

## A Ridge-Regularized Jackknifed Anderson-Rubin Test

Max-Sebastian Dovì, Anders Bredahl Kock & Sophocles Mavroeidis

**To cite this article:** Max-Sebastian Dovì, Anders Bredahl Kock & Sophocles Mavroeidis (2024) A Ridge-Regularized Jackknifed Anderson-Rubin Test, Journal of Business & Economic Statistics, 42:3, 1083-1094, DOI: [10.1080/07350015.2023.2290739](https://doi.org/10.1080/07350015.2023.2290739)

**To link to this article:** <https://doi.org/10.1080/07350015.2023.2290739>



© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC.



View supplementary material [↗](#)



Published online: 05 Jan 2024.



Submit your article to this journal [↗](#)



Article views: 629



View related articles [↗](#)



View Crossmark data [↗](#)

# A Ridge-Regularized Jackknifed Anderson-Rubin Test

Max-Sebastian Dovi<sup>a</sup>, Anders Bredahl Kock<sup>b</sup>, and Sophocles Mavroeidis<sup>b</sup> 

<sup>a</sup>International Monetary Fund, Washington, DC; <sup>b</sup>Department of Economics, University of Oxford, Oxford, UK

## ABSTRACT

We consider hypothesis testing in instrumental variable regression models with few included exogenous covariates but many instruments—possibly more than the number of observations. We show that a ridge-regularized version of the jackknifed Anderson and Rubin (henceforth AR) test controls asymptotic size in the presence of heteroscedasticity, and when the instruments may be arbitrarily weak. Asymptotic size control is established under weaker assumptions than those imposed for recently proposed jackknifed AR tests in the literature. Furthermore, ridge-regularization extends the scope of jackknifed AR tests to situations in which there are more instruments than observations. Monte Carlo simulations indicate that our method has favorable finite-sample size and power properties compared to recently proposed alternative approaches in the literature. An empirical application on the elasticity of substitution between immigrants and natives in the United States illustrates the usefulness of the proposed method for practitioners.

## ARTICLE HISTORY

Received September 2022  
Accepted November 2023

## KEYWORDS

High dimensional models;  
Instrumental variables; Ridge  
regression; Weak  
identification

## 1. Introduction

Instrumental variables (IVs) are commonly employed in economics and related fields to estimate causal effects from observational data. The use of a large number of IVs has gained popularity due to weak identification, where researchers aim to capture limited exogenous variation in endogenous covariates and obtain more precise inference. One example is Mendelian randomization in biology, where weakly associated genetic mutations are used as IVs, and the number of IVs can exceed the number of observations (Davies et al. 2015; Burgess and Thompson 2021). Other prominent examples are the granular IV approach (Gabaix and Koijen 2020) and the saturation approach to identify treatment effects (Blandhol et al. 2022).<sup>1</sup> However, it is crucial to have reliable inference methods that account for the potential lack of joint informativeness of a large number of weak IVs, especially when pretesting the strength of IVs is impractical or not available due to the excessive number of IVs compared to observations. This article contributes to the literature on the development of such methods.

We propose a ridge-regularized jackknifed Anderson-Rubin (RJAR) test to construct confidence sets for the coefficients of endogenous variables in weakly-identified and heteroscedastic IV models when the number of IVs is large. Jackknife-based

methods have recently been used in this context because they are applicable in an asymptotic framework where the number of IVs diverges with the number of observations. However, by relying on existing central limit theorems developed for standard projection matrices, these methods require that the number of IVs be less than the number of observations, and often perform poorly when the number of IVs is close to (but still less than) the number of observations. Recently proposed regularization approaches for inference under many IVs require strong identification or a sparse relationship between the endogenous variables and the IVs to work well. By combining jackknifing with ridge regularization, we provide a test that has the desired asymptotic size under heteroscedasticity, arbitrarily weak identification, and more IVs than observations, all while achieving good power both when the relationship between the endogenous variables and the IVs is sparse and when it is dense.

Throughout this article, we focus on Anderson and Rubin (1949, henceforth AR) tests. This is because, in addition to being robust to weak IVs, they are also robust to arbitrary relationships between the endogenous variables and the IVs, since they make no assumption on the first-stage projection of the endogenous variables on the IVs. In the context of AR tests with an increasing number of IVs, it is helpful to distinguish three different asymptotic regimes.

The first “moderately many” IVs regime allows the number of IVs,  $k_n$ , to grow with the sample size,  $n$ , but still requires it to be asymptotically negligible with respect to the sample size. Examples include Andrews and Stock (2007), who require  $k_n/n^{1/3} \rightarrow 0$ , and Phillips and Gao (2017), who require  $k_n/n \rightarrow 0$ .

The second “many” IVs regime allows the number of IVs to be of the same order of magnitude as the number of

<sup>1</sup>Recent examples of applications with many IVs include Mueller-Smith (2015), who uses judge dummies interacted with individual covariates as IVs to identify causal effects of incarceration; van Duijn and Rouwendal (2021), who derive an equation for residential sorting that produces many more IVs than observations; and Lönn and Schotman (2022), who derive structural asset-pricing models that lead to a vast number of moment conditions.

**Table 1.** Anderson and Rubin (1949) tests with many IVs: schematic comparison of main assumptions and results in the literature.

|   | $k_n$   | $Z$   | $\varepsilon$  |
|---|---|---|--|
| Kaffo and Wang (2017) and Anatolyev and Gospodinov (2011) | $k_n/n \rightarrow \tau, 0 \leq \tau < 1, n \rightarrow \infty$                             | $Z$ is fixed, obeys restrictive balanced design assumption  | iid, $\mathbb{E}[\varepsilon_i] = 0, \mathbb{E}[\varepsilon_i^4] < \infty$         |
| Belloni et al. (2012)                                     | $\log(\max(k_n, n)) = o(n^{1/3}), n \rightarrow \infty$                                     | $Z$ is fixed  | independent, $\mathbb{E}[\varepsilon_i] = 0, \mathbb{E}[\varepsilon_i^3] < \infty$ |
| Kapetanios, Khalaf, and Marcellino (2015)                 | Factor-dependent  | Potentially dependent data, $\mathbb{E}[Z_{ij}^4] < \infty, l = 1, \dots, k_n, \mathbb{E}[\varepsilon_i^4] < \infty$  |  |
| Carrasco and Tchuente (2016)                              | $n \rightarrow \infty$  | $Z_i \varepsilon_i$ is iid, $\mathbb{E}[Z_i \varepsilon_i] = 0, \mathbb{E}[X_{ig}^2] < \infty$  |  |
| Bun, Farbmacher, and Poldermans (2020)                    | $k_n/n \rightarrow \tau, 0 \leq \tau < 1, n \rightarrow \infty$                             | $Z_i \varepsilon_i$ is iid, $Z$ obeys restrictive balanced design assumption, $\mathbb{E}[Z_i \varepsilon_i] = 0, \mathbb{E}[Z_i^8 \varepsilon_i^8] < \infty$ |  |
| Crudu, Mellace, and Sándor (2020)                         | $k_n/n \rightarrow \tau, 0 \leq \tau < 1, k_n \rightarrow \infty$ as $n \rightarrow \infty$ | $Z$ is fixed, $P_{ii} \leq 1 - \delta, 0 < \delta < 1$  | independent, $\mathbb{E}[\varepsilon_i] = 0, \mathbb{E}[\varepsilon_i^4] < \infty$ |
| Mikusheva and Sun (2022)                                  | $k_n/n \rightarrow \tau, 0 \leq \tau < 1, k_n \rightarrow \infty$ as $n \rightarrow \infty$ | $Z$ is fixed, $P_{ii} \leq 1 - \delta, 0 < \delta < 1$  | independent, $\mathbb{E}[\varepsilon_i] = 0, \mathbb{E}[\varepsilon_i^6] < \infty$ |
| RJAR  | $r_n \rightarrow \infty$ as $n \rightarrow \infty$  | $Z$ is fixed, $\liminf_{n \rightarrow \infty} \frac{1}{r_n} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n})^2 > 0$   | independent, $\mathbb{E}[\varepsilon_i] = 0, \mathbb{E}[\varepsilon_i^4] < \infty$ |

NOTE:

$n$  is the number of observations,  $k_n$  is the number of IVs, and  $g$  is the number of endogenous variables.  $Z$  is the  $n \times k_n$  matrix of IVs.  $X$  is the  $n \times g$  matrix of endogenous variables.  $\varepsilon$  is the  $n \times 1$  vector of structural error terms.

$r_n := \text{rank}(Z)$ . See also (1).

