UNIVERSITÀ DI SIENA 1240

QUADERNI DEL DIPARTIMENTO

DI ECONOMIA POLITICA E STATISTICA

Federico Crudu
Giovanni Mellace
Zsolt Sándor

Inference in instrumental variables models with heteroskedasticity and many instruments

n. 761 – Novembre 2017

Inference in instrumental variables models with heteroskedasticity and many instruments*

Federico Crudu†

Università di Siena

Pontificia Universidad Católica de Valparaíso

CRENoS

Giovanni Mellace‡

University of Southern Denmark

Zsolt Sándor§

Sapientia University Miercurea Ciuc

April 2017

Abstract

This paper proposes a specification test for instrumental variable models that is robust to the presence of heteroskedasticity. The test can be seen as a generalization the Anderson-Rubin test. Our approach is based on the jackknife principle. We are able to show that under the null the proposed statistic has a Gaussian limiting distribution. Moreover, a simulation study shows its competitive finite sample properties in terms of size and power.

Key words: Instrumental variables, heteroskedasticity, many instruments, jackknife, specification tests, overidentification tests.

JEL classification: C12, C13, C23.

*Financial support from FONDECYT grant n. 11140433 (F. Crudu) and grant PN-II-ID-PCE-2012-4-0066 of the Romanian Ministry of National Education, CNCS-UEFISCDI (Z. Sándor) is gratefully acknowledged.

†federico.crudu@gmail.com

‡giome@sam.sdu.dk

§zsosan@gmail.com

1 Introduction

Instrumental variables (IV) techniques are ubiquitous in the contemporary econometric practice. The popularity of these methods, however, does not free them from criticisms. In fact, while there is an uncountable amount of papers using the IV framework for estimation and inference, an important strand of literature has focussed on its weaknesses and highlighted its failures. See e.g. the beautiful survey of Imbens (2014), see also Andrews & Stock (2007), Bekker (1994) and Bound et al. (1995).

From this critical approach, many authors have proposed alternatives and improvements over the standard IV framework. In particular, the work of Angrist & Krueger (1991) on the effect of schooling on earnings and the subsequent critique to that paper by Bound et al. (1995) have initiated a discussion on the role of weak instruments and many instruments on IV estimation and inference.

With respect to inference one of the most influential contributions is the paper by Anderson & Rubin (1949, henceforth AR) where they suggest conducting inference on the parameter (vector) by means of the limited information maximum likelihood (LIML) objective function. Notoriously, this approach has the advantage of being robust to the presence of weak instruments. However, when the number of instruments grows larger than the number of parameters, the performance of the AR test starts deteriorating (e.g. Anatolyev & Gospodinov, 2011). Kleibergen (2002) proposes a modification of the AR statistic that is robust to the presence of many instruments (see also Moreira, 2009). Moreira (2003), on the other hand, suggests replacing standard asymptotic critical values with conditional quantiles. The resulting conditional likelihood ratio (CLR) test enjoys excellent power properties (see Andrews et al., 2006). Other tests have been proposed by Steiger & Stock (1997), Wang & Zivot (1998) and Zivot et al. (1998). However, those tests tend to be conservative and are generally outperformed by the CLR test in terms of power (Stock et al., 2002).

The objective of this paper is to construct an AR-type statistic for linear IV models in the presence of many, potentially weak, instruments and heteroskedasticity. The starting point of our work is the paper by Bekker & Crudu (2015). The analysis is closely related to the papers by Anatolyev & Gospodinov (2011) and Chao et al. (2014). The plan of the paper is as follows: Section 2 introduces the model and the test statistics, Section 3 describes the main asymptotic results and the associated assumptions, Section 4 and Section **??** contain the simulation results and an empirical application on XXX data respectively, Section 5 concludes the paper. Proofs and and auxiliary results are relegated to the Appendix.

2 Model, assumptions and test statistics

Let us consider the model

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

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

where $\boldsymbol{y}$ is a vector of dimension n and $\boldsymbol{X}$ is a $n \times g$ matrix. Throughout the paper it is assumed that the $n \times k$ matrix of instruments $\boldsymbol{Z}$ is nonstochastic and $E[\boldsymbol{X}] = \boldsymbol{Z}\boldsymbol{\Pi}$. Such assumptions are made for convenience and may be generalized, see e.g. Chao et al. (2014), Hausman et al. (2012), Bekker (1994). The rows of the disturbance couple $(\boldsymbol{\varepsilon}, \boldsymbol{U})$, say $(\varepsilon_i, \boldsymbol{U}_i')$ $i = 1, \ldots, n$, are independent with zero mean and covariance matrices

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \sigma_i^2 & \boldsymbol{\sigma}_{i12} \\ \boldsymbol{\sigma}_{i21} & \boldsymbol{\Sigma}_{i22} \end{pmatrix} \tag{3}$$

while the covariance matrix of the rows $(y_i, \boldsymbol{X}_i')$ are

$$\boldsymbol{\Omega}_i = \begin{pmatrix} 1 & \boldsymbol{\beta}' \\ \boldsymbol{0} & \boldsymbol{I}_g \end{pmatrix} \boldsymbol{\Sigma}_i \begin{pmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{\beta} & \boldsymbol{I}_g \end{pmatrix}. \tag{4}$$

The symmetric jackknife instrumental variable estimator (SJIVE) introduced by Bekker & Crudu (2015) estimates consistently, in the many (weak) instruments sense, the parameter vector $\boldsymbol{\beta}$. The SJIVE is defined as

$$\widehat{\boldsymbol{\beta}}_{SJIVE} = \arg\min_{\boldsymbol{\beta}} Q_{SJIVE}(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}} \frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{C}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{B}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})} \tag{5}$$

and, given the projection matrix $\boldsymbol{P} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$ and the diagonal matrix $\boldsymbol{D}$ containing the diagonal elements of $\boldsymbol{P}$,

$$\begin{aligned}
&\boldsymbol{C} = \boldsymbol{A} - \boldsymbol{B} \\
&\boldsymbol{A} = \boldsymbol{P} + \boldsymbol{\Delta} \\
&\boldsymbol{B} = (\boldsymbol{I}_n - \boldsymbol{P})\boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1}(\boldsymbol{I}_n - \boldsymbol{P}) \\
&\boldsymbol{\Delta} = \boldsymbol{P}\boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1}\boldsymbol{P} - \frac{1}{2}\boldsymbol{P}\boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1} - \frac{1}{2}\boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{D})^{-1}\boldsymbol{P}.
\end{aligned}$$

To compute the SJIVE consider the partition $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$, where $\boldsymbol{X}_2 = \boldsymbol{Z}_2$ and $\boldsymbol{Z} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2)$. Define $\boldsymbol{C}^* = \boldsymbol{C} - \boldsymbol{A}\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{A}$, then

$$\widehat{\boldsymbol{\beta}}_{SJIVE} = \Big(\boldsymbol{X}'\boldsymbol{C}\boldsymbol{X} - \lambda\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}\Big)^{-1}\Big(\boldsymbol{X}'\boldsymbol{C}\boldsymbol{y} - \lambda\boldsymbol{X}'\boldsymbol{B}\boldsymbol{y}\Big) \tag{6}$$

where

$$\lambda = \lambda_{\min}\Big\{\Big((\boldsymbol{y}, \boldsymbol{X}_1)'\boldsymbol{B}(\boldsymbol{y}, \boldsymbol{X}_1)\Big)^{-1}\Big((\boldsymbol{y}, \boldsymbol{X}_1)'\boldsymbol{C}^*(\boldsymbol{y}, \boldsymbol{X}_1)\Big)\Big\}.$$

3 Asymptotic results

In this section we introduce a set of assumptions that are used to prove our asymptotic results. Furthermore, we generalize a result due to Anatolyev & Gospodinov (2011) to the heteroskedastic case and we introduce our main results.

