Robust detection of weak instruments

Matheus Silva

ARTICLE HISTORY
Compiled November 14, 2022

1. Introduction

Note: This is preliminary work submitted as a writing sample.

It is well known that standard statistical inference in instrumental variables models breaks down under weak identification. In this case, the researcher relies on robust inference methods that are often computationally expensive, potentially return wide confidence intervals, and do not allow to test many types of hypotheses - testing a subset of the structural parameters, for example. A pre-test for weak instruments may come in handy when there is suspicion the instruments are weak. If there is evidence that the instruments are not weak, the researcher can use standard statistical methodologies to carry out their hypotheses tests before turning to robust techniques.

The test for weak instruments developed by Stock and Yogo (2005) may perhaps be the most widely used procedure to address this issue. However, their results only apply in the case of homoskedastic error terms. Violations of this homoskedasticity assumption, which is the norm in economic data, make their test inapplicable from start. The problem of testing for weak instruments in the instrumental variables model remains open, with some progress made by Montiel-Olea and Pflueger (2013). They extend Stock and Yogo's (2005) results for the heteroskedastic case. However, their setup only contains only one endogenous variable, despite admitting finitely many instruments.

In this paper, I propose a test for weak identification that allows for finitely many instruments and endogenous variables and does not depend on the homoskedasticity assumption. This test is based on simulating of the distribution of the rank statistic in Kleibergen and Paap (2006) and avoids the issue of the naive bootstrap "mistakenly thinking" the model is identified with probability one (see Andrews et al. for a deeper discussion). The basic idea consists on testing the rank of the first-stage matrix of coefficients by bootstrapping the test statistic around a "worst-case scenario" as in Montiel-Olea and Pflueger (2013).

In the remainder of this paper I describe the model and the assumptions needed for the methodology to work. I proceed to describing the test statistic and how to perform the simulation procedure. I also provide Monte Carlo simulations to illustrate the simulations before applying my procedure to the AER papers that fit this setup.

TBD: Prove that the nonpivotal conditional bootstraps were developed by Booth et al (1992) work here. Apply the method to selected papers from Young's (2018, 2021) compilation of AER journals that consider more than one endogenous regressor.

2. The model

In what follows, I will use fairly standard notation for instrumental variable models with as few deviations as possible. I am interested in the following model:

$$y = X\beta + W\gamma + u \tag{1}$$

$$X = Z\Pi_T + W\delta + V \tag{2}$$

Where y is a $T \times 1$ vector that contains the response variable of interest, X is a $T \times K_X$ matrix of endogenous regressors ($K_X < \infty$ of those), Z is a $T \times K_Z$ matrix of instruments ($K_Z < \infty$ of those), and W is a $T \times K_W$ matrix of exogenous regressors. I will also assume that $K_Z > K_X$ since moments related to Nagar-type approximation may not exist in the case $K_Z = K_X$. The econometrician is primarily interested in the vector of parameters β (of dimension $K_X \times 1$). Π_T is a $K_Z \times K_X$ matrix of first-stage coefficients that depends on the sample size. Finally, u and V are stochastic disturbances; u has dimension $T \times 1$ and V is a $T \times K_X$ matrix.

It is also convenient to cast this model in its reduced form:

$$y^* = Z^*\Pi_T^* + E \tag{3}$$

Where $Y \equiv [y \vdots X]$, $Z^* = [W \vdots Z]$ ($\vdots$ represents the horizontal concatenation of matrices) and all other remaining matrices and quantities are obtained by stacking the structure of the model (1). I work under the following assumptions:

Assumption 2.1. (Assumptions related to rank)

a. Π_T is "local to incomplete rank." That is, $\Pi_T = \Pi_{LR} + C/\sqrt{T}$, where Π_{LR} is a $K_Z \times K_X$ rank-deficient matrix and C is a $K_Z \times K_X$ fixed, real matrix.

b. Π_T is "local to zero." That is, $\Pi_T = C/\sqrt{T}$, where C is a $K_Z \times K_X$ fixed, real matrix.