$P := Z(Z'Z)^{-1}Z'$ ,  $P^{\gamma_n} := Z(Z'Z + \gamma_n I_{k_n})^{-1}Z'$  for  $\gamma_n \geq 0$  if  $r_n = k_n$  and  $\gamma_n \geq \gamma_- > 0$  if  $r_n < k_n$ .

observations. Anatolyev and Gospodinov (2011) provide an AR test that remains valid for  $k_n/n \rightarrow \tau, 0 \leq \tau < 1, n \rightarrow \infty$ , provided that the error terms are homoscedastic and the IVs satisfy a restrictive balanced-design assumption.<sup>2</sup> Bun, Farbmacher, and Poldermans (2020) provide analogous results for the GMM version of the AR statistic under the assumption of iid data. These results were recently extended in the form of a jackknifed AR statistic by Crudu, Mellace, and Sándor (2020, henceforth CMS) and Mikusheva and Sun (2022, henceforth MS) to the case where errors are allowed to display arbitrary heteroscedasticity, and the only assumption on the IVs is that the diagonal entries of the projection matrix of the IVs are bounded away from unity from above.<sup>3</sup>

The third “very many” IVs regime allows the number of IVs to grow with  $n$ , and further allows the number of IVs to be greater than the number of observations. Belloni et al. (2012, henceforth BCCH) propose a Sup Score test that remains valid under mild conditions, and allows the number of IVs to increase exponentially with the sample size. Carrasco and Tchuente (2016, henceforth CT) propose a ridge-regularized AR statistic that allows for more IVs than observations under the assumption of iid data and homoscedasticity. Kapetanios, Khalaf, and Marcellino (2015) extend Bai and Ng (2010) by proposing weak-identification robust factor-based tests that in principle allow for the number of IVs to be larger than the number of observations, provided that the factor structure of the IVs is sufficiently strong.

Using the notation of the model introduced in the next section, Table 1 provides a schematic overview of the main assumptions and results in the literature on inference that is robust to many weak IVs.

Our test provides a twofold extension of the existing literature. First, it allows for valid inference under many IVs and heteroscedastic errors while further weakening the assumptions of similar tests proposed by CMS and MS. This is made possible by deriving the limiting behavior of the RJAR statistic from the bottom up, without relying on the existing results in Chao et al. (2012) or Hansen and Kozbur (2014). Second, this test allows for more IVs than observations under arbitrary heteroscedasticity of the error terms. The only other approach currently available in the literature that is robust to heteroscedastic error terms and more IVs than observations is the Sup Score test of BCCH. Simulations show that the RJAR test has power comparable to the Sup Score test of BCCH whenever the signal in the first stage is sparse (i.e., only a few of the IVs are informative), and substantially more power when the signal in the first stage is dense (i.e., not sparse). Simulations also show that the RJAR test is as powerful as, and in some cases more powerful than, the AR test of CT, the validity of which has been established only under homoscedastic error terms. Finally, we provide a comparison of the most recently proposed approaches to conducting inference in possibly heteroscedastic linear IV models when the number of IVs is not negligible with respect to the sample size. Indeed, using extensive simulation evidence and an empirical application based on Card (2009), we provide a comparison between these existing “state-of-the-art” methods in a controlled and comparable setting.

In the rest of the article, we use the following notation.  $I_p$  denotes the  $p \times p$  identity matrix. For an  $m \times 1$  vector  $a = [a_1, \dots, a_m]'$ ,  $\|a\|_2 := \sqrt{\sum_{j=1}^m a_j^2}$ . The entry  $(i, j)$  for an  $m \times p$  matrix  $A$  is denoted as  $A_{ij}$  for  $1 \leq i \leq m, 1 \leq j \leq p$ .  $\text{tr}(A) := \sum_{i=1}^m A_{ii}$  for any  $m \times m$  matrix  $A$  with entries given by  $A_{ij}$  for  $1 \leq i, j \leq m$ , and  $\|A\|_F := \sqrt{\text{tr}(A'A)}$  denotes the

<sup>2</sup>See Anatolyev and Yaskov (2017) and Crudu, Mellace, and Sándor (2020) for a discussion of the restrictiveness of the balanced-design assumption of Anatolyev and Gospodinov (2011). See also Kaffo and Wang (2017) for a bootstrapped version of the Anatolyev and Gospodinov (2011) AR test.

<sup>3</sup>Boot and Ligtenberg (2023), Matsushita and Otsu (2022), and Lim, Wang, and Zhang (2024) propose weak-identification robust jackknife-based procedures for inference with many IVs. However, these methods make additional assumptions on the first-stage projection of the endogenous variables on the IVs, and hence fall outside of the scope of this article.

Frobenius norm of  $A$ . The remaining notation follows standard conventions.

The structure of the article is as follows. Section 2 specifies the linear IV model considered throughout. Section 3 introduces the RJAR test, and provides the main asymptotic results. Section 4 provides simulation evidence on the size and power of our RJAR test and compares it with the tests proposed by BCCH, CT, CMS, and MS. Section 5 considers an empirical application based on Card (2009). Section 6 concludes. All proofs are given in the supplementary material.

## 2. Model

We consider the heteroscedastic linear IV model given by

$$y = X\beta + \varepsilon \quad (1a)$$

$$X = Z\Pi + V, \quad (1b)$$

where  $y$  is an  $n \times 1$  vector containing the dependent variable,  $X$  is an  $n \times g$  matrix containing the endogenous variables,  $\beta$  is a  $g \times 1$  coefficient vector,  $\varepsilon$  is an  $n \times 1$  vector containing the structural error terms,  $Z$  is an  $n \times k_n$  matrix containing the IVs,  $\Pi$  is a  $k_n \times g$  coefficient matrix, and  $V$  is an  $n \times g$  matrix containing the first-stage errors.  $k_n$  can diverge with  $n$ , but  $g$  is fixed. Also, let  $y_i$ ,  $X_i$ ,  $\varepsilon_i$ ,  $Z_i$ , and  $V_i$  denote the  $i$ th row of  $y$ ,  $X$ ,  $\varepsilon$ ,  $Z$ , and  $V$ , respectively. As in BCCH, CMS, and MS, we treat  $Z$  as fixed (non-stochastic),<sup>4</sup> and assume that any exogenous covariates have been partialled out (see the discussion of Assumption 3). We exclusively consider methods that allow for arbitrarily weak identification, that is, methods that control asymptotic size irrespective of the value of  $\Pi$ .

Inference is conducted on the coefficient vector  $\beta$  by testing hypotheses of the following type for a prespecified  $\beta_0 \in \mathbb{R}^g$ :

$$H_0 : \beta = \beta_0 \text{ versus } H_1 : \beta \neq \beta_0. \quad (2)$$

For a given non-randomized test of asymptotic size  $\alpha \in (0, 1)$ , a confidence set of asymptotic coverage  $1 - \alpha$  can be constructed as the collection of those  $\beta_0$  for which  $H_0$  in (2) is not rejected. For convenience, define

$$\begin{aligned} e(\beta_0) &:= [e_1(\beta_0), \dots, e_n(\beta_0)]', \\ e_i(\beta_0) &:= y_i - X_i\beta_0, \quad i = 1, \dots, n, \end{aligned} \quad (3)$$

which we refer to as the structural error under the null hypothesis.

It will become apparent below that so long as the error term,  $\varepsilon$ , remains additive, the linearity in the structural equation (1a) and the first-stage equation (1b) can be relaxed to allow for any (known) real-valued function, without affecting the asymptotic size of the RJAR test.

## 3. The RJAR Test

This section introduces our proposed RJAR test, derives its large sample properties under the null hypothesis in (2), and discusses

its relationship to the most closely related tests in the literature: the ridge-regularized AR test of CT, the jackknifed AR tests of CMS and MS, and the Sup Score test of BCCH.

### 3.1. Definition of the RJAR Test

The original AR test of the null hypothesis in (2) can be thought of as a test that the IVs are exogenous using the implied structural errors in (3). More specifically, the AR test is a Wald test of the significance of the IVs in the auxiliary regression of the structural errors under the null,  $e_i(\beta_0)$ , on the IVs,  $Z_i$ .<sup>5</sup> The weak-IV-robust AR test of the null hypothesis in (2) with asymptotic size  $\alpha$  rejects the null if and only if the AR statistic exceeds the  $1 - \alpha$  quantile of a  $\chi^2$  distribution with  $k_n$  degrees of freedom. When the number of IVs  $k_n$  grows with  $n$ , that is, when there are many (potentially weak) IVs, the  $\chi^2$  approximation of the original AR statistic breaks down. The recently proposed AR test of CT, which we shall briefly review in Section 3.3, uses ridge-regularization to allow for a growing number of IVs that may even exceed the sample size, but is not robust to arbitrary heteroscedasticity in the error terms. The recent papers by CMS and MS, which we shall also briefly review in Section 3.3, propose jackknifed versions of the AR test that remain valid when  $k_n$  grows with  $n$ , but require  $k_n < n$ . Our proposed method combines ridge regularization with jackknifing of the auxiliary regression of  $e_i(\beta_0)$  on  $Z_i$  to allow for a weak-identification robust test capable of dealing with both more IVs than observations and arbitrary heteroscedasticity in the error terms.

We first standardize the IVs in-sample (after partialling out any covariates) as in BCCH (p. 2393) so that

$$\frac{1}{n} \sum_{i=1}^n Z_{ij}^2 = 1 \quad \text{for } j = 1, \dots, k_n. \quad (4)$$

The RJAR test for the testing problem in (2) is then based on the following statistic:

$$RJAR_{\gamma_n}(\beta_0) := \frac{1}{\sqrt{r_n} \sqrt{\hat{\Phi}_{\gamma_n}(\beta_0)}} \sum_{i=1}^n \sum_{j \neq i} P_{ij}^{\gamma_n} e_i(\beta_0) e_j(\beta_0), \quad (5)$$

where  $r_n := \text{rank}(Z)$  (assumed to be positive without loss of generality),

$$\hat{\Phi}_{\gamma_n}(\beta_0) := \frac{2}{r_n} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n})^2 e_i^2(\beta_0) e_j^2(\beta_0), \quad (6)$$

and  $P^{\gamma_n} := Z(Z'Z + \gamma_n I_{k_n})^{-1}Z'$  is the ridge-regularized projection matrix with  $\gamma_n \geq 0$  if  $r_n = k_n$  and  $\gamma_n > 0$  if  $r_n < k_n$ .  $\gamma_n$  is a (sequence of) regularization parameter(s) whose choice we discuss below. We note that setting  $\gamma_n = 0$  when  $r_n = k_n$  corresponds to the case where a standard least-squares projection matrix  $Z(Z'Z)^{-1}Z'$  is used to construct the RJAR statistic. This is the standard jackknife statistic. We also note that the  $r_n$ -scaling in the denominator of the RJAR statistic cancels with the  $r_n$  in the denominator of  $\hat{\Phi}_{\gamma_n}(\beta_0)$ .