3.1 Assumptions

The assumptions we use are similar to those in Bekker & Crudu (2015). Additional assumptions are included to generalize some results due to Anatolyev & Gospodinov (2011).

Assumption 1. *The generic diagonal element of the projection matrix* $\boldsymbol{P}$*,* P_{ii}*, satisfies* $\max_i P_{ii} \leq 1 - 1/c_u$.

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

Let us define the signal matrix as $\boldsymbol{H} = \boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi}$. Moreover, we can partition $\boldsymbol{\Sigma}$ as we partitioned $\boldsymbol{\Sigma}_i$ and define $\boldsymbol{\Omega}$ as we defined $\boldsymbol{\Omega}_i$ in (4). Hence, $\lim_{n\to\infty} \frac{1}{k}\,\mathrm{E}[(\boldsymbol{y}, \boldsymbol{X})'\boldsymbol{B}(\boldsymbol{y}, \boldsymbol{X})] = \boldsymbol{\Omega}$ and $\mathrm{E}[\boldsymbol{X}'\boldsymbol{C}\boldsymbol{X}] = \boldsymbol{H}$.

Assumption 3. $plim_{n\to\infty} \frac{1}{k}(\boldsymbol{y}, \boldsymbol{X})'\boldsymbol{B}(\boldsymbol{y}, \boldsymbol{X}) = \boldsymbol{\Omega}$ *and* $plim_{n\to\infty} \boldsymbol{H}^{-1}\boldsymbol{X}'\boldsymbol{C}\boldsymbol{X} = \boldsymbol{I}_g$.

Assumption 4. $\mathrm{E}[\varepsilon_i^4] \leq c_u$.

Let $r_{\min} = \lambda_{\min}(\boldsymbol{H})$ and $r_{\max} = \lambda_{\max}(\boldsymbol{H})$ be the smallest eigenvalue and the largest eigenvalue of the signal matrix respectively.

Assumption 5. $r_{\min} \to \infty$.

Assumption 6. $k/r_{\min} \to \kappa$ *and* κ *being a constant.*

Assumption 7. $k/r_{\max} \to \infty$*,* $\sqrt{k}/r_{\min} \to 0$.

3.2 The Anderson-Rubin test

The Anderson-Rubin statistic is a popular choice to test a null hypothesis defined as $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. The statistic is defined as

$$AR = (n-k)\,\frac{\boldsymbol{\varepsilon}_0'\boldsymbol{P}\boldsymbol{\varepsilon}_0}{\boldsymbol{\varepsilon}_0'\left(\boldsymbol{I}_n - \boldsymbol{P}\right)\boldsymbol{\varepsilon}_0}$$

and it is chi square distributed with k degrees of freedom. In the many instruments context and in the presence of homoskedaticity, the behaviour of the AR test has been studied by Andrews & Stock (2007) and Anatolyev & Gospodinov (2011), among others. The following theorem generalizes the results in Lemma 1 of Anatolyev & Gospodinov (2011) to the heteroskedastic case. Let us define $\sigma^2_{(n)} = \frac{1}{n}\sum_{i=1}^{n}\sigma^2$ and $W_n = \frac{2}{k}\sum_{i\neq j}P^2_{ij}\sigma^2_i\sigma^2_j$.

Theorem 1. *Let us assume that Assumption 2 and Assumption 4 hold. Moreover, given* $\lim_{n\to\infty}\frac{k}{n} = \lambda$, *where* $0 < \lambda < 1$, *let* $\frac{1}{n}\sum_{i=1}^{n}(P_{ii} - \frac{k}{n})^2 \to 0$ *and assume that* $\lim_{n\to\infty}\sigma^2_{(n)} = \sigma^2_0$ *and* $\lim_{n\to\infty}W_n = W_0$ *exist. Then,*

$$\left(1 - \frac{k}{n}\right)\frac{\sigma^2_{(n)}}{\sqrt{W_n}}\sqrt{k}\left(\frac{AR}{k} - 1\right) \to_d N\left(0, 1\right). \tag{7}$$

We note that under homoskedasticity

$$\frac{\sigma^2_{(n)}}{\sqrt{W_n}} \to \frac{\sigma^2_i}{\sqrt{2\left(1-\lambda\right)\sigma^2_i}} = \frac{1}{\sqrt{2\left(1-\lambda\right)}},$$

so

$$\sqrt{k}\left(\frac{AR}{k} - 1\right) \to_d N\left(0, \frac{2}{1-\lambda}\right),$$

which is exactly as in Lemma 1 of Anatolyev & Gospodinov (2011).

Remark 1. *The asymptotic distribution result in Theorem 1 has two important implications. First, the fact that the statistic provided by Anatolyev & Gospodinov (2011) has the limit*

$$AR_{AG} = \sqrt{k}\left(\frac{AR}{k} - 1\right) \to_d N\left(0, \frac{W_0}{\sigma^4_0\left(1-\lambda\right)^2}\right)$$

means that this test statistic is not valid under heteroskedasticity. In fact the asymptotic *size of this test is*

$$\Pr\left(AR_{AG} > \Phi^{-1}(1-\alpha)\right) = \Pr\left(\frac{\sigma_0^2(1-\lambda)}{\sqrt{W_0}} AR_{AG} < \frac{\sigma_0^2(1-\lambda)}{\sqrt{W_0}}\Phi^{-1}(\alpha)\right)$$
$$\to \Phi\left(\frac{\sigma_0^2(1-\lambda)}{\sqrt{W_0}}\Phi^{-1}(\alpha)\right).$$

Second, the result in Theorem 1 also shows how to correct the statistic provided by Anatolyev & Gospodinov (2011) in order to become feasible when the errors are heteroskedastic. In fact we can easily construct statistics that approximate $\sigma^2_{(n)}$ *and* W_n*.*

3.3 Inference with heteroskedasticity and many instruments

The test statistic we propose is based on the numerator of the objective function in equation (5). This is,

$$Q(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{C}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}). \tag{8}$$