Assumption 2.1 provides two alternative ways to test

states that the matrix of first-stage coefficients is close to not having full rank. In this situation, statistical inference on the parameters of interest β depends on weak instrument asymptotics. This setup differs from Stock and Yogo's (2005), where $\Pi_T = C/\sqrt{T}$. Sanderson and Windmeijer (2016) study conditional F-statistics within this context under the homoskedasticity assumption .

Assumption 2.2. As $T \to \infty$,

1.
$$\begin{pmatrix} Z'u/\sqrt{T} \\ vec(Z'V)/\sqrt{T} \end{pmatrix} \to^d \mathcal{N}(0, \Sigma) \tag{4}$$

   Where $\Sigma = \begin{bmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ is a positive definite matrix.
2. There exists a sequence of positive definite estimates $\{\hat{\Sigma}(t)\}$, measurable with respect to $\{y_t, X_t, Z_t\}_{t=1}^T$, such that $\hat{\Sigma}(T) \to^p \Sigma$, as $T \to \infty$.
3. $\mathbb{E}\left[\begin{pmatrix} u_t \\ V_t \end{pmatrix} \begin{pmatrix} u_t & V_t' \end{pmatrix}\right] = \begin{pmatrix} \omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \equiv \Omega$, a positive definite matrix. The fourth moments of these error terms must also be finite in order for a law of large numbers to hold, so $[u;\ vec(V)']'[u;\ vec(V)]/T \to^p \Omega$.

Assumption 2.2 begins by stating that a Central Limit Theorem holds for the inner products of the instrument matrix and all the error terms in the model. As Montiel-Olea and Pflueger (2013) highlight, this first part of Assumption 2.2 is the main point of departure from Stock and Yogo's (2005) setup since Σ does not necessarily have a Kronecker-product form[1]. The second requirement is that the econometrician can estimate Σ consistently. Finally, a law of large numbers holds for the variance-covariance matrix of the error terms.

3. The weak instrument set

The focus of this paper is on the bias of the two-stage least square estimator of β,

$$\hat{\beta}_{2SLS} = (X'P_Z X)^{-1} X' P_Z y$$

As in Stock and Yogo (2005), the goal is to compare the relative bias of the ordinary and the two stage least square estimators depending on the strength of the instruments in the first stage. This comparison defines the weak instrument set. Even though testing the strength of the instruments in the first stage is a test for the rank of the matrix of the first-stage coefficients, the usual test statistics do not follow a standard chi-square distributions. In fact, the distributions of those test statistics depend on a noncentrality parameter for which there are no consistent estimators.

[1]Making the results on the matrix normal distribution unusable (see Gupta and XXXX, 2000.)

In regular situations, bootstrapping the test statistic provides a valid way to conduct hypothesis tests even in this non-pivotal case. However, usual bootstrap procedures do not work here because the parameter matrix of the first stage is on the verge of being rank-deficient, but it is not. Andrews et al (2018) intuitively explains this is due to the ordinary least squares estimates of Π_T - the first stage parameter matrix in the notation presented in Section 2 - have full rank with probability one, the bootstrap then "thinks" Π_T is identified. "Naïve" bootstrapped rank statistics (say Cragg and Donald (1993), Robin and Smith (2000), or Kleibergen and Paap (2006)) will then follow their usual chi-squared distributions. The bootstrap procedure I propose consists on resampling the Kleibergen and Paap (2006) rank statistic (rk statistic) for the matrix of first-stage coefficients conditional on the instruments being weak (which will be defined later). In other words, I restrict the bootstrap procedure to resample rk under the null hypothesis that Π_T is in a neighborhood of being rank-deficient. As Kleibergen and Paap (2006) show, the rk statistic is a generalization of other statistics for tests of matrix ranks that allows for heteroskedasticity and underidentification. Therefore. it is the statistic I am interested in using[2]. Its formula is:

$$rk(K_X - 1) = T\hat{\lambda}_{K_X-1}\hat{\Omega}^{-1,KP}_{K_X-1}\hat{\lambda}_{K_X-1} \tag{5}$$