<sup>4</sup>The treatment of IVs as non-stochastic is an assumption that has been commonly used in several of the papers that are most closely related to ours, see Table 1. With stochastic IVs, one can interpret the results as holding conditional on  $Z$ , see also CMS Footnote 1.

<sup>5</sup>The AR statistic is defined as  $e(\beta_0)'Z\hat{\Sigma}^{-1}Z'e(\beta_0)/n$ , where  $\hat{\Sigma}$  is a consistent estimator of  $\text{var}[Z'e(\beta_0)/\sqrt{n}]$ .

We set the regularization parameter  $\gamma_n$  to

$$\gamma_n^* := \max \arg \max_{\gamma_n \in \Gamma_n} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n})^2, \quad (7)$$

where  $\Gamma_n := \{\gamma_n \in \mathbb{R} : \gamma_n \geq 0 \text{ if } r_n = k_n, \text{ and } \gamma_n \geq \gamma_- > 0 \text{ if } r_n < k_n\}$  for some constant  $\gamma_-$  not depending on  $n$ . The existence of  $\gamma_n^*$  is shown in the supplementary material. We let  $\gamma_n^*$  be an element of  $\arg \max_{\gamma_n \in \Gamma_n} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n})^2$  out of conservativeness, that is, to make [Assumption 3](#) as plausible as possible given the IVs. Furthermore, we let  $\gamma_n^*$  be the maximal element of this set because the maximizer is not necessarily unique without imposing additional assumptions on the singular values and left-singular vectors of the IVs (although in practice we only found unique maximizers). We choose to take the maximum of the maximizers to make the smallest eigenvalues of the ridge-regularized Gram matrix,  $Z'Z + \gamma_n I_{k_n}$ , as far away from zero as possible when  $r_n < k_n$ . To see that  $\gamma_n^*$  maximizes the smallest eigenvalue of  $Z'Z + \gamma_n I_{k_n}$  among the elements of  $\arg \max_{\gamma_n \in \Gamma_n} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n})^2$ , notice that due to the symmetry of  $Z'Z + \gamma_n I_{k_n}$ , its smallest eigenvalue can be expressed as

$$\min_{a \in \mathbb{R}^{k_n} : \|a\|_2=1} a'(Z'Z + \gamma_n I_{k_n})a = \min_{a \in \mathbb{R}^{k_n} : \|a\|_2=1} a'(Z'Z)a + \gamma_n = \gamma_n,$$

where the first equality used that  $a'a = 1$ , and the second equality used that the smallest eigenvalue of  $Z'Z$  is 0 when the rank  $r_n$  of  $Z$  is less than  $k_n$ .

We now have all the ingredients to define our new test.

**Definition 3.1.** The RJAR test rejects  $H_0 : \beta = \beta_0$  in (2) at significance level  $\alpha \in (0, 1)$  if and only if

$$RJAR_{\gamma_n^*}(\beta_0) > Q(1 - \alpha), \quad (8)$$

where  $RJAR_{\gamma_n^*}(\beta_0)$  is defined in (5),  $\gamma_n^*$  is defined in (7), and  $Q(1 - \alpha)$  is the  $(1 - \alpha)$  quantile of the Standard Normal distribution.

### 3.2. Asymptotic Properties of the RJAR Test

We make the following assumptions to derive the limiting distribution of  $RJAR_{\gamma_n^*}(\beta_0)$  under the null hypothesis in (2).

**Assumption 1.**  $\{\varepsilon_i\}_{i \in \mathbb{N}}$  is a sequence of independent random variables satisfying  $\mathbb{E}[\varepsilon_i] = 0$ ,  $\inf_{i \in \mathbb{N}} \text{var}[\varepsilon_i] > 0$ , and  $\sup_{i \in \mathbb{N}} \mathbb{E}[\varepsilon_i^4] < \infty$ .

**Assumption 2.**  $r_n = \text{rank}(Z) \rightarrow \infty$  as  $n \rightarrow \infty$ .

**Assumption 3.** 1. If  $r_n = k_n$ , then there exists a  $\gamma_n \in [0, \infty)$  such that  $\liminf_{n \rightarrow \infty} \frac{1}{r_n} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n})^2 > 0$ .  
2. If  $r_n < k_n$ , then there exists a  $\gamma_- > 0$  and a  $\gamma_n \in [\gamma_-, \infty)$  such that  $\liminf_{n \rightarrow \infty} \frac{1}{r_n} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n})^2 > 0$ .

**Assumption 1** is a mild condition on the structural error terms, and allows for conditional heteroscedasticity. It is the

same as in CMS. It is slightly less restrictive than the one in MS (who require finite sixth moments on the structural error terms), and slightly more restrictive than the one in BCCH (who require finite third moments).<sup>6</sup>

**Assumption 2** is a weak technical assumption. It implies that both  $k_n$  and  $n$  diverge. It also allows for the sum of the number of IVs and the number of exogenous covariates to be larger than the number of observations, provided that the number of exogenous covariates that have been partialled out be sufficiently small (so that the rank of the matrix of partialled IVs continues to diverge).<sup>7</sup> This assumption is weaker than the restriction on the dimensionality in MS and CMS, who require  $r_n = k_n$ , and  $k_n < n$  for each  $n \in \mathbb{N}$ , as well as  $k_n \rightarrow \infty$  as  $n \rightarrow \infty$ . BCCH prove asymptotic size control of their Sup Score test under the assumption that  $\log k_n = o(n^{1/3})$ .

#### 3.2.1. Assumption 3 when $r_n = k_n$

**Assumption 3** is a high-level assumption on the number of IVs and their correlation structure. **Assumption 3** implies that  $\gamma_n^*$  satisfies  $\liminf_{n \rightarrow \infty} \frac{1}{r_n} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n^*})^2 > 0$ . In the absence of exogenous covariates and when  $k_n < n$  (as in CMS and MS), **Assumption 3** is weaker than the *balanced-design* assumption in CMS and MS which requires for  $P := Z(Z'Z)^{-1}Z'$  that

$$\max_{1 \leq i \leq n} P_{ii} \leq 1 - \delta \quad \text{for all } n \in \mathbb{N}, \quad (9)$$

for some  $0 < \delta < 1$ . This is stated formally in part 1 of the following Proposition.

**Proposition 3.1.** Suppose (as in CMS and MS) that  $\text{rank}(P) = k_n$ .

1. If there exists a  $\delta \in (0, 1)$  such that  $\max_{1 \leq i \leq n} P_{ii} \leq 1 - \delta$  for all  $n \in \mathbb{N}$  then **Assumption 3** is satisfied.
2. For any  $\delta \in (0, 1)$ , let  $\mathcal{A}_n(\delta) := \{i \in \{1, \dots, n\} : P_{ii} \geq 1 - \delta\}$ . If there exists a  $\delta' \in (0, 1)$  such that  $\frac{1}{k_n} |\mathcal{A}_n(\delta')| \rightarrow 0$  then **Assumption 3** is satisfied.

Part 2. of **Proposition 3.1** implies that **Assumption 3** can be satisfied—even when the balanced-design assumption is not—as long as the number of diagonal elements of  $P$  “close to 1” increases slower than  $k_n$ . Note also that if for some  $\delta' \in (0, 1)$  it holds that  $\max_{1 \leq i \leq n} P_{ii} \leq 1 - \delta'$  for all  $n \in \mathbb{N}$ , then  $\mathcal{A}_n(\delta') = \emptyset$ .

Thus, part 1. of **Proposition 3.1** actually follows from Part 2. However, we have chosen to keep part 1. as a separate statement to explicitly state that the balanced-design assumption implies **Assumption 3**.

#### 3.2.2. Assumption 3 When $r_n < k_n$

When the number of instruments exceeds the sample size, that is when  $k_n > n$ , one has that  $r_n \leq n$ . The following proposition

<sup>6</sup>It is difficult to allow for autocorrelated errors in our framework, since our proof relies on casting the RJAR statistic as a degenerate U-statistic, which requires that  $\varepsilon_i$  be independent of  $\varepsilon_j$  for  $i \neq j$ .

<sup>7</sup>More formally, suppose that  $W$  is an  $n \times l_n$  matrix of covariates with rank  $l_n$ , and define  $M_W := I_n - W(W'W)^{-1}W'$ . Furthermore, let  $\tilde{Z}$  be the  $n \times k_n$  matrix of IVs prior to partialling out the controls  $W$ , and let  $\tilde{r}_n := \text{rank}(\tilde{Z})$ , such that  $Z = M_W \tilde{Z}$  and  $r_n := \text{rank}(Z) \leq \min(n - l_n, \tilde{r}_n)$ . Therefore,  $l_n$  can diverge as  $n \rightarrow \infty$  without contradicting  $r_n \rightarrow \infty$ , which means that a growing number of exogenous controls can be accommodated.



provides sufficient conditions for [Assumption 3](#) to be satisfied with probability one in this situation in case  $Z$  is random.