Consider testing the following null hypothesis

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \tag{9}$$

where $\boldsymbol{\beta}$ is the true parameter. The test statistic is defined as

$$T_1 = \frac{1}{\sqrt{k}} \frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_0)'\boldsymbol{C}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_0)}{\sqrt{\widehat{V}(\boldsymbol{\beta}_0)}}, \quad \widehat{V}(\boldsymbol{\beta}_0) = \frac{2}{k}\boldsymbol{\varepsilon}_0^{(2)\prime}\boldsymbol{C}^{(2)}\boldsymbol{\varepsilon}_0^{(2)} \tag{10}$$

where $\boldsymbol{\varepsilon}_0 = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_0$ and the superscript "$^{(2)}$" indicates the elementwise product of two conformable matrices or vectors. Sometimes we are interested only in performing inference on a subset of parameters. In particular, we would like to test the coefficients associated to the endogenous variables. Let us now define the parameter vector as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ and suppose we want to test the following null hypothesis

$$H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}. \tag{11}$$

In this case, in order to compute the T_1 test, we need a consistent estimator for $\boldsymbol{\beta}_2$. Under the null, a consistent estimator is, for example,

$$\widetilde{\boldsymbol{\beta}}_2 = \left(\boldsymbol{X}_2'\boldsymbol{C}\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\boldsymbol{C}\left(\boldsymbol{y} - \boldsymbol{X}_1\boldsymbol{\beta}_{10}\right) \tag{12}$$

and the corresponding T_1 statistic is

$$T_1 = \frac{1}{\sqrt{k}}\frac{\widetilde{\boldsymbol{\varepsilon}}'\boldsymbol{C}\widetilde{\boldsymbol{\varepsilon}}}{\sqrt{\widehat{V}(\boldsymbol{\beta}_{10}, \widetilde{\boldsymbol{\beta}}_2)}}), \quad \widehat{V}(\boldsymbol{\beta}_{10}, \widetilde{\boldsymbol{\beta}}_2) = \frac{2}{k}\widetilde{\boldsymbol{\varepsilon}}^{(2)'}\boldsymbol{C}^{(2)}\widetilde{\boldsymbol{\varepsilon}}^{(2)} \tag{13}$$

where $\widetilde{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}_1\boldsymbol{\beta}_{10} - \boldsymbol{X}_2\widetilde{\boldsymbol{\beta}}_2$.
Let us now consider a nominal level α and let $q_{1-\alpha}^{N(0,1)}$ be the $(1-\alpha)$-th quantile of the Normal distribution. Then, the null hypothesis is rejected if $T_1 \geq q_{1-\alpha}^{N(0,1)}$.
Let us now define the following statistic

$$T_2 = \frac{1}{\sqrt{k}}\frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_0)'(\boldsymbol{P} - \boldsymbol{D})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_0)}{\sqrt{\widehat{V}(\boldsymbol{\beta}_0)}}, \quad \widehat{V}(\boldsymbol{\beta}_0) = \frac{2}{k}\boldsymbol{\varepsilon}_0^{(2)'}(\boldsymbol{P} - \boldsymbol{D})^{(2)}\boldsymbol{\varepsilon}_0^{(2)}. \tag{14}$$

Similarly to 15 we have

$$T_2 = \frac{1}{\sqrt{k}}\frac{\widetilde{\boldsymbol{\varepsilon}}'(\boldsymbol{P} - \boldsymbol{D})\widetilde{\boldsymbol{\varepsilon}}}{\sqrt{\widehat{V}(\boldsymbol{\beta}_{10}, \widetilde{\boldsymbol{\beta}}_2)}}), \quad \widehat{V}(\boldsymbol{\beta}_{10}, \widetilde{\boldsymbol{\beta}}_2) = \frac{2}{k}\widetilde{\boldsymbol{\varepsilon}}^{(2)'}(\boldsymbol{P} - \boldsymbol{D})^{(2)}\widetilde{\boldsymbol{\varepsilon}}^{(2)} \tag{15}$$

where $\widetilde{\boldsymbol{\beta}}_2$ is some consistent estimator of $\boldsymbol{\beta}_2$. The following theorems provide the asymptotic distribution for T_1 and T_2.

Theorem 2. *If Assumptions 1, 4 are satisfied,* $\sum_{i=1}^{n}|P_{ij}| < c_u$ *(see Van Hasselt, 2010),* $\lim \frac{k}{n} > 0$ *and* $\sigma_i^2 \geq \underline{\sigma}^2$ *for any* i*, then*

1. *under* $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ *we have* $T_1 \rightarrow_d N(0,1)$,

2. *under* $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$ *and if the additional Assumptions 3, 5 and 6 or 7 are satisfied we have* $\widetilde{\boldsymbol{\beta}}_2 \rightarrow_p \boldsymbol{\beta}_2$ *and* $T_1 \rightarrow_d N(0,1)$.

Theorem 3. *If Assumptions 1, 4 are satisfied,* $\lim \frac{k}{n} > 0$ *and* $\sigma_i^2 \geq \underline{\sigma}^2$ *for any* i*, then*

1. *under* $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ *we have* $T_2 \rightarrow_d N(0,1)$,

2. *under* $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$ *and if the additional Assumptions 3, 5 and 6 or 7 are satisfied we have* $\widetilde{\boldsymbol{\beta}}_2 \rightarrow_p \boldsymbol{\beta}_2$ *and* $T_2 \rightarrow_d N(0,1)$.

The limit distribution of T_2 holds without the condition $\sum_{i=1}^{n} |P_{ij}| < c_u$ (Van Hasselt, 2010) that is used for T_1. Under the conditions of Theorem 2, it holds that $V_n \geq c_u > 0$ for any n.

Remark 2. *The test statistic* T_1 *proposed in this paper is also valid under homoskedasticity. Its asymptotic distribution requires the assumption that the main diagonal elements* P_{ii}, $i = 1, ..., n$, *of the projection matrix* $\boldsymbol{P} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$ *should be bounded away from* 1. *The test statistic proposed by Anatolyev & Gospodinov (2011) requires the stronger assumption that the main diagonal elements of* $\boldsymbol{P}$ *converge to* $\lambda = \lim \frac{k}{n}$. *Therefore, even under homoskedasticity* T_1 *has broader applicability than the test statistic proposed by Anatolyev & Gospodinov (2011). This difference in the assumptions required comes from the fact that the former test statistic does not involve the diagonal elements of* $\boldsymbol{P}$ *while the latter statistic does.*

4 Monte Carlo simulations

We study the finite sample properties of T_1 and T_2 in terms of size and power.[1] We make inference on the full parameter vector (see Figure 1 and Figure 3) and on the sole parameter associated to the endogenous variable (see Figure 2 and Figure 4). The T_1 test

[1]The results associated to the T_2 test are omitted as they are nearly identical to the those of the T_1 test.

is compared to the AR test is the version proposed by Anatolyev & Gospodinov (2011) both in the homoskedastic version (AR_{AG}) and in the heteroskedastic version (HAR_{AG}) as proposed in Theorem 1. This comparison is useful for two reasons: first, we are able to see whether or not our test works in the homoskedastic context; second, it gives us a clear idea of how much we gain by using a proper test in the heteroskedastic case.[2] In order to implement the HAR_{AG} test we need to estimate W_0 and σ_0^2. More specifically, we define

$$\widehat{W}_n = \frac{2}{k}\boldsymbol{\varepsilon}_0^{(2)\prime}(\boldsymbol{P}-\boldsymbol{D})^2\boldsymbol{\varepsilon}_0^{(2)}, \qquad \widehat{\sigma}^2_{(n)} = \frac{\boldsymbol{\varepsilon}_0'\boldsymbol{\varepsilon}_0}{n}.$$

4.1 Data generating process

We use the same Monte Carlo set up as Hausman et al. (2012). The data generating process is given by

$$\boldsymbol{y} = \boldsymbol{\iota}\gamma + \boldsymbol{x}\beta + \boldsymbol{\varepsilon}$$
$$\boldsymbol{x} = \boldsymbol{z}\pi + \boldsymbol{v}$$

where $\gamma = \beta = 1$. The strength of the instruments changes according to the concentration parameter $\mu^2 = n\pi^2$. Moreover, $n = 800$, $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0},\ \boldsymbol{I}_n)$ and independently $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$. The disturbances vector $\boldsymbol{\varepsilon}$ is generated as

