An Identification and Dimensionality Robust Test for Linear IV

Manu Navjeevan

University of California,
Los Angeles

January, 2024

Setup

Consider the linear IV model

$$y_i = x_i'\beta + z_{1i}'\Gamma + \epsilon_i, \quad \mathbb{E}[\epsilon_i | z_i] = 0$$

Researcher observes $(y_i, x_i, z_i)'$

- Outcome $y_i \in \mathbb{R}$;
  - e.g income
- Endogenous Regressor(s) $x_i \in \mathbb{R}^{d_x}$;
  - e.g years of education
- Instruments $z_i = (z_{1i}, z_{2i})' \in \mathbb{R}^{d_c} \times \mathbb{R}^{d_z - d_c}$.
  - e.g demographic characteristics, quarter of birth

Research Question

I propose a new test for

$$H_0 : \beta = \beta_0 \ \text{ vs. } \ H_1 : \beta \neq \beta_0$$

when

- Identification is arbitrarily weak;
- Errors are heteroskedastic;
- The number of instruments is potentially large, $d_z \gg n$.
  - Dimensionality of controls is fixed, $d_c \ll n$.

Existing Weak IV Robust Tests

- **Low Dimensional:** Anderson and Rubin (1949), Staiger and Stock (1997), Kleibergen (2002, 2005), Moreira (2003, 2009), Newey and Windmeijer (2009), Andrews (2016).
  - Analyses treat d_z as fixed or growing slowly with sample size. Tests control size under heteroskedasticity when $d_z^3/n \to 0$ (Andrews and Stock, 2007).

Existing Weak IV Robust Tests

- **Low Dimensional:** Anderson and Rubin (1949), Staiger and Stock (1997), Kleibergen (2002, 2005), Moreira (2003, 2009), Newey and Windmeijer (2009), Andrews (2016).
  - Analyses treat d_z as fixed or growing slowly with sample size. Tests control size under heteroskedasticity when $d_z^3/n \to 0$ (Andrews and Stock, 2007).
- **Many Instruments:** Crudu et al. (2021), Mikusheva and Sun (2021), Matsushita and Otsu (2022), Lim et al. (2022).
  - Allow $d_z/n \to \varrho \in [0, 1)$ but use Chao et al. (2012) CLT that requires $d_z \to \infty$.
  - If $d_z \to \infty$ slowly, asymptotic approximation may be poor in finite samples.

Existing Weak IV Robust Tests

- **Low Dimensional:** Anderson and Rubin (1949), Staiger and Stock (1997), Kleibergen (2002, 2005), Moreira (2003, 2009), Newey and Windmeijer (2009), Andrews (2016).
  - Analyses treat d_z as fixed or growing slowly with sample size. Tests control size under heteroskedasticity when $d_z^3/n \to 0$ (Andrews and Stock, 2007).
- **Many Instruments:** Crudu et al. (2021), Mikusheva and Sun (2021), Matsushita and Otsu (2022), Lim et al. (2022).
  - Allow $d_z/n \to \varrho \in [0, 1)$ but use Chao et al. (2012) CLT that requires $d_z \to \infty$.
  - If $d_z \to \infty$ slowly, asymptotic approximation may be poor in finite samples.
- **High Dimensional:** Belloni et al. (2012), Gautier and Rose (2021), Mikusheva (2023).
  - $d_z \gg n$ allowed under strong identification but limited work when identification is weak.

Motivation

Even when $d_z < n$, it can be unclear which test, if any, is applicable

1. Bastos et al. (2018). *Export Destinations and Input Prices*, AER.

   Interested in effect of Portugese firm export destinations on prices paid for inputs.

   - Instrument for firm export destinations using interactions of exchange rate movements and initial export destinations.

   First-stage F-statistic on 100 instruments (≈ 2.5) indicates weak identification.

   - Authors validate results using Anderson-Rubin test, arguments about direction of weak IV bias.

   Which identification robust test to use?

   - $d_z^3 = 1,000,000 \gg n = 45,659$;
   - $d_z = 100 \overset{?}{\to} \infty$.

