2}(\boldsymbol{Z}\Pi + \boldsymbol{V}^{\perp})'\boldsymbol{P}_U\boldsymbol{C}_Z(\lambda)\boldsymbol{U} + N^{-1/2}\Sigma_{vu}\Sigma_{uu}^{-1}\boldsymbol{U}'\boldsymbol{C}_Z(\lambda)\boldsymbol{U}\\
&= N^{-1/2}\Sigma_{vu}\Sigma_{uu}^{-1}\boldsymbol{U}'\boldsymbol{C}_Z(\lambda)\boldsymbol{U} + o_p(1),\\
N^{-1/2}\bar{\boldsymbol{X}}'\boldsymbol{P}_{\hat{U}}\boldsymbol{C}_Z(\lambda)\boldsymbol{U} &= N^{-1/2}\bar{\boldsymbol{X}}'\hat{\boldsymbol{U}}(\boldsymbol{U}'\boldsymbol{U})^{-1}\hat{\boldsymbol{U}}'\boldsymbol{C}_Z(\lambda)\boldsymbol{U}\\
&\quad + \left(\frac{\bar{\boldsymbol{X}}'\hat{\boldsymbol{U}}}{N}\right)\Sigma_{uu}^{-1}(\Sigma_{uv} + \Sigma_{vu})\Sigma_{uu}^{-1}\left(\frac{\hat{\boldsymbol{U}}'\boldsymbol{C}_Z(\lambda)\boldsymbol{U}}{N}\right)N^{1/2}\left(\hat{\beta}_{\text{LIML}} - \beta\right) + o_p(1)\\
&= N^{-1/2}\bar{\boldsymbol{X}}'\boldsymbol{P}_U\boldsymbol{C}_Z(\lambda)\boldsymbol{U} + o_p(1).
\end{aligned}$$

For the numerator we thus find

$$n_{\boldsymbol{P}\text{−CIVE}} = N^{-1/2}\text{tr}\left\{\Sigma_{uu}^{-1}(\boldsymbol{Z}\Pi + \boldsymbol{V}^{\perp})'\boldsymbol{C}_Z(\lambda)\boldsymbol{U}\right\} + o_p(1).$$

Combining this with (A.8) gives the desired result in $N^{1/2}(\hat{\beta}\boldsymbol{P}\text{−CIVE}-\beta)\overset{a}{\sim}\mathcal{N}(0, \nu_{\boldsymbol{P}\text{−CIVE}})$, where $\nu_{\boldsymbol{P}\text{−CIVE}} = \nu_{\text{ML}}$ as given in (9). Finally,

$$\begin{aligned}
\text{tr}\left\{\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}}\boldsymbol{C}_Z^2(\hat{\lambda})\boldsymbol{M}_{\hat{U}}\bar{\boldsymbol{X}}(\hat{\boldsymbol{U}}\prime\hat{\boldsymbol{U}})^{-1}\right\} &= \text{tr}\left\{\Sigma_{uu}^{-1}\left(\boldsymbol{Q} + \lambda\Sigma_{vv|u}\right)\right\} + o_p(1),\\
\text{tr}\left\{\bar{\boldsymbol{X}}'\boldsymbol{M}_{\hat{U}}\boldsymbol{C}_Z(\hat{\lambda})\bar{\boldsymbol{X}}(\hat{\boldsymbol{U}}\prime\hat{\boldsymbol{U}})^{-1}\right\} &= \text{tr}\left(\Sigma_{uu}^{-1}\boldsymbol{Q}\right) + o_p(1),
\end{aligned}$$

and so $\hat{\nu}_{\text{ML}}$ as defined in (15) is many-instruments consistent for $\nu_{\boldsymbol{P}\text{−CIVE}}$.

References

Akashi, K., Kunitomo, N. (2012). Some properties of the LIML estimator in a dynamic panel structural equation. *Journal of Econometrics* 166(2):167–183.
Alvarez, J., Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71(4):1121–1159.
Anderson, T. W., Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. The *Annals of Mathematical Statistics* 20(1):46–63.
Arellano, M. (2016). Modelling optimal instrumental variables for dynamic panel data models. *Research in Economics* 70(2):238–261.
Arellano, M., Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies* 58(2):277–297.
Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica* 62(3):657–681.
Bekker, P. A., Crudu, F. (2015). Jackknife instrumental variable estimation with heteroskedasticity. *Journal of Econometrics* 185(2):332–342.
Bekker, P. A., Wansbeek, T. J. (2016). Simple many-instruments robust standard errors through concentrated instrumental variables. *Economics Letters* 149:52–55.
Bun, M., Sarafidis, V. (2015). Dynamic panel data models. In: Baltagi, B. H. ed., *The Oxford Handbook of Panel Data*, Oxford: Oxford University Press, pp. 373–396.

Hausman, J. A., Newey, W. K., Woutersen, T., Chao, J. C., Swanson, N. R. (2012). Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics* 3(2):211–255.

Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics*68(1):53–78.

Moral-Benito, E. (2013). Likelihood-based estimation of dynamic panels with predetermined regressors. *Journal of Business & Economic Statistics* 31(4):451–472.

Wansbeek, T. J., Prak, D. R. J. (2017). LIML in the static linear panel data model. *Econometric Reviews* 36(1–3): 385–395.

Ziliak, J. P. (1997). Efficient estimation with panel data when instruments are predetermined: An empirical comparison of moment-condition estimators. *Journal of Business & Economic Statistics* 15(4):419–431.