The set of weak instruments is based on the relative bias of the two-stage least squares estimator ($\hat{\beta}_{2SLS}$) relative to a benchmark, similar to Montiel-Olea and Pflueger (2013). Let RBm^* be the worst possible two-stage least squares bias with respect to a benchmark BM,

$$RBm^*_{2SLS} = sup_{\beta\in\mathbb{R}^{K_X},\Pi^*\in\mathcal{M}^{(K_Z+K_W)\times K_X}} \frac{\mathbb{E}(\hat{\beta}_{2SLS}-\beta)'\Sigma_{XX}\mathbb{E}(\hat{\beta}_{2SLS}-\beta)}{BM} \tag{6}$$

Where Π^* is the parameter matrix of the first-stage regression, $\mathcal{M}^{(K_Z+K_W)\times K_X}$ is the space of matrices with dimension $(K_Z + K_W) \times K_X$, and $\Sigma_{XX} = (X'X)^{-1}$.

Following Montiel-Olea and Pflueger's work (2013), I use an approximation of the bias of the two-stage least squares estimator in the numerator and a bound for that quantity as the denominator. I use Phillips's (2000) generalized approximation[3] of the bias of the two-stage least squares which does not depend on a covariance matrix that can be written as a Kronecker product. Let $\hat{\beta}^i_{2SLS}$ be the two-stage least squares estimator of the coefficient associated with the $i-th$ endogenous regressor. Phillips (2000) shows that:

$$\mathbb{E}(\hat{\beta}^i_{2SLS} - \beta^i) = tr(\mathcal{F}_i \times \hat{\Omega}^*)$$

[2]The rk statistic is equivalent to the Cragg-Donald's (1993) statistic in the setup Stock and Yogo (2005) study.

[3]In the homoskedastic case, this approximation is equivalent to Nagar's (1959) approximation.

Where $\hat{\Omega}^* = Cov(vec(\hat{\Pi}^*))$ is the heteroskedasticity-robust estimate of the covariance matrix of the reduced form coefficients and $\mathcal{F}_i$ is

$$\mathcal{F}_i \equiv (\xi_i + \xi_i')/2$$
$$\xi_i \equiv (H' \otimes I)\{(X^{*\prime} X^*)^{-1} \otimes M -$$
$$- [(X^{*\prime} X^*)^{-1} X^{*\prime} Z^* \otimes Z^{*\prime} X^* (X^{*\prime} X^*)^{-1}] I'^*\}(e_i \otimes I)(\beta' \otimes I)$$
$$M \equiv Z^{*\prime}(I - X^*(X^{*\prime} X^*)^{-1} X^{*\prime}) Z^*$$

Where I^* is a commutation matrix (Magnus and Neudecker, 1979) and H is a selection matrix. Z^*, X^*0... **WRITE THIS**

It is worthwhile to notice that this is not the asymptotic covariance matrix; so, if the instruments are not weak, $\mathcal{F}_i \to^p c_i$ (where c_i is some constant matrix) and $\hat{\Omega}^* \to^p 0$, meaning the bias vanishes asymptotically. With weak instruments, however, the convergence of $\hat{\Pi}^*$ is offset by the increase in $\mathcal{F}_i$, rendering the two-stage least squares estimators inconsistent.

With respect to the benchmark, I use Fang, Loparo, and Feng's (1994) result that:

$$\mathbb{E}(\hat{\beta}^i_{2SLS} - \beta^i) = tr(\mathcal{F}_i \times \hat{\Omega}^*) \geq mineig(\mathcal{F}_i) \times tr(\hat{\Omega}^*)$$

The interpretation of this is that the bias of each coefficient is bounded by the "size" of the estimate of the covariance matrix (the larger, the higher the lower bound) and the strength of the instruments of the first stage (the weaker, the larger the bias).

Remark 1. Even though $tr(\mathcal{F}_i \times \hat{\Omega}^*) \geq mineig(\mathcal{F}_i) \times tr(\hat{\Omega}^*)$, that does not necessarily mean that $\frac{tr(\mathcal{F}_i \times \hat{\Omega}^*)}{mineig(\mathcal{F}_i) \times tr(\hat{\Omega}^*)} \geq 1$ because the quantities may have opposite signs (which flips the inequality once divided).