**Proposition 3.2.** Let the entries of  $Z$  be iid with a mean zero and variance one. In addition, let  $p > 2$  satisfy  $\mathbb{E}|Z_{11}|^{4p} < \infty$  and let  $k_n = \tau n$  for  $\tau \in [1, \infty)$ . Then, if the distribution of  $Z_{11}$  is absolutely continuous with respect to the Lebesgue measure, there exists a constant  $\eta > 0$  such that for  $\gamma_{n,\eta} = \eta n$  it holds with probability one that

$$\liminf_{n \rightarrow \infty} \frac{1}{r_n} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_{n,\eta}})^2 > 0.$$

Note that [Assumption 3](#) is satisfied in particular when the entries of  $Z$  are iid Gaussian. The latter was used in Hansen and Kozbur (2014) to provide a set of sufficient conditions for their assumptions to be satisfied. The proof of [Proposition 3.2](#) also sheds further light on the exact value of  $\eta$ . To summarize, [Propositions 3.1](#) and [3.2](#) show that [Assumption 3](#) is likely to be satisfied; this can be the case even in settings in which the assumptions underlying previous methods are violated.

If very small values of  $\frac{1}{r_n} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n})^2$  are observed, then [Assumption 3](#) may be questionable. If this is the case, practitioners could make [Assumption 3](#) more plausible by suitably modifying the matrix of IVs,  $Z$ . This could include the dropping of IVs (either randomly or motivated by economic reasoning, but not informed by the in-sample correlation of the IVs with the endogenous variable), or the (iterative) removal of certain observations. If sufficient data is available to split the sample, a data-driven selection of IVs on one split of the sample for use on the other is also possible.

### 3.2.3. Asymptotic Normality of the RJAR Test Statistic

We are now in a position to state the asymptotic distribution of the RJAR test under the null hypothesis.

**Theorem 3.1.** Suppose [Assumptions 1–3](#) and the null hypothesis in (2) hold. Consider any sequence  $\gamma_n$  such that  $\liminf_{n \rightarrow \infty} \frac{1}{r_n} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n})^2 > 0$ . Then, the statistic  $RJAR_{\gamma_n}(\beta_0)$  defined in (5) satisfies

$$RJAR_{\gamma_n}(\beta_0) \xrightarrow{d} \mathcal{N}[0, 1].$$

**Corollary 3.1.** Under [Assumptions 1–3](#) and the null hypothesis in (2), the RJAR test given in [Definition 3.1](#) (i.e., using  $\gamma_n = \gamma_n^*$  for all  $n \in \mathbb{N}$ ) has asymptotic size  $\alpha$ .

Notice that we did not impose any assumption on the coefficients of the instruments  $\Pi$  in the first-stage regression in (1). Therefore, the RJAR test is robust to arbitrarily weak identification.

### 3.3. Closest Alternatives in the Literature

The RJAR test combines two existing approaches in the literature. It uses ridge-regularization as in CT to allow for  $r_n < k_n$ , and jackknifing as in CMS and MS to allow for arbitrary heteroscedasticity in the error terms.

The RJAR test is similar to the ridge-regularized AR test proposed by CT given by

$$AR_{CT} = \frac{ne(\beta_0)' P^\theta e(\beta_0)}{e(\beta_0)' (I_n - P^\theta) e(\beta_0)},$$

where  $P^\theta = Z(Z'Z + \theta I_k)^{-1}$  for some fixed scalar  $\theta$  that does not depend on  $n$ , where  $\theta \geq 0$  if  $r_n = k_n$ , and  $\theta > 0$  if  $r_n < k_n$ . CT show that under the assumption of homoscedastic error terms,  $AR_{CT}$  converges to an infinite sum of weighted  $\chi_1^2$  distributions. Since the limiting distribution depends on an infinite number of unobserved weights, CT propose a bootstrap procedure to derive the critical values for  $AR_{CT}$ . The distinguishing features of the RJAR test are the data-driven and unique penalty parameter used to regularize the projection matrix (we find substantial sensitivity of the performance of  $AR_{CT}$  to different values of  $\theta$ ), the robustness to arbitrary heteroscedasticity in the error terms, and its computational speed (it uses standard Normal asymptotic critical values instead of the bootstrap).

Robustness to arbitrary heteroscedasticity in the error terms is achieved through jackknifing, as in CMS and MS. The distinguishing feature of the RJAR test to the tests proposed in CMS and MS is the use of a ridge-regularized projection matrix  $P^{\gamma_n}$ , which makes the RJAR test applicable also when  $k_n > r_n$ , unlike the aforementioned two tests that use the standard least-squares projection matrix  $Z'(Z'Z)^{-1}Z'$ , and cannot be computed when  $k_n > r_n$ .

CMS assume  $r_n = k_n < n$ , and propose the jackknifed AR statistic given by

$$AR_{CMS}(\beta_0) := \frac{1}{\sqrt{k_n} \sqrt{\hat{\Phi}_{CMS}(\beta_0)}} \sum_{i=1}^n \sum_{j \neq i} C_{ij} e_i(\beta_0) e_j(\beta_0),$$

where  $C := A - B$ ,  $A := P + \Delta$ ,  $B := (I_n - P)D(I_n - D)^{-1}(I_n - P)$ ,  $\Delta := PD(I_n - D)^{-1}P - \frac{1}{2}PD(I_n - D)^{-1} - \frac{1}{2}D(I_n - D)^{-1}P$ , and  $D$  is the diagonal matrix containing the diagonal elements of  $P$ .  $\hat{\Phi}_{CMS}(\beta_0) := \frac{2}{k_n} \sum_{i=1}^n \sum_{j \neq i} C_{ij}^2 e_i^2(\beta_0) e_j^2(\beta_0)$ . Under the null hypothesis in (2), CMS show that  $AR_{CMS}(\beta_0)$  converges to a Standard Normal distribution. The CLT underlying this result is a modified version of Lemma A2 of Chao et al. (2012) proposed in Bekker and Cruadu (2015, Appendix A.4).

MS also assume that  $r_n = k_n < n$ , and propose a different jackknifed AR statistic than CMS that can be obtained from  $RJAR_{\gamma_n}(\beta_0)$  in (5) by setting  $\gamma_n = 0$ , and replacing  $\hat{\Phi}_{\gamma_n}(\beta_0)$  with

$$\hat{\Phi}_{MS}(\beta_0) := \frac{2}{k_n} \sum_{i=1}^n \sum_{j \neq i} \frac{P_{ij}^2}{M_{ii}M_{jj} + M_{ij}^2} [e_i(\beta_0)M_{ii}e(\beta_0)] [e_j(\beta_0)M_{jj}e(\beta_0)], \quad (10)$$

where  $M = I_n - P$ , and  $M_i$  is the  $i$ th row of  $M$ . The reason why the unregularized jackknifed AR test in MS uses  $\hat{\Phi}_{MS}(\beta_0)$  instead of the variance estimator given in (6) evaluated at  $\gamma_n = 0$ , is because, according to their Theorem 4, and the discussion in Section 4.2.1, the former yields higher power than the latter. It follows that the unregularized version of our RJAR test, which arises when  $\gamma_n^* = 0$  in (7), will be dominated by the jackknifed AR test of MS in terms of power, and practitioners may prefer the latter when  $r_n = k_n$  and the balanced-design assumption is satisfied. This can have implications for the finite-sample

performance of the jackknifed AR test of MS compared to the RJAR, which we investigate in [Section 4](#).

The only other existing test that allows for  $r_n < k_n$  and arbitrary heteroscedasticity is the Sup Score test of BCCH. BCCH first standardize the IVs as in (4), and then propose the Sup Score statistic given by

$$S(\beta_0) = \max_{1 \leq j \leq k_n} \frac{\left| \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i(\beta_0) Z_{ij} \right|}{\sqrt{\frac{1}{n} \sum_{i=1}^n (e_i(\beta_0))^2 Z_{ij}^2}}.$$

BCCH propose to use the critical value  $v_\alpha = c_{BCCH} \times \mathcal{Q}(1 - \alpha/(2k_n))$ ,  $c_{BCCH} > 1$  and  $\alpha \in (0, 1)$ . They show that comparing their Sup Score statistic to this critical value yields a test of the null hypothesis in (2) that has asymptotic size less than or equal to  $\alpha$ . Being a supremum-norm test suggests that the BCCH Sup Score test will work well with a sparse first stage (i.e., where only a few elements of  $\Pi$  are zero), but may have lower power than the RJAR test when the first stage is dense. This is verified in the simulations in [Section 4](#).

#### 4. Simulations

We now investigate the size and power properties of the RJAR test and compare them to those of the tests proposed in CT, CMS, MS, and BCCH. We take our simulation setup from Hansen and Kozbur (2014) who in turn take theirs from BCCH. The DGP is given by

$$y_i = X_i \beta + \varepsilon_i \quad (11a)$$

$$X_i = Z_i' \pi + v_i, \quad (11b)$$

for  $i = 1, \dots, n = 100$ . The IVs  $Z_i$  are independent and identically Gaussian with mean 0 and  $\text{var}[Z_{il}] = 0.3$  and  $\text{corr}[Z_{il}, Z_{im}] = 0.5^{|l-m|}$ .<sup>8</sup> The error terms are given by

$$\begin{bmatrix} \varepsilon_i \\ v_i \end{bmatrix} \sim \mathcal{N} \left[ 0, \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & \sigma_v^2 \end{bmatrix} \right]$$

with  $\sigma_\varepsilon^2 = 2$ ,  $\sigma_v^2 = 1$  and  $\sigma_{\varepsilon v} = 0.6\sigma_\varepsilon\sigma_v$ .