Motivation

Even when $d_z < n$, it can be unclear which test, if any, is applicable

2. Gilchrist and Sands (2016). *Something to Talk About: Social Spillovers in Movie Consumption*, JPE.

   Consider effect of strong opening weekend on ticket sales in later weeks. Instrument for opening weekend sales using national weather conditions.

   - Unusually poor weather conditions may lead people to choose to watch a movie instead of "chilling" outside.

   Start with 52 weather instruments, then use LASSO select up to three. F-statistic on selected instruments ranges from 15-38, F-statistic on all instruments is ≈ 3.5.

   - Unclear whether F-statistic on selected instruments is interpretable.

   Which identification robust test to use?

   - $d_z^3 \approx 140,000 \gg n = 1671$;
   - $d_z = 52 \xrightarrow{?} \infty$

Motivation

Even when $d_z < n$, it can be unclear which test, if any, is applicable

3. Derenoncourt (2022). *Can You Move to Opportunity? Evidence from the Great Migration*, AER.

   Considers effect of Black Americans migration rates from 1940 - 1970 on income mobility gap in current day. Instrument for Black migration shares using "supply-side" variation.

   - Focus on industry conditions for industries with historically higher than average Black employment rates.

   Initial instrument set consists of 9 instruments, then use post-LASSO to estimate first-stage. F-statistic on selected variables is 14.78, on all variables is 11.68.

   - Stock and Yogo (2005) cutoff for $d_z = 9$ and no more than a 15% size distortion is 14.01.

   Which identification robust test to use?

   - $d_z^3 = 729 > n = 239$;
   - $d_z = 9 \xrightarrow{?} \infty$.

Contribution

1. Test can be applied in any of settings mentioned above.
   - Relies on a nuisance parameter that is easy to estimate using "out-of-the-box" methods.
   - Incorporate first stage information using ridge regression.
2. Limiting $\chi^2(d_x)$ distribution of test statistic is derived via direct gaussian approximation.
   - Number of instruments can be larger than n, existing CLTs cannot be applied;
   - Limiting distribution of test statistic is pivotal and does not require $d_z \to \infty$.
3. To improve power against certain alternatives, I propose a combination with sup-score test of Belloni et al. (2012).

Contribution

n	d_z	ϱ	New Test	And. Rubin	J. AR	J. LM
200	30	0.3	0.0498	0.0096	0.1090	0.0318
		0.6	0.0562	0.0088	0.1104	0.0292
	75	0.3	0.0488	0.0168	0.1144	0.0380
		0.6	0.0516	0.0122	0.1166	0.0390
500	30	0.3	0.0554	0.0174	0.0940	0.0272
		0.6	0.0570	0.0206	0.0984	0.0280
	75	0.3	0.0500	0.0274	0.1028	0.0470
		0.6	0.0522	0.0230	0.1002	0.0434
Average			0.0526	0.0169	0.1057	0.0354

Table 1: Simulated size of various tests with nominal level $\alpha = 0.05$ under weak identification and heteroskedastic errors. Parameter ϱ controls degree of endogeneity

Details

Prior Literature

In addition to previously mentioned results, contribute to the following literatures:

- **Weak Identification:** Nelson and Startz (1990), Bound et al. (1995), Wang and Zivot (1998), Stock and Wright (2000), Andrews et al. (2006), Hansen et al. (2008), Andrews and Cheng (2012), Cheng (2008, 2015), Chen and Fang (2019), Andrews and Guggenberger (2019), Cheng et al. (2022), Andrews and Mikusheva (2016, 2022, 2023).
- **Many/High-Dimensional Instruments:** Bekker (1994), Hahn (2002), Chao and Swanson (2005), Han and Phillips (2006), Anatolyev (2012), Adusumilli (2017), Phillips and Gao (2017), Hao and Lee (2023).
- **Gaussian Approximation:** Lindeberg (1922), Chatterjee (2006), Pouzo (2015), Chernozhukov et al. (2013, 2017), Graham (2017).