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

where $\rho = 0.3$, $\psi = 0.86$ and conditional on $\boldsymbol{z}$, independent of $\boldsymbol{v}$, $\boldsymbol{w}_1 \sim \mathcal{N}(\boldsymbol{0},\ \text{Diag}(\boldsymbol{z})^2)$ and $\boldsymbol{w}_2 \sim \mathcal{N}(\boldsymbol{0},\ \psi^2\boldsymbol{I}_n)$. Moreover, $\phi \in \{0, 1.38072\}$. The instrument matrix $\boldsymbol{Z}$ is given by matrices with rows $(1, z_i, z_i^2, z_i^3, z_i^4)$ and $(1, z_i, z_i^2, z_i^3, z_i^4, z_i b_{1i}, \ldots, z_i b_{\ell i})$, $\ell = 5, 15, 35, 55, 75, 95$, respectively, where, independent of other random variables, the elements $b_{1i}, \ldots, b_{\ell i}$ are i.i.d. Bernoulli distributed with $p = 1/2$. We replicate our experiments 5000 times. The

[2] The AR_{AG} test works quite well in the homoskedastic case but it is not designed to work in the presence of heteroskedasticity.

size properties of T_1 and T_2 investigated by means of PP-plots as described in Davidson & MacKinnon (1998). When using the T_1 test we consider both $H_0 : (\gamma, \beta)' = (1, 1)'$ and $H_0 : \beta = 1$.

Our simulation results in Figures 1 to 4 confirm that the T_1 test (and the T_2 test) works well both in the homoskedastic case and the heteroskedastic case. We notice that in the homoskedastic case it performs similarly to the AR_{AG} test. Moreover, we notice that there is a trade off between size and power with respect to k. This is, as k grows the empirical size approaches the nominal size, but the power curves tend to get wider. This is not necessarily a justification for using a small k as it would imply that the test rejects too often.

5 Conclusion

This paper introduces a specification test for the parameters of a linear model in the presence of endogeneity, heteroskedasticity and many, potentially weak instruments. The test is easy to build as it is based on the numerator of the SJIVE estimator proposed by Bekker & Crudu (2015). It is possible to show that, after appropriate rescaling, the limiting distribution of the test statistic is standard normal. Simulation evidence shows that, in finite samples, the proposed test outperforms its competitors.

A Appendix

The central limit theorem (CLT) used in the proofs, and described below is the CLT for quadratic forms of Keleijan & Prucha (2001, Theorem 1) We also use a special case of the CLT of Chao et al. (2012).

Theorem A.1 (Central Limit Theorem by Keleijan & Prucha (2001))**.** *Consider the quadratic form* $Q = \boldsymbol{\varepsilon}'\boldsymbol{A}\boldsymbol{\varepsilon}$ *such that* (*a*) $\mathrm{E}\left[\varepsilon_i\right] = 0$ *and* $\varepsilon_1, ..., \varepsilon_n$ *are independent;* (*b*) $\boldsymbol{A}$ *symmetric, with diagonal elements* $a_{ii} = 0$ *and* $\sup_{j,n} \sum_{i=1}^{n} |a_{ij}| < \infty$*;* (*c*) $\mathrm{E}\left[|\varepsilon_i|^{2+\eta}\right] < \infty$*, for* *some* $\eta > 0$; *(d)* $\frac{1}{n} \operatorname{Var}[Q] \geq c > 0$. *Then*

$$\frac{Q}{\sqrt{\operatorname{Var}[Q]}} \rightarrow_d N(0,1).$$

Theorem A.2 (Central Limit Theorem by Chao et al. (2012))**.** *Consider the quadratic form* $Q = \sum_{i \neq j} a_{ij} \varepsilon_i \varepsilon_j$ *such that (a)* $\operatorname{E}[\varepsilon_i] = 0$ *and* $\varepsilon_1, \ldots, \varepsilon_n$ *are independent; (b)* $\boldsymbol{A}$ *symmetric, idempotent,* $a_{ii} \leq c_u < 1$ *and rank* $(\boldsymbol{A}) = k$*, where* $k \to \infty$ *as* $n \to \infty$*; (c)* $\operatorname{E}[\varepsilon_i^4] < \infty$*; (d)* $\frac{1}{n} Var[Q] \geq c_u > 0$ *. Then*

$$\frac{Q}{\sqrt{VarQ}} \rightarrow_d N(0,1).$$

Let $\boldsymbol{C}$ be a $n \times n$ matrix with zero diagonal whose elements depend on $\boldsymbol{Z}$. Suppose also that for appropriate positive numbers $\underline{c}$ and $\overline{c}$

$$\underline{c}|P_{ij}| \leq |C_{ij}| \leq \overline{c}|P_{ij}| \text{ for any } i \neq j.$$

Proof of Theorem 1. Under $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ we have

$$\sqrt{k}\left(\frac{AR}{k} - 1\right) = \frac{\frac{1}{\sqrt{k}}\left(\frac{n-k}{k}\boldsymbol{\varepsilon}'\boldsymbol{P}\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}'\left(\boldsymbol{I} - \boldsymbol{P}\right)\boldsymbol{\varepsilon}\right)}{\frac{1}{k}\boldsymbol{\varepsilon}'\left(\boldsymbol{I} - \boldsymbol{P}\right)\boldsymbol{\varepsilon}} = \frac{n}{k}\frac{\frac{1}{\sqrt{k}}\left(\boldsymbol{\varepsilon}'\boldsymbol{P}\boldsymbol{\varepsilon} - \frac{k}{n}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right)}{\frac{1}{k}\boldsymbol{\varepsilon}\left(\boldsymbol{I} - \boldsymbol{P}\right)\boldsymbol{\varepsilon}}. \tag{16}$$

Note that

$$\frac{1}{\sqrt{k}}\left(\boldsymbol{\varepsilon}'\boldsymbol{P}\boldsymbol{\varepsilon} - \frac{k}{n}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right) = \frac{1}{\sqrt{k}}\sum_{i \neq j} P_{ij}\varepsilon_i\varepsilon_j + \frac{1}{\sqrt{k}}\sum_i \left(P_{ii} - \frac{k}{n}\right)\varepsilon_i^2 \equiv E_1 + E_2. \tag{17}$$

We can apply the CLT from Kelejian and Prucha (2001) to the quadratic form

$$R = \sum_{i \neq j} P_{ij}\varepsilon_i\varepsilon_j$$

involved in E_1. We obtain that

$$\frac{R}{\sqrt{kW_n}} \rightarrow_d N(0,1),$$

where

$$W_n = \frac{\mathrm{Var}[R]}{k} = \frac{2}{k}\sum_{i\neq j} P_{ij}^2\sigma_i^2\sigma_j^2$$

with the property that

$$\frac{1}{n}\mathrm{Var}[R] = \frac{2}{n}\sum_{i\neq j} P_{ij}^2\sigma_i^2\sigma_j^2 \geq \frac{2\underline{\sigma}^4}{n}\sum_{i\neq j} P_{ij}^2 \geq \frac{2\underline{\sigma}^4}{n}\frac{k}{c_u},$$

(the latter inequality from (23)), which is bounded away from 0. [3] Consequently, W_n is bounded between two positive numbers. We obtain that $E_1/\sqrt{W_n} \rightarrow_d N(0,1)$.