$\pi = \zeta \kappa$ , where  $\kappa$  is a vector of zeros and ones that varies with the type of DGP considered (sparse or dense, as modeled below), and  $\zeta$  is some scalar that ensures that for a given concentration parameter,  $\mu^2$ , the following relationship is satisfied:

$$\mu^2 = \frac{n\pi' \mathbb{E}[Z_i Z_i'] \pi}{\sigma_v^2}.$$

This implies that

$$\zeta = \sqrt{\frac{\sigma_v^2 \mu^2}{n\kappa' \mathbb{E}[Z_i Z_i'] \kappa}}.$$

To illustrate how the sparsity structure of the first stage in (11b) can affect the size and power of the studied tests, we consider both a sparse first stage and a dense first stage. Sparsity in the first stage is modeled by setting  $\kappa = [\iota_5', 0_{k_n-5}']'$ , where  $\iota_q$  is

a  $q \times 1$  vector of ones, and  $0_q$  is a  $q \times 1$  vector of zeros. Denseness in the first stage is modeled by setting  $\kappa = [\iota_{0.4k_n}', 0_{0.6k_n}']'$ .

We consider  $k_n = 30, 90, 190$ . In the context of [Assumption 3](#), we search over values greater than 1 when choosing  $\gamma_n^*$  in case  $r_n < k_n$ . For the case of 30 IVs, the RJAR test does not impose any regularization ( $\gamma_n^* = 0$ ). For the case of 90 IVs,  $\gamma_n^* = 12.048$ . We note that in the latter case  $\sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n^*})^2 = 11.123 > 8.845 = \sum_{i=1}^n \sum_{j \neq i} (P_{ij})^2$ . This shows that even in the case where  $r_n < n$ , ridge regularization can make [Assumption 3](#) strictly more plausible. For the case of 190 IVs,  $\gamma_n^* = 109.187$ .

The variance estimator of MS occasionally yields a negative value. These cases are conservatively interpreted as a failure to reject the null hypothesis. As recommended by BCCH,  $c_{BCCH} = 1.1$ . As in the simulation section in CT, we set  $\theta = 0.05$ . The number of Monte Carlo replications is 10,000.

The supplementary material provides additional simulation evidence on inference with heteroscedastic error terms. The supplementary material provides additional simulation evidence on projection-based inference.

##### 4.1. Size

[Figure 1](#) shows the simulation results with a sparse first stage for tests of size 0.01 to 0.99, that is the rejection frequency under  $H_0 : \beta_0 = 1$ . As far as the illustration of the tests' size properties is concerned, the dense first stage yields virtually the same results, and is hence omitted. Since all tests are robust to weak IVs, the rejection frequencies of the tests are not affected by the strength of identification. For the case of 30 IVs, the AR tests of CT, CMS, and MS and the RJAR test have correct size, while the BCCH Sup Score test is undersized.

For the case of 90 IVs, the AR test of CT and the RJAR test control size. The AR test of CMS appears to control size for common small nominal test sizes (e.g., 0.05 or 0.1). The AR test of MS appears to be generally oversized. For example, at nominal level 0.05, the rejection frequency of the test is 0.189.<sup>9</sup> The BCCH Sup Score test continues to be undersized.

For the case of 190 IVs, only the AR test of CT, the BCCH Sup Score test, and the RJAR test are feasible. As before, the BCCH Sup Score test is undersized, while the AR test of CT and the RJAR test have correct size.

##### 4.2. Power

[Figures 2–4](#) show the power of the tests when the number of IVs and the sparsity pattern of the first stage is varied. It is still the case that  $H_0 : \beta_0 = 1$ .

For the case of 30 IVs ([Figure 2](#)), the AR tests of CT and CMS have similar power to the RJAR test, while MS is slightly more powerful than the RJAR test. The BCCH Sup Score test is less powerful than all other tests.

For the case of 90 sparse IVs ([Figure 3](#)), the RJAR test is slightly more powerful than the BCCH Sup Score test. The AR test of MS fails to control the size, while the AR tests of CT and

<sup>8</sup>Simulations in the supplementary material show that the results remain qualitatively unchanged if uncorrelated Gaussian IVs are considered instead.

<sup>9</sup>Simulations in the supplementary material show that if  $n$  and  $k_n$  are increased to about 1000 and 900, respectively, there is virtually no more size distortion.

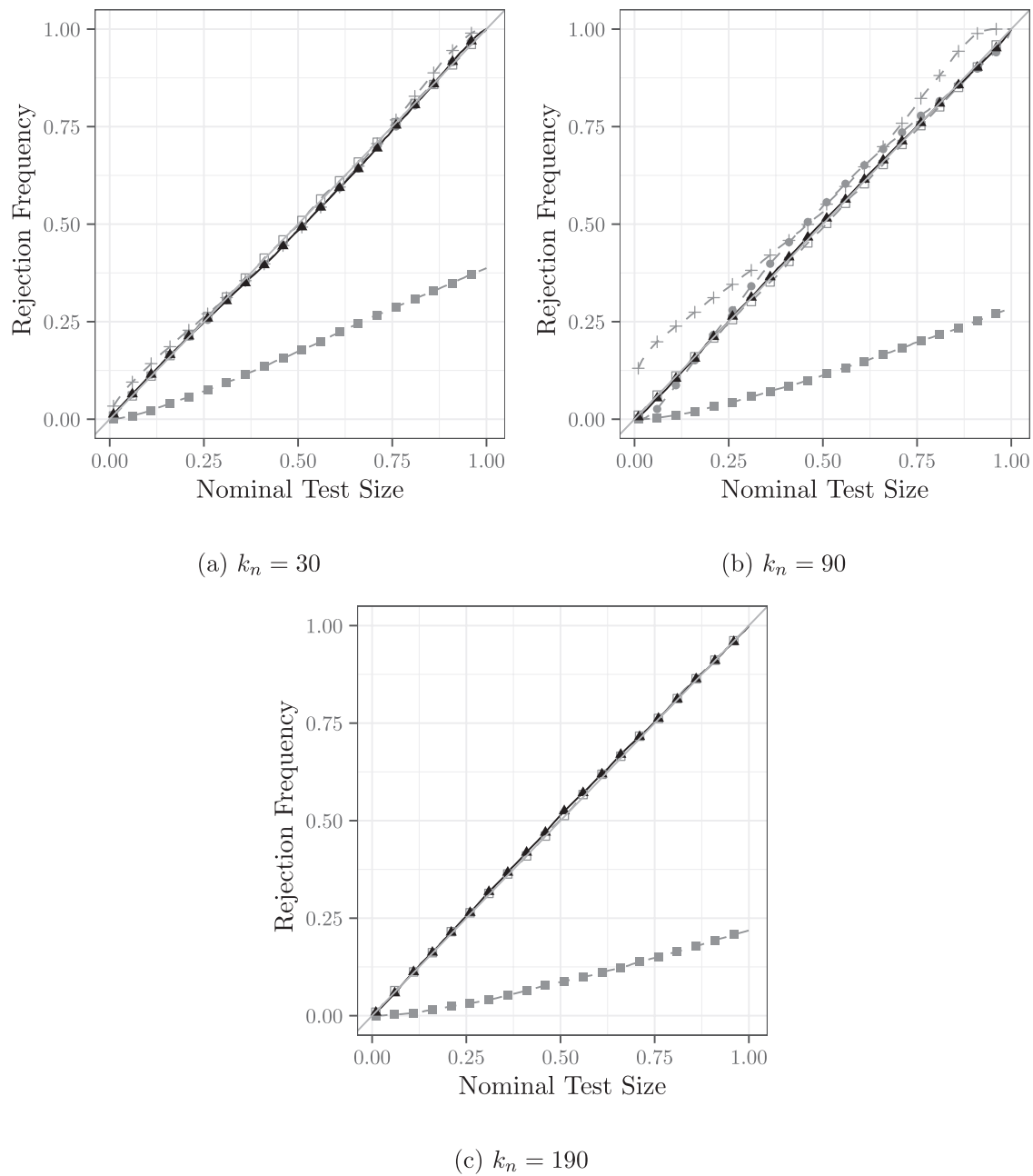


Figure 1. PP Plots for Sparse IVs, homoscedastic errors,  $\beta = 1$ ,  $\mu^2 = 0$ ,  $H_0 : \beta_0 = 1$ .

CMS exhibit power properties substantially worse than those of the BCCH Sup Score test and the RJAR test. For the case of 90 dense IVs (Figure 3), the RJAR test is substantially more powerful than all other alternatives. For the case of 190 sparse and dense IVs (Figure 4), the RJAR test is more powerful than the BCCH Sup Score test.<sup>10</sup> The RJAR test is also more powerful than the AR test of CT. Thus, for all the DGPs that are considered

here, the RJAR test is as powerful as existing methods whenever these are applicable, and sometimes much more powerful.