Table of Contents

Model

Focus on the case where $d_x = 1$. With first stage, can write model as

$$
\begin{aligned}
y_i &= x_i\beta + \epsilon_i \\
x_i &= \underbrace{\mathbb{E}[x_i \mid z_i]}_{\Pi_i} + v_i \qquad\qquad \mathbb{E}[(\epsilon_i, v_i)'|z_i] = 0
\end{aligned}
$$

- Controls z_{1i} assumed to be partialled out of (y_i, x_i);
- Random variables $\{(z_i, \epsilon_i, v_i)'\}_{i=1}^n$ independent and identically distributed;

Additionally, define the null errors $\epsilon_i(\beta_0) \coloneqq y_i - x_i\beta_0$.

Model

Focus on the case where $d_x = 1$. With first stage, can write model as

$$
\begin{aligned}
y_i &= x_i\beta + \epsilon_i \\
x_i &= \underbrace{\mathbb{E}[x_i \mid z_i]}_{\Pi_i} + v_i \qquad \mathbb{E}[(\epsilon_i, v_i)'|z_i] = 0
\end{aligned}
$$

- Controls z_{1i} assumed to be partialled out of (y_i, x_i);
- Random variables $\{(z_i, \epsilon_i, v_i)'\}_{i=1}^n$ independent and identically distributed;

Additionally, define the null errors $\epsilon_i(\beta_0) \coloneqq y_i - x_i\beta_0$.

Remark.

All results hold conditionally on a realization of the instruments. Having described basic model, treat them as fixed moving forward.

Testing Procedure

Proposed test statistic is similar in spirit to, but structurally distinct from, a jackknife version of the K-statistic (Kleibergen, 2002, Kleibergen, 2005).

- Kleibergen K-statistic influential in “low-dimensional” literature.
- Adaptating Kleibergen idea for setting with large instruments requires significant modification of test statistic.

Ideal Test Statistic

Ideal test would use first stage to test null hypothesis.

$$\text{Ideal}(\beta_0) := \frac{\left(\sum_{i=1}^n \epsilon_i(\beta_0)\Pi_i\right)^2}{\sum_{i=1}^n \epsilon_i^2(\beta_0)\Pi_i^2} \rightsquigarrow \chi^2(1)$$

- Use of first stage leads to an efficient test (Chamberlain, 1987);
- Limiting distribution is straightforward to derive.

Ideal Test Statistic

Since true first stage is unknown, could try to instead estimate $\widehat{\Pi}_i$ using $x_1, \ldots, x_n$;

$$\mathrm{Ideal}(\beta_0) := \frac{\left(\sum_{i=1}^n \epsilon_i(\beta_0)\widehat{\Pi}_i\right)^2}{\sum_{i=1}^n \epsilon_i^2(\beta_0)\widehat{\Pi}_i^2}$$

- Under weak identification, distribution of $\widehat{\Pi}_i$ is relevant to limiting distribution of test statistic;
- Cannot approximate limiting behavior without knowledge of underling DGP.

Kleibergen Idea

Partial out $\epsilon_i(\beta_0)$ from x_i;

$$r_i = x_i - \frac{\text{Cov}(\epsilon_i(\beta_0), x_i)}{\text{Var}(\epsilon_i(\beta_0))} \epsilon_i(\beta_0)$$

Then estimate $\widehat{\Pi}_i$ using OLS of r_i on z_i. Under homoskedasticity, resulting first-stage estimates are uncorrelated with $\epsilon_i(\beta_0)$.

Kleibergen K-Statistic

K-statistic is constructed using these first stage estimates

$$K(\beta_0) := \frac{\left(\sum_{i=1}^{n} \epsilon_i(\beta_0)\widehat{\Pi}_i\right)^2}{\widehat{\mathrm{Var}}(\epsilon(\beta_0)) \sum_{i=1}^{n} \widehat{\Pi}_i^2}$$

To analyze limiting behavior, apply CLT to numerators and denominators and treat variables as if they are normally distributed.

- $\{\widehat{\Pi}_i\}_{i\in[n]} \perp \{\epsilon_i(\beta_0)\}_{i\in[n]}$ since uncorrelated jointly gaussian variables are independent.
- Conditional on $\{\widehat{\Pi}_i\}_{i\in[n]}$, $K(\beta_0) \sim \chi^2(1)$ under H_0; unconditional distribution also $\chi^2(1)$.