Regarding E_2 we note that by Assumption 4

$$\mathrm{Var}[E_2] = \frac{1}{k}\sum_i \left(P_{ii} - \frac{k}{n}\right)^2 var\left(\varepsilon_i^2\right) \leq \frac{c_u}{k}\sum_i \left(P_{ii} - \frac{k}{n}\right)^2.$$

Using the assumption $\frac{1}{k}\sum_i \left(P_{ii} - \frac{k}{n}\right)^2 \rightarrow 0$ (Anatolyev and Gospodinov use $\frac{1}{k}\sum_i \left(P_{ii} - \lambda\right)^2 \rightarrow 0$, where $\lambda = \lim \frac{k}{n}$), we obtain that $\mathrm{Var}[E_2] = o(1)$. Consequently, $E_2 = o_p(1)$; therefore,

$$\frac{E_1 + E_2}{\sqrt{W_n}} \rightarrow_d N(0,1). \tag{18}$$

Regarding the denominator involved in (16) we observe that

$$\frac{1}{k}\boldsymbol{\varepsilon}'(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{\varepsilon} = \frac{1}{k}\left(1 - \frac{k}{n}\right)\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \frac{1}{k}\boldsymbol{\varepsilon}'\left(\boldsymbol{P} - \frac{k}{n}\boldsymbol{I}\right)\boldsymbol{\varepsilon}.$$

[3]Note that Anatolyev & Gospodinov (2011) prove this from the assumption that $\frac{1}{k}\sum_i \left(P_{ii} - \lambda\right)^2 \rightarrow 0$. It seems that this implies our Assumption 1 but not the other way around, so our Assumption 1 is weaker.

The second term is just the expression from (17) divided by $\sqrt{k}$, that is,

$$\frac{1}{k}\boldsymbol{\varepsilon}'\left(\boldsymbol{I}-\boldsymbol{P}\right)\boldsymbol{\varepsilon}=\frac{1}{k}\left(1-\frac{k}{n}\right)\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}-\frac{1}{\sqrt{k}}\left(E_1+E_2\right)=\frac{1}{k}\left(1-\frac{k}{n}\right)\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}+O_p\left(\frac{1}{\sqrt{k}}\right).$$

Using Assumption 4 and the Law of Large Numbers, using the notation

$$\sigma^2_{(n)}=\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2$$

we have that

$$\frac{1}{n}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}-\sigma^2_{(n)}=O_p\left(\frac{1}{\sqrt{k}}\right). \tag{19}$$

Consequently,

$$\frac{1}{k}\boldsymbol{\varepsilon}'\left(\boldsymbol{I}-\boldsymbol{P}\right)\boldsymbol{\varepsilon}=\frac{n}{k}\left(1-\frac{k}{n}\right)\sigma^2_{(n)}+O_p\left(\frac{1}{\sqrt{k}}\right).$$

Now, from equation (16) and the fact that $\frac{n}{k}\left(1-\frac{k}{n}\right)\sigma^2_{(n)}$ is bounded between two positive numbers, we have

$$\begin{aligned}\sqrt{k}\left(\frac{AR}{k}-1\right)&=\frac{n}{k}\frac{\frac{1}{\sqrt{k}}\left(\boldsymbol{\varepsilon}'\boldsymbol{P}\boldsymbol{\varepsilon}-\frac{k}{n}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right)}{\frac{n}{k}\left(1-\frac{k}{n}\right)\sigma^2_{(n)}}+\frac{n}{k}\frac{\frac{1}{\sqrt{k}}\left(\boldsymbol{\varepsilon}'\boldsymbol{P}\boldsymbol{\varepsilon}-\frac{k}{n}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right)}{\frac{n}{k}\left(1-\frac{k}{n}\right)\sigma^2_{(n)}}\left(\frac{\frac{n}{k}\left(1-\frac{k}{n}\right)\sigma^2_{(n)}}{\frac{1}{k}\boldsymbol{\varepsilon}'\left(\boldsymbol{I}-\boldsymbol{P}\right)\boldsymbol{\varepsilon}}-1\right)\\&=\frac{E_1+E_2}{\left(1-\frac{k}{n}\right)\sigma^2_{(n)}}+o_p\left(1\right).\end{aligned}$$

Therefore, collecting the above results we obtain that

$$\left(1-\frac{k}{n}\right)\frac{\sigma^2_{(n)}}{\sqrt{W_n}}\sqrt{k}\left(\frac{AR}{k}-1\right)=\frac{E_1+E_2}{\sqrt{W_n}}+o_p\left(1\right),$$

which by (18) implies that

$$\left(1-\frac{k}{n}\right)\frac{\sigma^2_{(n)}}{\sqrt{W_n}}\sqrt{k}\left(\frac{AR}{k}-1\right)\to_d N\left(0,1\right). \tag{20}$$

Suppose that $\lim_{n\to\infty} \sigma^2_{(n)} = \sigma^2_0$ and $\lim_{n\to\infty} W_n = W_0$ exist.

$\square$

Lemma A.1. *Let* $\widehat{V}(\boldsymbol{\beta}_0) = \frac{2}{k}\boldsymbol{\varepsilon}_0^{(2)\prime}\boldsymbol{C}^{(2)}\boldsymbol{\varepsilon}_0^{(2)}$. *If Assumptions 1, 4 hold,* $\widehat{V}(\boldsymbol{\beta}_0) - V_n = O_p\left(\frac{1}{\sqrt{k}}\right)$*; consequently* $\widehat{V}(\boldsymbol{\beta}_0) - V_n \to_p 0$*, where*

$$V_n = \frac{2}{k}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}^2\sigma_i^2\sigma_j^2.$$

Proof. Let $\eta_i = \varepsilon_i^2 - \sigma_i^2$; then

$$\widehat{V}(\boldsymbol{\beta}_0) - V_n = \frac{2}{k}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}^2\left(\varepsilon_i^2\varepsilon_j^2 - \sigma_i^2\sigma_j^2\right) = \frac{2}{k}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}^2\left(\eta_i\eta_j + \sigma_i^2\eta_j + \sigma_j^2\eta_i\right).$$

So

$$\begin{aligned}\left|V_n - \widehat{V}(\boldsymbol{\beta}_0)\right| &\le \frac{2}{k}\left|\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}^2\eta_i\eta_j\right| + \frac{2}{k}\left|\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}^2\sigma_i^2\eta_j\right| + \frac{2}{k}\left|\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}^2\sigma_j^2\eta_i\right| \\ &\equiv A_1 + A_2 + A_3.\end{aligned}$$

Since

$$\mathrm{E}\left[\eta_i^2\right] = \mathrm{E}\left[\varepsilon_i^4\right] - \sigma_i^4,$$

from Assumption 4 we have $\mathrm{E}\left[\eta_i^2\right] \le c_u$. So

$$\mathrm{E}\left[A_1^2\right] = \frac{8}{k^2}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}^4 E\left[\eta_i^2\right]\mathrm{E}\left[\eta_j^2\right] \le \frac{8c_u^2}{k^2}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}^4.$$

Note that for $i \ne j$ we have

$$C_{ij} = \frac{P_{ij}}{2}\left(\frac{1}{1-P_{ii}} + \frac{1}{1-P_{jj}}\right),$$

which from Assumption 1 implies

$$|C_{ij}| = \frac{|P_{ij}|}{2}\left(\frac{1}{1-P_{ii}} + \frac{1}{1-P_{jj}}\right) \leq c_u\,|P_{ij}| \quad \text{for any } i,j, \tag{21}$$

so

$$\mathrm{E}\left[A_1^2\right] \leq \frac{8c_u^6}{k^2}\sum_{i=1}^{n}\sum_{j=1}^{n}P_{ij}^4.$$