## 5. Empirical Application

We consider an empirical application based on Card (2009). The coefficient of interest is given by  $\beta_s$  in the following model:

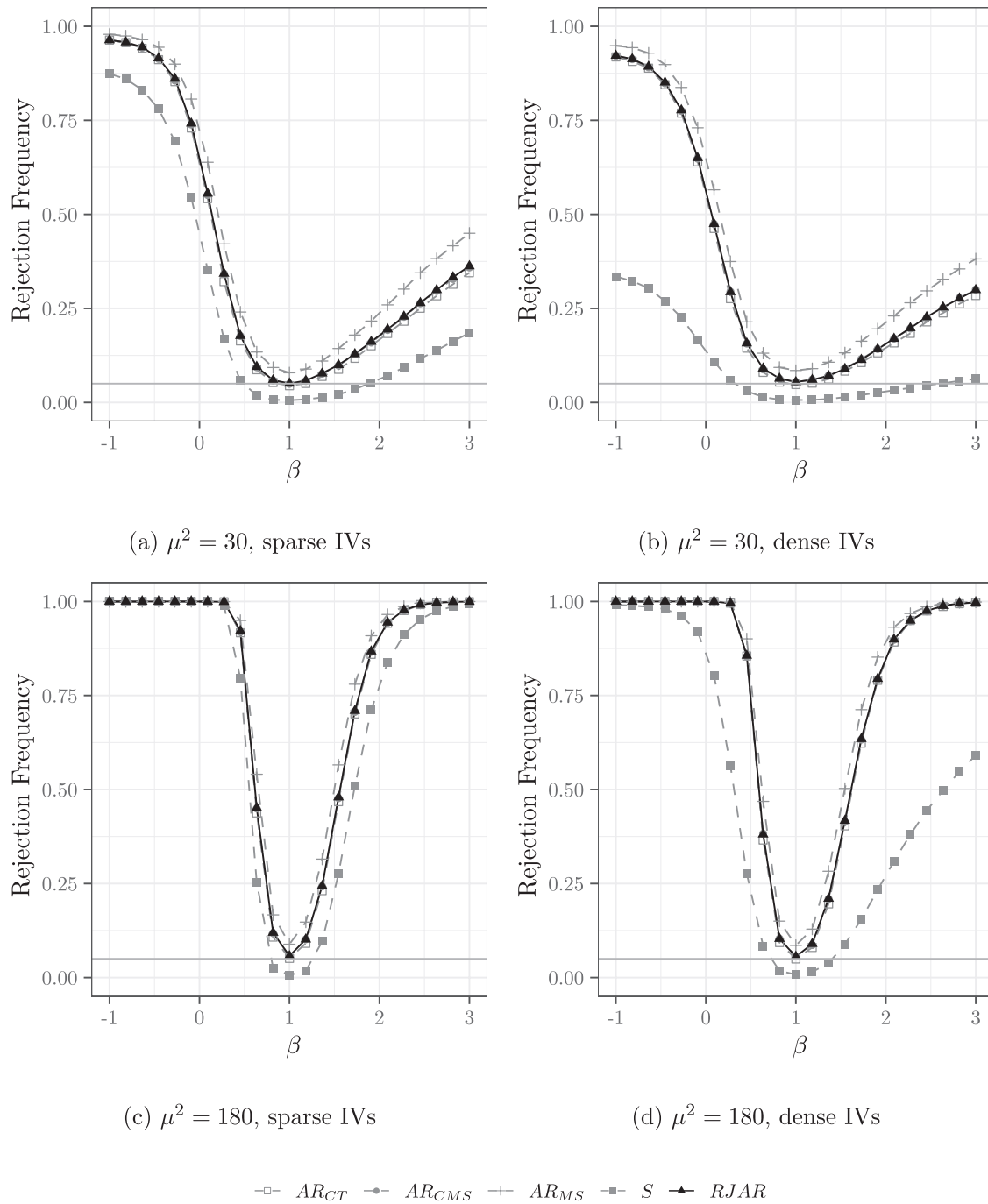
$$y_{is} = \beta_s X_{is} + \delta'_s W_i + \varepsilon_{is}, \quad (12)$$

where  $y_{is}$  is the difference between residual log wages for immigrant and native men in skill group  $s$  in city  $i$ ,<sup>11</sup>  $X_{is}$  is the log ratio

<sup>10</sup>The poor power of the BCCH Sup Score is likely a consequence of the Union Bound underlying their Lemma 5. We note that the poor power properties of the BCCH Sup Score are acknowledged in BCCH, and are what leads them to recommend using their main estimator (Post-LASSO two-stage least squares) whenever weak identification is not a concern, and the first stage is sufficiently sparse.

<sup>11</sup>As discussed in Card (2009, p. 11, footnote 17), residual wages are wages once observed characteristics of the entire U.S. workforce are controlled for.



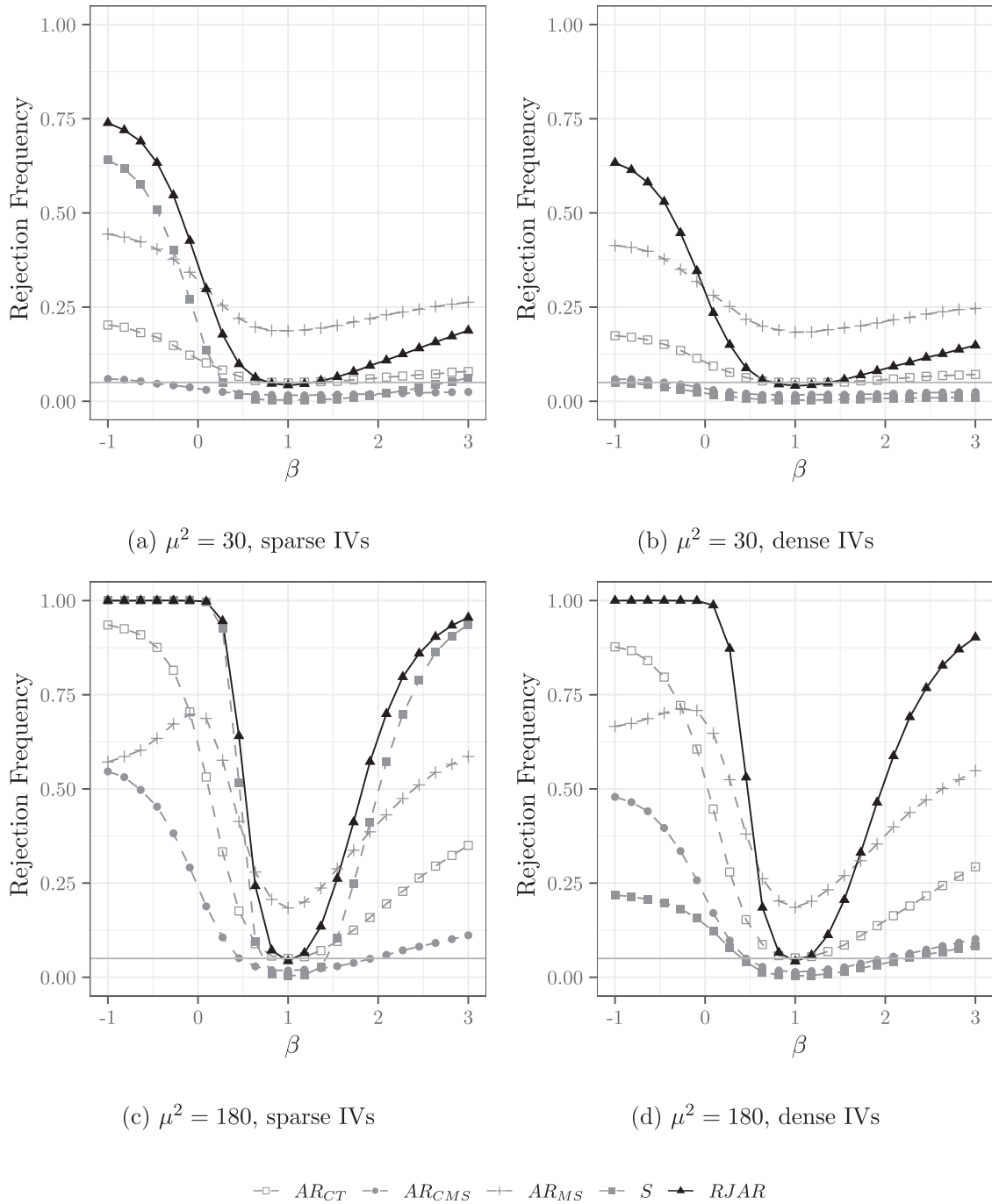


**Figure 2.** Power curves for 30 IVs. Nominal test size of 5% indicated by the grey horizontal line.  $H_0 : \beta_0 = 1$ .

of immigrant to native hours worked in skill group  $s$  of both men and women in city  $i$ , and  $W_i$  is a vector of city-level controls with coefficient vector  $\delta_s$ , and  $\varepsilon_{is}$  is the structural error. In the context of the production function specified in Card (2009, sec. I),  $\beta_s$  can be interpreted as the (negative) inverse elasticity of substitution between immigrants and natives in the United States in their respective skill group. As in Card (2009), we consider two skill groups  $s = h, c$  (high school or college equivalent) separately.