Extending Kleibergen

When d_z is large and errors are heteroskedastic run into the following issues

1. Cannot apply CLTs to examine limiting behavior of test statistic,
2. OLS may be poorly behaved or not well defined if d_z is large,
3. Kleibergen (2005) extension for heteroskedastic errors requires estimating a $d_z \times d_z$ matrix
   - Cannot be consistently estimated when d_z large.

Modified Endogenous Variable

I use versions of x_i that are conditionally uncorrelated with $\epsilon_i(\beta_0)$;

$$r_i := x_i - \rho(z_i)\epsilon_i(\beta_0), \quad \rho(z_i) := \frac{\text{Cov}(\epsilon_i(\beta_0), x_i|z_i)}{\text{Var}(\epsilon_i(\beta_0)|z_i)}.$$

Notice that $\text{Cov}(\epsilon_i(\beta_0), r_i|z_i) = 0$. Will use $r_1, \ldots, r_n$ to construct first stage estimates and test statistic.

Estimating Nuisance Parameter

Parameter $\rho(z_i)$ is conditional slope parameter from OLS of x_i on $\epsilon_i(\beta_0)$. Under H_0 it solves the population problem;

$$\rho(z_i) = \arg\min_{\tilde{\rho}(z_i)} \mathbb{E}\left[\left(x_i - \epsilon_i(\beta_0)\tilde{\rho}(z_i)\right)^2\right]$$

If $\rho(z_i) = b(z_i)'\gamma + \xi_i$ for a basis $b(z_i) \in \mathbb{R}^{d_b}$ then

$$\gamma = \arg\min_{\tilde{\gamma}} \mathbb{E}\left[\left(x_i - \epsilon_i(\beta_0)b(z_i)'\tilde{\gamma}\right)^2\right]$$

Estimation

Can estimate γ via

$$\widehat{\gamma} = \arg\min_{\gamma} \frac{1}{n} \sum_{i=1}^{n} (x_i - \epsilon_i(\beta_0) b(z_i)' \gamma)^2 + \lambda \|\gamma\|_1 \tag{1}$$

- Eqn. (1) is a simple LASSO regression of x_i on $\epsilon_i(\beta_0) b(z_i)$.
- $\widehat{\gamma}$ converges to γ under approximate sparsity, even if $d_b \gg n$.
  - If errors are homoskedastic, $\rho(z_i)$ is sparse in any basis with a constant term.

For each $i = 1, \ldots, n$ define $\hat{r}_i := x_i - \epsilon_i(\beta_0) b(z_i)' \widehat{\gamma}$.

Test Statistic

Using $\hat{r}_1, \ldots, \hat{r}_n$ construct a jackknife-linear estimate of the first stage.

$$\widehat{\Pi}_i := \sum_{j \neq i} h_{ij} \hat{r}_j$$

The weights h_{ij} derive from matrix $H \in \mathbb{R}^{n \times n}$ which depends only on the instruments $\mathbf{z} = (z_1', \ldots, z_n')' \in \mathbb{R}^{n \times d_z}$. In paper, take H to be the ridge regression hat matrix.

$$H = \mathbf{z}(\mathbf{z}'\mathbf{z} + \lambda^\star I_{d_z})^{-1}\mathbf{z}$$

However, any other form of H is permissible so long as a balanced design condition is met.

Ridge Penalty Alternate H

Test Statistic

Using $\widehat{\Pi}_1, \ldots, \widehat{\Pi}_n$ construct the test-statistic

$$JK(\beta_0) := \frac{\left(\sum_{i=1}^n \epsilon_i(\beta_0)\widehat{\Pi}_i \right)^2}{\sum_{i=1}^n \epsilon_i^2(\beta_0)\widehat{\Pi}_i^2}$$

The test statistic is similar in spirit to a jackknife version of the K-statistic, but the construction is distinct.