Since $\boldsymbol{P}$ is idempotent $\boldsymbol{P}^4 = \boldsymbol{P}$, so based on the main diagonal elements we have

$$P_{hh} \geq \left(\sum_{i=1}^{n}P_{hi}^2\right)^2 \geq \sum_{i=1}^{n}P_{hi}^4 + \sum_{i=1}^{n}\sum_{j=1}^{n}P_{hi}^2P_{hj}^2,$$

so

$$\sum_{i=1}^{n}\sum_{j=1}^{n}P_{ij}^4 \leq tr\left(\boldsymbol{P}\right) = k \quad \text{and} \quad \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{h=1}^{n}P_{hi}^2P_{hj}^2 \leq k. \tag{22}$$

Therefore,

$$\mathrm{E}\left[A_1^2\right] \leq \frac{8c_u^6}{k}.$$

Now, by Cauchy-Schwarz $\left(\mathrm{E}\left[\varepsilon_i^2\right]\right)^2 \leq \mathrm{E}\left[\varepsilon_i^4\right]$, thus $\sigma_i^2 \leq \sqrt{c_u}$, so from Assumption 4, (21) and (22)

$$\begin{aligned}\mathrm{E}\left[A_2^2\right] &= \frac{4}{k^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{h=1}^{n}C_{hi}^2C_{ij}^2\sigma_h^2\sigma_j^2\,\mathrm{E}\left[\eta_i^2\right] \leq \frac{4c_u^2}{k^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{h=1}^{n}C_{hi}^2C_{ij}^2 \\ &\leq \frac{4c_u^6}{k^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{h=1}^{n}P_{hi}^2P_{ij}^2 \leq \frac{4c_u^6}{k}.\end{aligned}$$

We can obtain a similar inequality for A_3, so by the Markov and triangle inequalities we obtain that $\widehat{V}\left(\boldsymbol{\beta}_0\right) - V_n = O_p\left(\frac{1}{\sqrt{k}}\right)$, therefore, $\widehat{V}\left(\boldsymbol{\beta}_0\right) - V_n \rightarrow_p 0$. □

For given $\boldsymbol{\beta}_0$ consider

$$T_1 = \frac{1}{\sqrt{k}} \frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_0)' \boldsymbol{C} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_0)}{\sqrt{\widehat{V}(\boldsymbol{\beta}_0)}},$$

where

$$\widehat{V}(\boldsymbol{\beta}_0) = \frac{2}{k} \boldsymbol{\varepsilon}_0^{(2)\prime} \boldsymbol{C}^{(2)} \boldsymbol{\varepsilon}_0^{(2)}, \quad \boldsymbol{\varepsilon}_0 = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_0.$$

Proof of Theorem 2. Let us consider part 1. Under the null hypothesis we have

$$\begin{aligned} &\mathrm{E}[\boldsymbol{\varepsilon}' \boldsymbol{C} \boldsymbol{\varepsilon}] = 0, \\ &\mathrm{Var}[\boldsymbol{\varepsilon}' \boldsymbol{C} \boldsymbol{\varepsilon}] = E\left[(\boldsymbol{\varepsilon}' \boldsymbol{C} \boldsymbol{\varepsilon})^2\right] = 2 \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij}^2 \sigma_i^2 \sigma_j^2 \equiv k V_n. \end{aligned}$$

We use the CLT in Theorem A.1. Assumption (a) is clearly satisfied; (c) is satisfied due to Assumption 4. The last statement in (b) is satisfied because by (21) and by assumption,

$$\sum_{i=1}^{n} |C_{ij}| \leq c_u \sum_{i=1}^{n} |P_{ij}| < c_u \quad \text{for any } j.$$

Regarding (d) note that

$$\frac{1}{n} \mathrm{Var}[Q] \equiv \frac{k}{n} V_n = \frac{2}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij}^2 \sigma_i^2 \sigma_j^2 \geq \frac{2\underline{\sigma}^4}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij}^2,$$

where

$$\begin{aligned} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij}^2 &= \sum_{i \neq j} \frac{P_{ij}^2}{4} \left(\frac{1}{1 - P_{ii}} + \frac{1}{1 - P_{jj}} \right)^2 \geq \sum_{i \neq j} \frac{P_{ij}^2}{4} (1+1)^2 = \sum_{i \neq j} P_{ij}^2 \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij}^2 - \sum_{i=1}^{n} P_{ii}^2 = tr(\boldsymbol{P}) - \sum_{i=1}^{n} P_{ii}^2 = k - \sum_{i=1}^{n} P_{ii}^2. \end{aligned}$$

By Assumption 1

$$\sum_{i=1}^{n} P_{ii}^2 \leq \max P_{ii} \sum_{i=1}^{n} P_{ii} \leq (1 - 1/c_u)\, tr\,(\boldsymbol{P}) = (1 - 1/c_u)\, k. \tag{23}$$

So

$$\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij}^2 \geq k/c_u,$$

therefore,

$$\frac{1}{n} \operatorname{Var}[Q] \geq \frac{2\underline{\sigma}^4}{n} \frac{k}{c_u},$$

which is bounded away from 0 if $\lim \frac{k}{n} > 0$. In this case we can apply the CLT and complete the proof for part 1. Let us now consider part 2. Under the null we can construct the estimator

$$\widetilde{\boldsymbol{\beta}}_2 = \left(\boldsymbol{X}_2' \boldsymbol{C} \boldsymbol{X}_2\right)^{-1} \boldsymbol{X}_2' \boldsymbol{C} \left(\boldsymbol{y} - \boldsymbol{X}_1 \boldsymbol{\beta}_{10}\right) = \left(\boldsymbol{L}' \boldsymbol{X}' \boldsymbol{C} \boldsymbol{X} \boldsymbol{L}\right)^{-1} \boldsymbol{L}' \boldsymbol{X}' \boldsymbol{C} \left(\boldsymbol{y} - \boldsymbol{X}_1 \boldsymbol{\beta}_{10}\right)$$

where the fixed matrix $\boldsymbol{L}$ is such that $\boldsymbol{X}\boldsymbol{L} = \boldsymbol{X}_2$. Under the null we find

$$\widetilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 = \left(\boldsymbol{L}' \boldsymbol{X}' \boldsymbol{C} \boldsymbol{X} \boldsymbol{L}\right)^{-1} \boldsymbol{L}' \boldsymbol{X}' \boldsymbol{C} \boldsymbol{\varepsilon}.$$

Consistency follows straightforwardly from Assumptions 3, 5 and 6 or 7 and results in Theorem 1 of Bekker & Crudu (2015). □

Proof of Theorem 3. Under the null hypothesis we have

$$\begin{aligned} E\left[\boldsymbol{\varepsilon}' (\boldsymbol{P} - \boldsymbol{D})\, \boldsymbol{\varepsilon}\right] &= 0, \\ Var\left(\boldsymbol{\varepsilon}' (\boldsymbol{P} - \boldsymbol{D})\, \boldsymbol{\varepsilon}\right) &= E\left[\left(\boldsymbol{\varepsilon}' (\boldsymbol{P} - \boldsymbol{D})\, \boldsymbol{\varepsilon}\right)^2\right] = 2 \sum_{i \neq j} P_{ij}^2 \sigma_i^2 \sigma_j^2. \end{aligned}$$