The model I described in Section 2 is defined by the set of matrices of first-stage coefficients, and the covariance matrices of the error terms, $W \equiv \{\Pi, \Sigma, \Omega\}$. If it were possible to consistently estimate the multivariate version of the concentration matrix, one could easily test this hypothesis. The test for weak instruments consists in checking whether W generates a "large" relative bias with respect to the benchmark, which is the measure of weakness we are interested in. Define $\mathcal{W}_{2SLS,bias} \equiv \{W : RBm^* \geq \tau\}$, $\tau \in \mathbb{R}_+$, as the set of matrices W that generates a high relative bias of the two-stage least squares compared to the benchmark. The hypothesis we want to test is:

$$\begin{cases} H_0 : W \in \mathcal{W}_{2SLS,bias} & \text{(The instruments are weak)} \\ H_A : W \notin \mathcal{W}_{2SLS,bias} & \text{(The instruments are not weak)} \end{cases}$$

The test statistic is the rk statistic for the rank of the matrix Π in (5). The decision rule is to reject the null hypothesis when $rk \notin \{rk(W) : W \in \mathcal{W}_{2SLS,bias}\}$.

3.1. *Critical values*

As Andrews et al [literature review] point out, the naive bootstrap fails to recover the correct distribution because the OLS estimators "think" the model is identified with probability one. In this case, the rk statistic follows a χ^2 distribution with degrees of freedom dependent on number of endogenous variables. To avoid this, I approach the bootstrap problem using a conditional bootstrap such as the one in Booth et al (1990/2?). I compute the distribution under the null hypothesis with the following algorithm:

Algorithm 1: Proposed bootstrap procedure

Input: A dataset $\{X_i, Z_i, W_i, y_i\}_i$, the total number of bootstrap repetitions B, and a cutoff τ

Result: Bootstrap distributions of the rk statistic

Steps:

1. Run the first-stage regression with the original data and store $\hat{\Pi}$ and its covariance matrix.
2. Calculate rk.

while $b \leq B$ **do**

  Resample from $\{X_i, Z_i, W_i, y_i\}$ with replacement.

  With the resampled data, calculate RBm_b^* by solving the optimization problem in (6).

  Using the solution from the optimization problem, calculate rk_b.

  Store RBm_b^* and rk_b.

end

With the bootstrapped rk statistics, one can condition on the optimal relative bias being greater than a the threshold τ and then nonparametrically estimate its distribution.

4. Monte Carlo simulations

I now present a few Monte Carlo simulations to illustrate the distribution of rk under the null of weak instruments.

Setup #1

$$\Pi = \begin{bmatrix} 0 & 0.02/\sqrt{n} \\ 0.01/\sqrt{n} & 0 \\ 0.3/\sqrt{n} & 0.2/\sqrt{n} \end{bmatrix}$$

$$\beta = [0;0]$$

$$\gamma = \delta = 0$$

$$e_{1i} \sim \mathcal{N}(0,2)$$

$$e_{2i}|z_i \sim \mathcal{N}(0, \max(0.2, (-(2z_{1i}-3)^2+20)/20, 1))$$

$$e_{3i}|z_i \sim \mathcal{N}(0, \max(0.2, (0.5+3|z_{2i}|)^3/2000)$$

$$e_{4i} \sim \mathcal{N}(0,2)$$

$$e_{5i} \sim \mathcal{N}(0,2)$$

$$V_{1i} = e_{1i} + e_{2i}$$

$$V_{2i} = e_{3i} + e_{4i}$$

$$V = [V1 \vdots V2]$$

$$u = e_{1i} + e_{3i} + e_{5i}$$

$$z_{1i} \sim \mathcal{N}(0,1)$$

$$z_{2i} \sim \mathcal{N}(0,1)$$

$$z_{3i} \sim \mathcal{N}(0,1)$$

Please find the results in the next page. In this case, we know that, under the alternative hypothesis, the distribution of the test statistic follows a $\chi^2(2)$ distribution (from Kleibergen and Paap, 2006.)