Balanced Design Condition

Balanced design condition requires that the average second moment of the first stage estimators is on the same order as the maximum second moment

$$\text{Balanced Design:} \quad \frac{\max_i \mathbb{E}[(\sum_{j\neq i} h_{ij} r_j)^2]}{\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[(\sum_{j\neq i} h_{ij} r_j)^2]} \quad \text{is bounded from above}$$

- Eliminates H matrices that are all zeroes or nearly all zeroes.
- Ensures that distribution of test statistic is not governed by a single observation.

Verifying Balanced Design

Size Control

Theorem 1

Suppose that moment, balanced design, and estimation assumptions hold. Then, under H_0, $JK(\beta_0) \rightsquigarrow \chi^2(1)$.

Proof Strategy.

Limiting distribution is derived in two main steps.

1. Show CDF of an infeasible statistic, constructed with the true $\rho(z_i)$, can be approximated by CDF of a gaussian analog statistic.
    - Approximation holds in local neighborhoods of null, allows for analysis of local power.
2. Show that the difference between feasible and infeasible statistics converges to zero.
    - Simple statement, but not immediate as some standard tools are lost in first step of argument.

Gaussian Approximation

For each $i \in [n]$, let $(\tilde{\epsilon}_i(\beta_0), \tilde{r}_i)'$ be generated

(a) independently of all other variables in the model and;

(b) with the same mean and covariance matrix as $(\epsilon_i(\beta_0), r_i)'$.

Define $\widehat{\Pi}_i^I = \sum_{j \neq i} h_{ij} r_j$, $\tilde{\Pi}_i = \sum_{j \neq i} h_{ij} \tilde{r}_j$, and

$$JK_I(\beta_0) := \frac{\left(\sum_{i=1}^n \epsilon_i(\beta_0) \widehat{\Pi}_i^I \right)^2}{\sum_{i=1}^n \epsilon_i^2(\beta_0) (\widehat{\Pi}_i^I)^2} \qquad JK_G(\beta_0) := \frac{\left(\sum_{i=1}^n \tilde{\epsilon}_i(\beta_0) \tilde{\Pi}_i \right)^2}{\sum_{i=1}^n \mathbb{E}[\epsilon_i^2(\beta_0)] \tilde{\Pi}_i^2}$$

Uncorrelated normal random variables are independent; under H_0 distribution of $JK_G(\beta_0)$ conditional on $(\tilde{r}_1, \ldots, \tilde{r}_n)$ is $\chi^2(1)$ and so unconditional distribution is also $\chi^2(1)$.

Gaussian Approximation

Let N_I and D_I denote scaled versions of the numerator and denominator $JK_I(\beta_0)$. Likewise, N_G and D_G represent scaled versions of the numerator and denominator of $JK_G(\beta_0)$ so that

$$JK_I(\beta_0) = \frac{N_I}{D_I} \text{ and } JK_G(\beta_0) = \frac{N_G}{D_G}$$

Interpolation argument is difficult on test statistics. Instead for any $a \geq 0$ work with

$$JK_I^a = N_I - aD_I \text{ and } JK_G^a = N_G - aD_G$$

Notice that $\{JK_I(\beta_0) \leq a\} = \{JK_I^a \leq 0\}$ and $\{JK_G(\beta_0) \leq a\} = \{JK_G^a \leq 0\}$.

Gaussian Approximation

One-by-one replace each pair $(\epsilon_i(\beta_0), r_i)$ in the expression of JK_I^a with $(\tilde{\epsilon}_i(\beta_0), \tilde{r}_i)$

$$
\begin{aligned}
JK_I^a = & JK^a\Big((\epsilon_1(\beta_0), r_1), (\epsilon_2(\beta_0), r_2), \ldots, (\epsilon_{n-1}(\beta_0), r_{n-1}), (\epsilon_n(\beta_0), r_n) \Big) \\
& \downarrow \\
& JK^a\Big((\tilde{\epsilon}_1(\beta_0), \tilde{r}_1), (\epsilon_2(\beta_0), r_2), \ldots, (\epsilon_{n-1}(\beta_0), r_{n-1}), (\epsilon_n(\beta_0), r_n) \Big) \\
& \vdots \\
& JK^a\Big((\tilde{\epsilon}_1(\beta_0), \tilde{r}_1), (\tilde{\epsilon}_2(\beta_0), \tilde{r}_2), \ldots, (\tilde{\epsilon}_{n-1}(\beta_0), \tilde{r}_{n-1}), (\epsilon_n(\beta_0), r_n) \Big) \\
& \downarrow \\
& JK^a\Big((\tilde{\epsilon}_1(\beta_0), \tilde{r}_1), (\tilde{\epsilon}_2(\beta_0), \tilde{r}_2), \ldots, (\tilde{\epsilon}_{n-1}(\beta_0), \tilde{r}_{n-1}), (\tilde{\epsilon}_n(\beta_0), \tilde{r}_n) \Big) = JK_G^a
\end{aligned}
$$