We verify whether the conditions of the CLT in Theorem A.2 are met: (a) is clearly satisfied; (c) is satisfied due to Assumption 4. The statement in (b) is satisfied because of the properties of the projection matrix $\boldsymbol{P}$ and by Assumption 1. Regarding (d) note that

$$\frac{1}{n}\operatorname{Var}[Q]=\frac{2}{n}\sum_{i\neq j}P_{ij}^2\sigma_i^2\sigma_j^2\geq\frac{2\underline{\sigma}^4}{n}\sum_{i\neq j}P_{ij}^2$$

where

$$\sum_{i\neq j}P_{ij}^2=\sum_{i=1}^{n}\sum_{j=1}^{n}P_{ij}^2-\sum_{i=1}^{n}P_{ii}^2=tr\left(\boldsymbol{P}\right)-\sum_{i=1}^{n}P_{ii}^2=k-\sum_{i=1}^{n}P_{ii}^2.$$

By Assumption 1

$$\sum_{i=1}^{n}P_{ii}^2\leq\max P_{ii}\sum_{i=1}^{n}P_{ii}\leq\left(1-1/c_u\right)tr\left(\boldsymbol{P}\right)=\left(1-1/c_u\right)k. \tag{24}$$

So

$$\frac{1}{n}\operatorname{Var}[Q]\geq\frac{2\underline{\sigma}^4}{c_u}\frac{k}{n},$$

which is bounded away from 0 if $\lim\frac{k}{n}>0$. In this case we can apply the CLT and complete the proof. □

References

Anatolyev, S. & Gospodinov, N. (2011). Specification testing in models with many instruments. *Econometric Theory*, *27*, 427–441.

Anderson, T. W. & Rubin, H. (1949). Estimators of the parameters of a single equation in a complete set of stochastic equations. *The Annals of Mathematical Statistics*, *21*, 570–582.

Andrews, D. W. K., Moreira, M. J. & Stock, J. H. (2006). Optimal two-sided invariant similar tests for instrumental variable regression. *Econometrica*, *73*, 715–752.

Andrews, D. W. K. & Stock, J. (2005). Inference with weak instruments. In R. Blundell, W. K. Newey & T. Persson (Eds.), *Advances in economics and econometrics, theory and applications: Ninth world congress of the econometric society* (Vol. III). Cambridge: Cambridge University Press.

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

Angrist, J. D. & Krueger, A. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, *106*, 979–1014.

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

Bekker, P. A. & Crudu, F. (2015). Jackknife instrumental variable estimation with heteroskedasticity. *The Journal of Econometrics*, *185*, 332–342.

Bound, J., Jaeger, D. A. & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, *90*, 443–450.

Chao, J. C., Hausman, J. A., Newey, W. K., Swanson, N. R. & Woutersen, T. (2014). Testing overidentifying restrictions with many instruments and heteroskedasticity. *The Journal of Econometrics*, *178*, 15–21.

Chao, J. C., Swanson, N. R., Hausman, J. A., Newey, W. K. & Woutersen, T. (2012). Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments. *Econometric Theory*, *28*, 42–86.

Davidson, R. & MacKinnon, J. G. (1998). Graphical methods for investigating the size and power of hypothesis tests. *The Manchester School*, *66*, 1–26.

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*, 211–255.

Imbens, G. W. (2014). Instrumental variables: An econometricians perspective. *Statistical Science*, *29*, 323–358.

Keleijan, H. H. & Prucha, I. R. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *The Journal of Econometrics*, *104*, 219–257.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, *70*, 1781–1803.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica*, *71*, 1027–1048.

Moreira, M. J. (2009). Tests with correct size when instruments can be arbitrarily weak. *The Journal of Econometrics*, *152*, 131–140.

Steiger, D. & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*, 557–586.

Stock, J. H., Wright, J. H. & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics*, *20*, 518–529.

Van Hasselt, M. (2010). Many instruments asymptotic approximations under nonnormal error distributions. *Econometric Theory*, *26*, 633–645.

Wang, J. & Zivot, E. (1998). Inference on a structural parameter in instrumental variables regression with weak instruments. *Econometrica*, *66*, 1389–1404.

Zivot, E., Startz, R. & Nelson, C. R. (1998). Valid confidence intervals and inference in the presence of weak instruments. *International Economic Review*, *39*, 1119–1144.

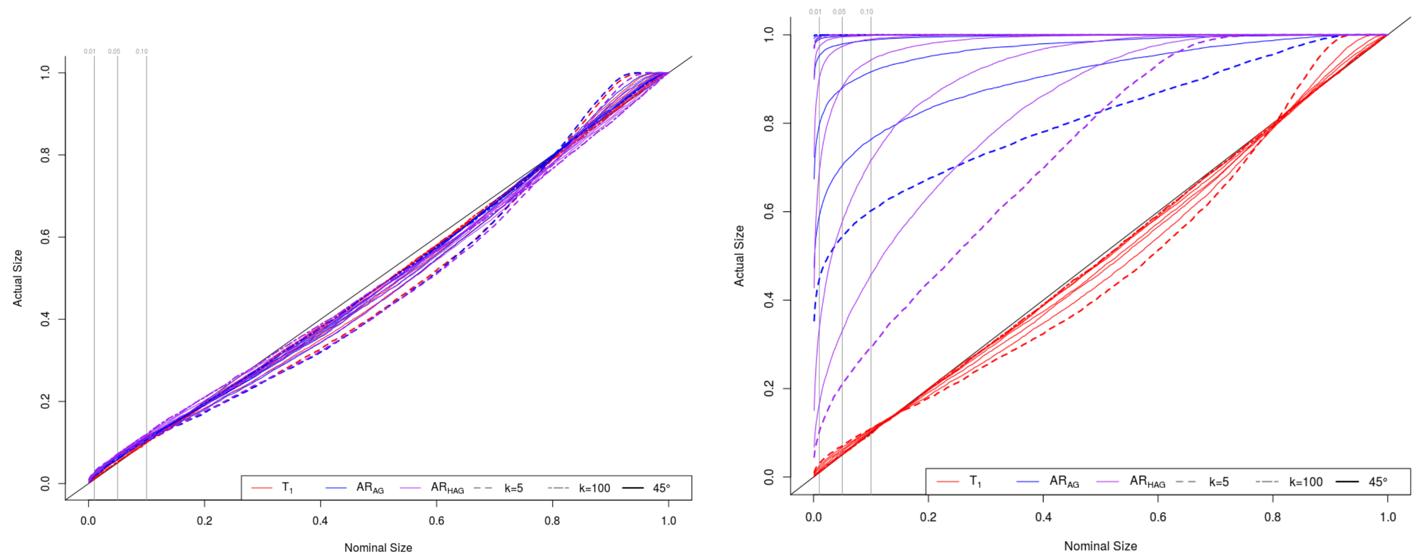

Figure 1: PP-plots of T_1, AR_{AG}, and AR_{HAG} with homoskedasticity (left) and heteroskedasticity (right), $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$.