Card (2009) raises the concern that unobserved factors in a city may lead to both higher wages and higher employment levels of immigrants relative to natives, causing  $X_{is}$  to be

endogenous. Card (2009) proposes to use the ratio of the number of immigrants from country  $l$  in city  $i$  to the total number of immigrants from foreign country  $l$  in the United States as an IV. The rationale for these IVs is that existing immigrant enclaves are likely to attract additional immigrant labor through social and cultural channels unrelated to labor market outcomes. We consider two sets of IVs. First, we consider the original setup of Card (2009), using as IVs the  $k_n = 38$  different countries of origin of the immigrants. Second, motivated by the saturation approach of Blandhol et al. (2022), we consider the setup where these 38 original IVs are interacted with the  $q = 9$  available controls (including a constant). This yields  $k_n = 342$  IVs. In



**Figure 3.** Power curves for 90 IVs. Nominal test size of 5% indicated by the grey horizontal line.  $H_0 : \beta_0 = 1$ .

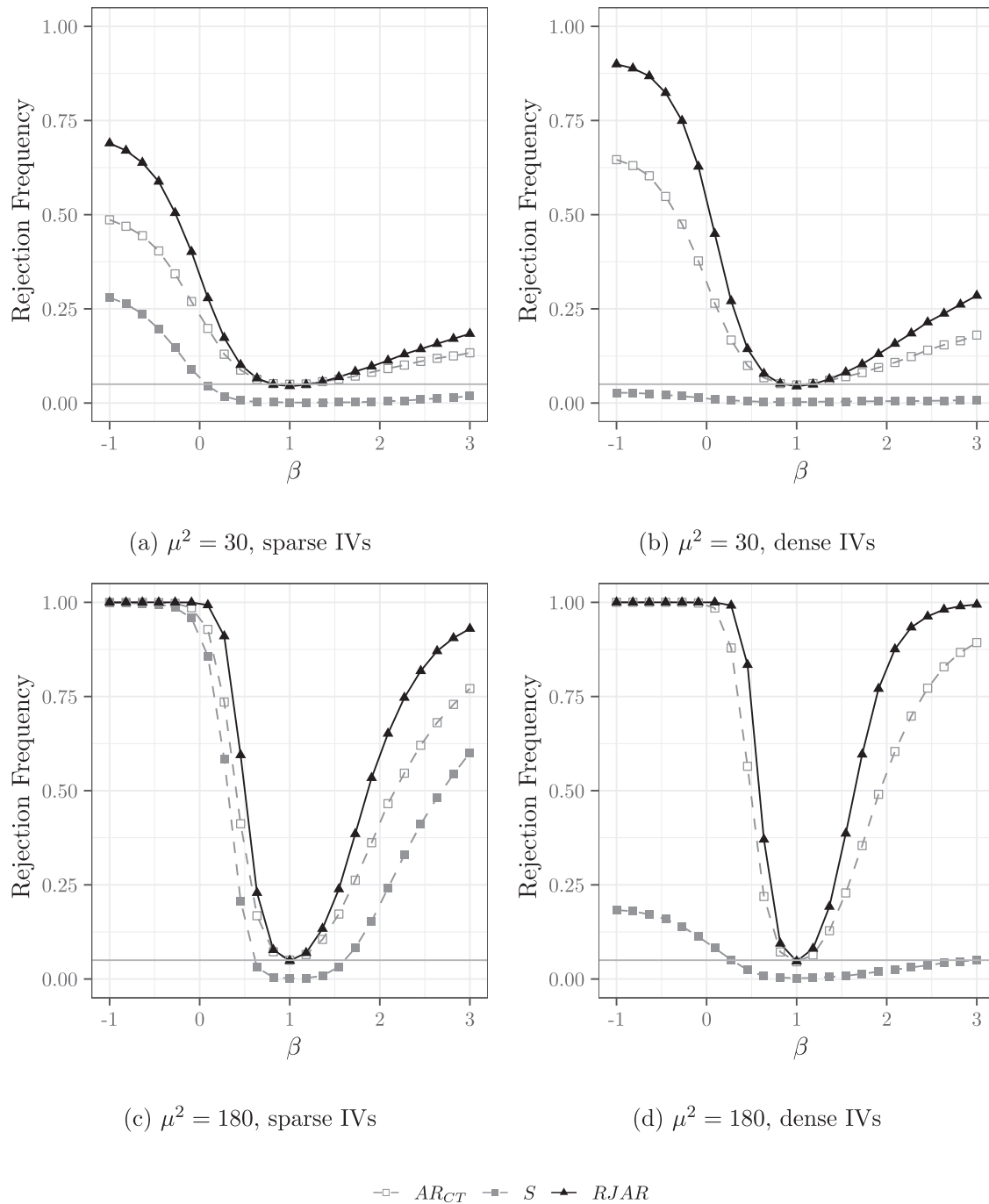
both cases, the number of observations (i.e., the number of cities) is  $n = 124$ .<sup>12</sup>

We construct (weak-identification robust) confidence sets for  $\beta_s$  by inverting the AR tests of CT, CMS, and MS, the Sup Score test, and the RJAR test. Thus, the 95% confidence set for any test is obtained as the collection of  $\beta_{s,0}$  for which that test does not reject the null at 5% level of significance. As in the simulation exercise in Section 4, we search over values greater than 1 when

choosing  $\gamma_n^*$  in case  $r_n < k_n$ . Again  $c_{BCH} = 1.1$  and  $\theta = 0.05$ . The number of bootstrap replications for the AR test of CT is set to 2,500. A grid of 100 values for  $\beta_{s,0}$  is used for  $s = h, c$ . Data is taken from a single cross section, as made available by Goldsmith-Pinkham, Sorkin, and Swift (2020).

Figure 5 shows the confidence sets when  $k_n = 38$  for high-school workers and college workers. We find that  $\gamma_n^* = 0$ , implying that no regularization is needed. This is in line with our simulations in Section 4, where we found that regularization was not needed to maximize the sum in Assumption 3 when  $k_n/n = 0.3$ . Furthermore,  $\max_{1 \leq i \leq n} P_{ii} = 0.944$ , and  $r_n^{-1} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n^*})^2 = 0.513$ . We also point out that only three diagonal entries of  $P$  are

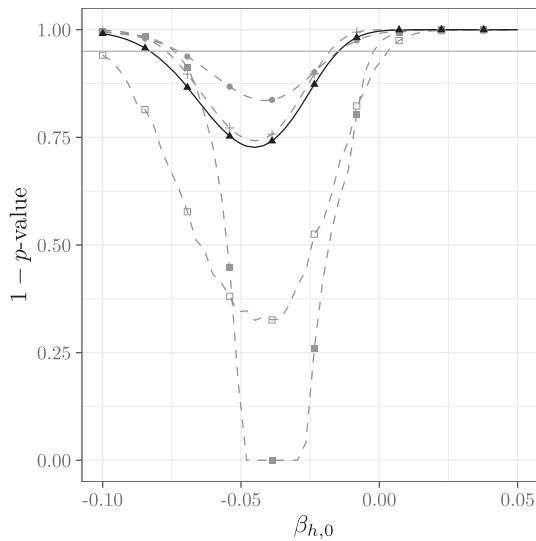
<sup>12</sup>We note that the controls and the IVs are at the city level, and hence the same for the applications to high-school workers and college workers. Therefore, the projection matrix and the ridge-regularized projection matrix will also be the same across these two skill groups.



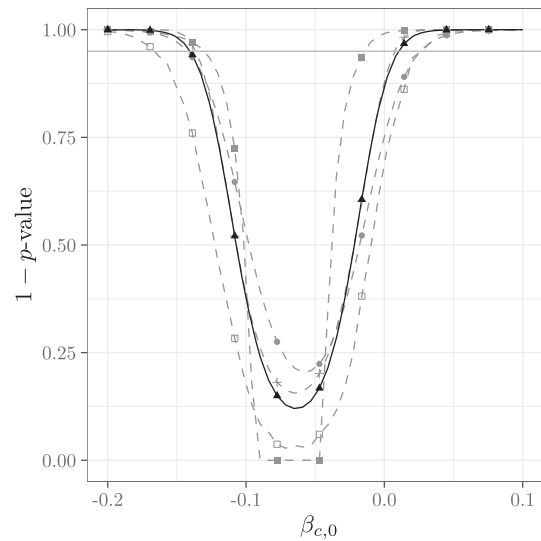
**Figure 4.** Power curves for 190 IVs. Nominal test size of 5% indicated by the gray horizontal line.  $H_0 : \beta_0 = 1$ .

larger than 0.9, which suggests that [Assumption 3](#) is reasonably satisfied by part 2. of [Proposition 3.1](#). The 95% confidence sets for each test are given by all the points below the grey horizontal line. The confidence sets for both skill groups broadly confirm the results in [Card \(2009\)](#). We find that the confidence sets for high-school workers is smallest for the jackknifed AR statistic of CMS, whereas the BCCH Sup Score test yields the smallest confidence interval for the application to college workers. For both cases, the AR test of CT yields the largest confidence interval. Based on the power results in [Section 4](#), this suggests a very sparse first stage for college workers, that is, a few nationalities being highly predictive of inflows of immigrant labor.

[Figure 6](#) shows the confidence sets when  $k_n = 342$  for high-school workers and college workers, respectively. Since  $r_n = n - q = 124 - 9 = 115 < 342 = k_n$ , the jackknifed AR statistics of CMS and MS are not applicable. We find that  $\gamma_n^* = 5.299$  and  $r_n^{-1} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{Y_n^*})^2 = 0.106$ . The 95% confidence sets obtained by inverting the AR test of CT are empty for both  $\beta_h$  and  $\beta_c$ . This could be due to heteroscedasticity in the error terms. In line with the simulation results on power reported in [Section 4](#), the RJAR test yields smaller confidence intervals than the BCCH Sup Score test. The qualitative conclusions with respect to the case of 38 IVs remain unchanged.