And show that <u>sum</u> of one step distributional changes is negligible as $n \to \infty$.

Gaussian Approximation

Apply Glivenko-Cantelli type argument to show approximation holds uniformily over $a \in \mathbb{R}$.

Lemma 1

Assume moment and balanced design assumptions hold. Then, in local neighborhoods of H_0;

$$\sup_{a \in \mathbb{R}} \left| \Pr\left(JK_I(\beta_0) \leq a\right) - \Pr\left(JK_G(\beta_0) \leq a\right) \right| \to 0 \tag{2}$$

Local Neighborhoods and Infeasible Consistency

Managing Estimation Error

Final step is to show that the difference between feasible $JK(\beta_0)$ and infeasible $JK_I(\beta_0)$ is negligible. Define

$$\Delta_N = \text{Scaled Numerator of } JK(\beta_0) - \underbrace{\text{Scaled Numerator of } JK_I(\beta_0)}_{N_I}$$

$$\Delta_D = \text{Scaled Denominator of } JK(\beta_0) - \underbrace{\text{Scaled Denominator of } JK_I(\beta_0)}_{D_I}$$

To argue that $|JK(\beta_0) - JK_I(\beta_0)| \to_p 0$ need to argue

1. $(\Delta_N, \Delta_D)' \to_p 0$;
2. $1/D_I$ is bounded in probability.
   - Difficult since D_I does not have a limiting distribution.

Managing Estimation Error

Need new argument to argue $1/D_I$ bounded in probability. I show that

$$\Pr\left(D_I \leq \delta_n\right) \to 0$$

for any sequence $\delta_n \searrow 0$.

1. Show distribution of D_I can be approximated by distribution of D_G

Managing Estimation Error

Need new argument to argue $1/D_I$ bounded in probability. I show that

$$\Pr\left(D_I \leq \delta_n\right) \to 0$$

for any sequence $\delta_n \searrow 0$.

1. Show distribution of D_I can be approximated by distribution of D_G
2. D_G is a quadratic form in $\tilde{r} := (\tilde{r}_1, \ldots, \tilde{r}_n)'$. Let $\Sigma_0 = \operatorname{diag}(\mathbb{E}[\epsilon_1^2(\beta_0)], \ldots, \mathbb{E}[\epsilon_n^2(\beta_0)])$

$$D_G = \text{Scaling Factor } \times \ \tilde{r}' H \Sigma_0 H' \tilde{r}$$

If $\operatorname{Var}(D_G) \not\to 0$, show Götze et al. (2019) results can be applied to bound density.

Managing Estimation Error

Need new argument to argue $1/D_I$ bounded in probability. I show that

$$\Pr\left(D_I \leq \delta_n\right) \to 0$$

for any sequence $\delta_n \searrow 0$.

1. Show distribution of D_I can be approximated by distribution of D_G
2. D_G is a quadratic form in $\tilde{r} := (\tilde{r}_1, \ldots, \tilde{r}_n)'$. Let $\Sigma_0 = \text{diag}(\mathbb{E}[\epsilon_1^2(\beta_0)], \ldots, \mathbb{E}[\epsilon_n^2(\beta_0)])$

   $$D_G = \text{Scaling Factor } \times \ \tilde{r}' H \Sigma_0 H' \tilde{r}$$

   If $\text{Var}(D_G) \not\to 0$, show Götze et al. (2019) results can be applied to bound density.
3. If $\text{Var}(D_G) \to 0$, use Chebyshev's inequality and $\mathbb{E}[D_G]$ bounded away from zero.