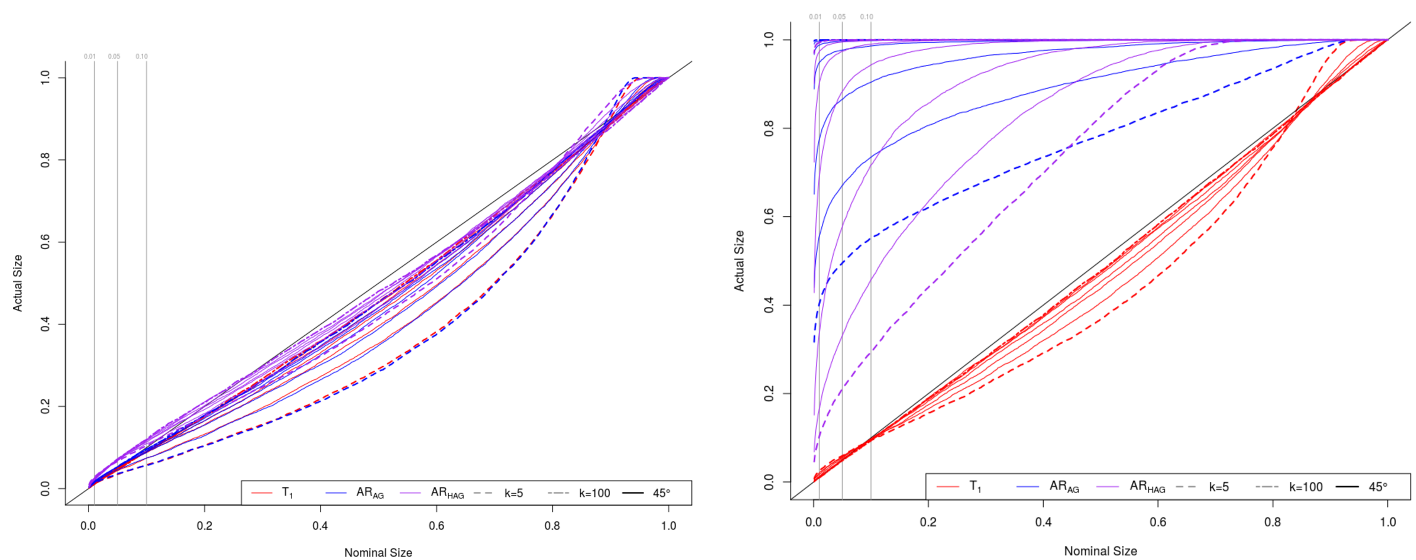

Figure 2: PP-plots of T_1 and AR_{AG} with homoskedasticity (left) and heteroskedasticity (right), $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$.

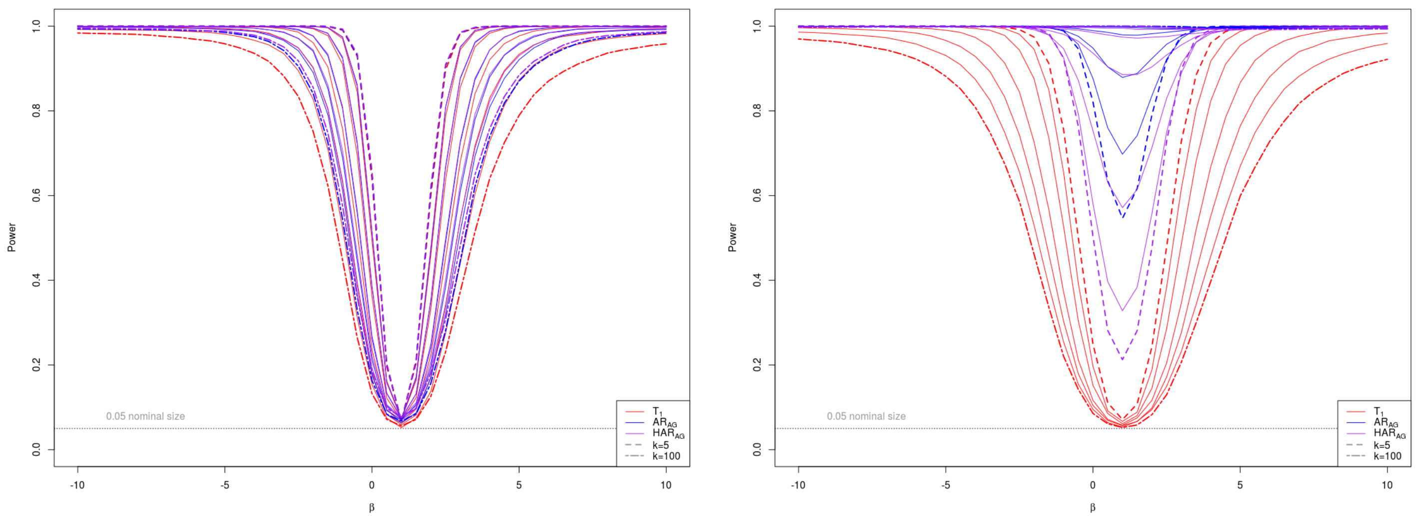

Figure 3: Power of T_1 and AR_{AG} with homoskedasticity (left) and heteroskedasticity (right), $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$.

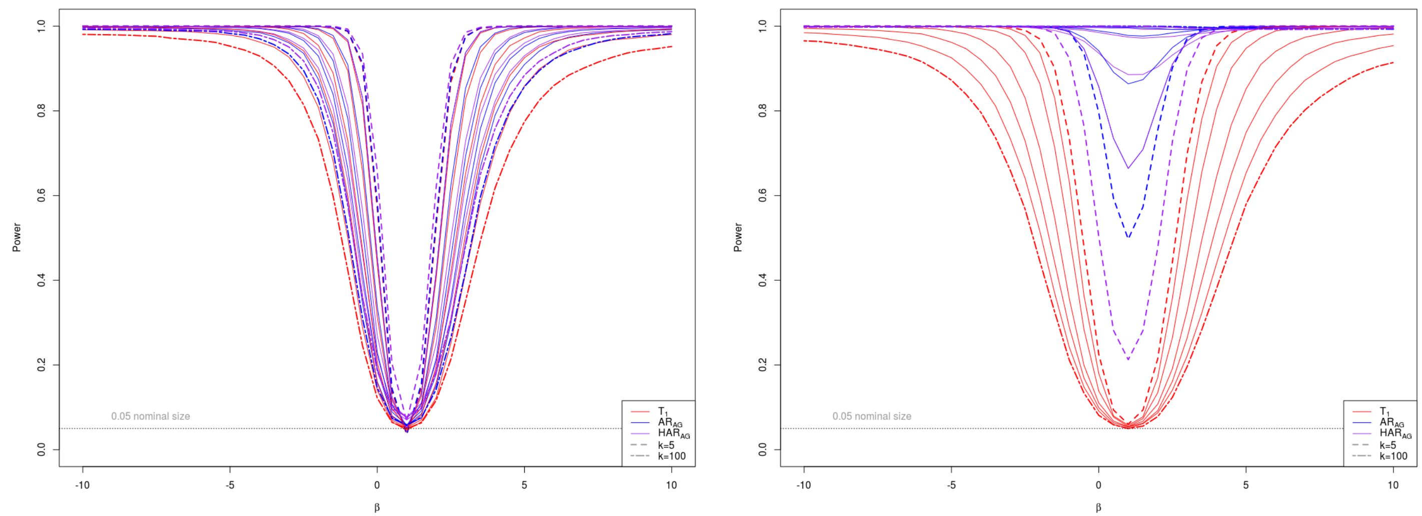

Figure 4: Power of T_1 and AR_{AG} with homoskedasticity (left) and heteroskedasticity (right), $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$.