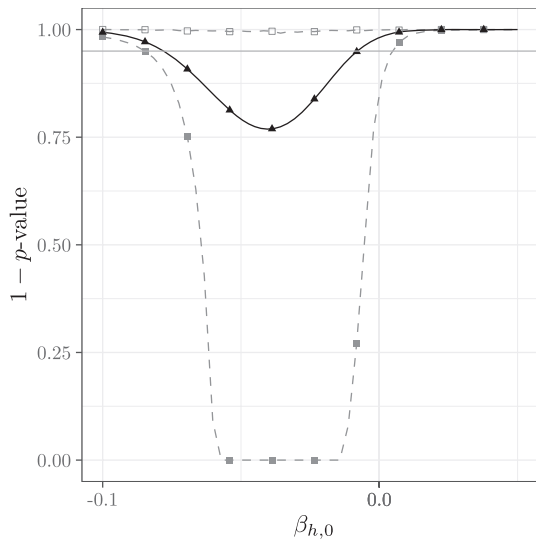
(a) High-school workers



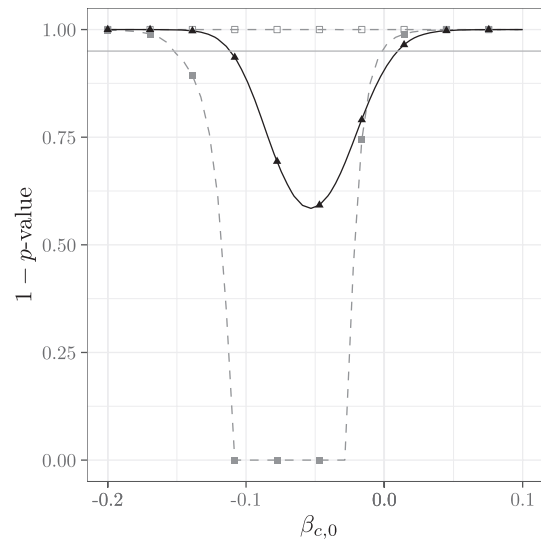
(b) College workers

—□—  $AR_{CT}$  —○—  $AR_{CMS}$  —△—  $AR_{MS}$  —■—  $S$  —▲—  $RJAR$

**Figure 5.** 95 % confidence sets for  $\beta_5$  for the application in (12) with  $k_n = 38$  IVs.  $\max_i p_{ii} = 0.944$ .  $\gamma_n^* = 0$ ,  $r_n^{-1} \sum_{i=1}^n \sum_{j \neq i} (p_{ij}^{\gamma_n^*})^2 = 0.513$ .



(a) High-school workers



(b) College workers

—□—  $AR_{CT}$  —■—  $S$  —▲—  $RJAR$

**Figure 6.** 95 % confidence sets for  $\beta_5$  for the application in (12) with  $k_n = 342$  IVs.  $\gamma_n^* = 5.299$ ,  $r_n^{-1} \sum_{i=1}^n \sum_{j \neq i} (p_{ij}^{\gamma_n^*})^2 = 0.106$ .

## 6. Conclusion

We contributed to the literature on (very) many IVs in the cross-sectional linear IV model by proposing a new, ridge-regularized jackknifed AR test. Our test compares favorably with existing methods in the literature both theoretically, by allowing for

high-dimensional IVs and weakening a common assumption on the IVs' projection matrix, and practically, by having correct asymptotic size and displaying favorable power properties even when the number of IVs approaches or exceeds the number of observations.

## Supplementary Materials

The supplementary material contains the proofs of the theoretical results reported in Section 3 of the article and the additional Monte Carlo simulation results mentioned in Section 4 of the article.

## Acknowledgments

The views expressed herein are those of the authors and should not be attributed to the IMF, its Executive Board, or its management. We thank Frank Kleibergen, Damian Kozbur, Anna Mikusheva, Bent Nielsen and Alexei Onatski for useful comments and discussions.

## Disclosure Statement

The authors report there are no competing interests to declare.

## Funding

This research is funded by the European Research Council via Consolidator grant number 647152 and the German National Merit Foundation.

## ORCID

Sophocles Mavroeidis  <http://orcid.org/0000-0002-8851-8044>

## References

- Anatolyev, S., and Gospodinov, N. (2011), "Specification Testing in Models with Many Instruments," *Econometric Theory*, 27, 427–441. [1084]
- Anatolyev, S., and Yaskov, P. (2017), "Asymptotics of Diagonal Elements of Projection Matrices Under Many Instruments/Regressors," *Econometric Theory*, 33, 717–738. [1084]
- Anderson, T., and Rubin, H. (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *The Annals of Mathematical Statistics*, 20, 46–63. [1083,1084]
- Andrews, D., and Stock, J. (2007), "Testing with Many Weak Instruments," *Journal of Econometrics*, 138, 24–46. [1083]
- Bai, J., and Ng, S. (2010), "Instrumental Variable Estimation in a Data Rich Environment," *Econometric Theory*, 26, 1577–1606. [1084]
- Bekker, P., and Crudu, F. (2015), "Jackknife Instrumental Variable Estimation with Heteroskedasticity," *Journal of Econometrics*, 185, 332–342. [1087]
- Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012), "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain," *Econometrica*, 80, 2369–2429. [1084]
- Blandhol, C., Bonney, J., Mogstad, M., and Torgovitsky, A. (2022), "When is TSLS Actually LATE?" *Becker Friedman Institute Working Papers* (2022-16), 1–68. [1083,1090]
- Boot, T., and Ligtenberg, J. (2023), "Identification- and Many Instrument-Robust Inference via Invariant Moment Conditions," arXiv preprint arXiv:2303.07822. [1084]
- Bun, M., Farbmacher, H., and Poldermans, R. (2020), "Finite Sample Properties of the GMM Anderson-Rubin Test," *Econometric Reviews*, 39, 1042–1056. [1084]
- Burgess, S., and Thompson, S. G. (2021), *Mendelian Randomization: Methods for Causal Inference Using Genetic Variants*, Boca Raton, FL: CRC Press. [1083]
- Card, D. (2009), "Immigration and Inequality," *American Economic Review Papers and Proceedings*, 99, 1–21. [1084,1085,1089,1090,1092]
- Carrasco, M., and Tchuente, G. (2016), "Regularization Based Anderson Rubin Tests for Many Instruments," *University of Kent, School of Economics Discussion Papers* (1608), 1–34. [1084]
- Chao, J., Swanson, N., Hausmann, J., Newey, W., and Woutersen, T. (2012), "Asymptotic Distribution of JIVE in a Heteroskedastic IV Regression with Many Instruments," *Econometric Theory*, 28, 42–86. [1084,1087]
- Crudu, F., Mellace, G., and Sándor, Z. (2020), "Inference in Instrumental Variable Models with Heteroskedasticity and Many Instruments," *Econometric Theory*, 77, 1–30. [1084]
- Davies, N. M., von Hinke Kessler Scholder, S., Farbmacher, H., Burgess, S., Windmeijer, F., and Smith, G. D. (2015), "The Many Weak Instruments Problem and Mendelian Randomization," *Statistics in Medicine*, 34, 454–468. [1083]
- Gabaix, X., and Koijen, R. (2020), "Granular Instrumental Variables," *NBER Working Paper Series* (28204), 1–98. [1083]
- Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020), "Bartik Instruments: What, When, Why, and How," *American Economic Review*, 110, 2586–2624. [1091]
- Hansen, C., and Kozbur, D. (2014), "Instrumental Variables Estimation with Many Weak Instruments Using Regularized JIVE," *Journal of Econometrics*, 182, 290–308. [1084,1087,1088]
- Kaffo, M., and Wang, W. (2017), "On Bootstrap Validity for Specification Testing with Many Weak Instruments," *Economics Letters*, 157, 107–111. [1084]
- Kapetanios, G., Khalaf, L., and Marcellino, M. (2015), "Factor-Based Identification-Robust Inference in IV Regressions," *Journal of Applied Econometrics*, 31, 821–842. [1084]
- Lim, D., Wang, W., and Zhang, Y. (2024), "A Conditional Linear Combination Test with Many Weak Instruments," *Journal of Econometrics*, 238, 105602. [1084]
- Lönn, R., and Schotman, P. (2022), "Empirical Asset Pricing with Many Assets and Short Time Series," SSRN, 1–53. [1083]
- Matsushita, Y., and Otsu, T. (2022), "A Jackknife Lagrange Multiplier Test with Many Weak Instruments," *Econometric Theory*, 1–24. [1084]
- Mikusheva, A., and Sun, L. (2022), "Inference with Many Weak Instruments," *Review of Economic Studies*, 1–37 (forthcoming). [1084]
- Mueller-Smith, M. (2015), "The Criminal and Labor Market Impacts of Incarceration," *mimeo*, 1–59, University of Michigan. [1083]
- Phillips, P. C. B., and Gao, W. Y. (2017), "Structural Inference from Reduced Forms with Many Instruments," *Journal of Econometrics*, 199, 96–116. [1083]
- van Duijn, M., and Rouwendal, J. (2021), "Sorting Based on Urban Heritage and Income: Evidence from the Amsterdam Metropolitan Area," *Regional Science and Urban Economics*, 90, 1–18. [1083]