Managing Estimation Error

Lemma 2

Suppose Lemma 1 conditions hold and $(\Delta_N, \Delta_D)' \to_p 0$. Then $|JK(\beta_0) - JK_I(\beta_0)| \to_p 0$.

Theorem 2

Suppose moment, balanced design, and consistency assumptions hold. Then, in strengthened local neighborhoods of H_0,

$$\sup_{a \in \mathbb{R}} \left| \Pr(JK(\beta_0) \le a) - \Pr(JK_G(\beta_0) \le a) \right| \to 0$$

In particular, under H_0, $JK(\beta_0) \rightsquigarrow \chi^2(1)$.

Table of Contents

Power Properties

Theorem 2 allows us to examine, in local neighborhoods of H_0, the limiting behavior of $JK(\beta_0)$ through the gaussian analog statistic $JK_G(\beta_0)$. Conditional on $\tilde{\Pi} := (\tilde{\Pi}_1, \ldots, \tilde{\Pi}_n)$;

$$JK_G(\beta_0) \sim A(\tilde{\Pi}) \cdot \underbrace{\chi^2(1; \mu(\tilde{\Pi}))}_{\text{Noncentral } \chi^2(1)}$$

where

$$A(\tilde{\Pi}) = \frac{\sum_{i=1}^n \mathrm{Var}(\epsilon_i(\beta_0))\tilde{\Pi}_i^2}{\sum_{i=1}^n \{\tilde{\Pi}_i^2(\beta - \beta_0)^2 + \mathrm{Var}(\epsilon_i(\beta_0))\}\tilde{\Pi}_i^2}$$

$$\mu^2(\tilde{\Pi}) = (\beta - \beta_0)^2 \frac{\left(\sum_{i=1}^n \Pi_i \tilde{\Pi}_i\right)^2}{\sum_{i=1}^n \{\Pi_i^2(\beta - \beta_0)^2 + \mathrm{Var}(\epsilon_i(\beta_0))\}\tilde{\Pi}_i^2}$$

Power Properties

Under local alternatives, $\Pi_i^2(\beta - \beta_0)^2 \to 0$ so that $A(\tilde{\Pi}) \to 1$ and $|\mu^2(\tilde{\Pi}) - \mu_\infty^2(\tilde{\Pi})| \to 0$ where

$$\mu_\infty^2(\tilde{\Pi}) = (\beta - \beta_0)^2 \frac{\left(\sum_{i=1}^n \Pi_i \tilde{\Pi}_i\right)^2}{\sum_{i=1}^n \mathrm{Var}(\epsilon_i(\beta_0))\tilde{\Pi}_i^2}$$

Will analyze power through $\mu_\infty^2(\tilde{\Pi})$.

1. Numerator of $\tilde{\mu}_\infty^2(\tilde{\Pi})$ suggests that power is maximized when the first stage estimate, $\tilde{\Pi}_i$, is close to the true first stage, Π_i.
   - Reflects efficiency bound of Chamberlain (1987).
2. Denominator of $\tilde{\mu}_\infty^2(\tilde{\Pi})$ suggests having estimates, $\tilde{\Pi}_i$, with smaller second moments may increase power.
   - Guides recommendation of ridge regression to construct $\widehat{\Pi}_i$.

Power Properties

Unfortunately, estimates of Π_i based on $r = (r_1, \ldots, r_n)$ may be biased as mean of r_i differs from that of x_i under H_1;

$$\mathbb{E}[r_i] = \Pi_i + \rho(z_i)\Pi_i(\beta - \beta_0)$$

Bias is particularly adverse when $(\beta - \beta_0) = -1/\rho(z_i)$ in which case $\mathbb{E}[r_i] = 0$.

- In "low-dimensional" literature, this is dealt with by combining K-statistic with Anderson-Rubin based on conditioning statistic;
  - Moreira (2003), Kleibergen (2005), Andrews (2016).
- Will take a similar approach, but need to find correct conditioning and mixing statistics.

Combination Test

Combine $JK(\beta_0)$ test with sup-score test of Belloni et al. (2012). Level $(1-\alpha)$ sup-score test rejects if

$$S(\beta_0) := \sup_{\ell \in [d_z]} \left| \frac{\sum_{i=1}^n \epsilon_i(\beta_0) z_{\ell i}}{\left(\sum_{i=1}^n z_{\ell i}^2 \right)^{1/2}} \right|$$

is larger than the bootstrap critical value;

$$c_{1-\alpha}^S := (1-\alpha) \text{ quantile of } \sup_{\ell \in [d_z]} \left| \frac{\sum_{i=1}^n e_i \epsilon_i(\beta_0) z_{\ell i}}{\left(\sum_{i=1}^n z_{\ell i}^2 \right)^{1/2}} \right| \text{conditional on } \{(y_i, x_i, z_i)\}_{i=1}^n$$

where $e_1, \ldots, e_n$ are i.i.d standard normals generated independently of the data.

Combination Test

Sup-score does not incorporate first-stage information but does not suffer power decline in any particular direction. Combination test decides whether to run sup-score or jackknife K based on conditioning statistic.

$$C = \sup_{i \in [n]} \left| \frac{\sum_{j \neq i} h_{ij} \hat{r}_j}{\left(\sum_{j \neq i} h_{ij}^2 \right)^{1/2}} \right|.$$

Conditioning statistic attempts to detect whether $\mathbb{E}[\widehat{\Pi}_i^I] = 0$ for all $i \in [n]$;

Combination Test

Combination test can be summarized by threshold τ;

$$T(\beta_0;\tau) = \begin{cases} \mathbf{1}\left\{S(\beta_0) > c^S_{1-\alpha}\right\} & \text{if } C \leq \tau \\ \mathbf{1}\left\{JK(\beta_0) > \chi^2_{1-\alpha}(1)\right\} & \text{otherwise} \end{cases}$$

where $\chi^2_{1-\alpha}(1)$ is the $(1-\alpha)$ quantile of the $\chi^2(1)$ distribution. In practice, take τ to be the 75$^{\text{th}}$ quantile of conditioning statistic under assumption that $\mathbb{E}[\widehat{\Pi}^I_i] = 0$ for all $i \in [n]$.

Theorem 3

Suppose the conditions of Theorem 2 hold along with strengthened moment and balanced design conditions. Further, assume $\log^M(d_z n)/n \to 0$ for a defined constant M. Then, the test $T(\beta_0;\tau)$ has asymptotic size α for any choice of cutoff τ.

Simulating Quantile

Combination Test

Proof of Theorem 3 follows the basic structure;

1. Establish that quantiles of $(JK(\beta_0), S(\beta_0), C)$ can be jointly uniformly approximated by quantiles of gaussian analogs $(JK_G(\beta_0), S_G(\beta_0), C_G)$.

Combination Test

Proof of Theorem 3 follows the basic structure;

1. Establish that quantiles of $(JK(\beta_0), S(\beta_0), C)$ can be jointly uniformly approximated by quantiles of gaussian analogs $(JK_G(\beta_0), S_G(\beta_0), C_G)$.
2. Under H_0, $JK_G(\beta_0) \perp C_G$ and $S_G(\beta_0) \perp C_G$.

Combination Test

Proof of Theorem 3 follows the basic structure;

1. Establish that quantiles of $(JK(\beta_0), S(\beta_0), C)$ can be jointly uniformly approximated by quantiles of gaussian analogs $(JK_G(\beta_0), S_G(\beta_0), C_G)$.
2. Under H_0, $JK_G(\beta_0) \perp C_G$ and $S_G(\beta_0) \perp C_G$.
3. Using independence, does not matter if we look at C before deciding to run $JK(\beta_0)$ or $S(\beta_0)$ test.
   - Only use marginal independence for thresholding test. More sophisticated combinations would require joint independence; $(JK_G(\beta_0), S_G) \perp C_G$.

Simulation Study

I present simulated power curves following a DGP similar to that of Matsushita and Otsu (2022). Main features:

1. Heteroskedastic laplacian errors (ϵ_i, v_i)
   - Parameter ϱ controls degree of endogeneity, with $\varrho = 0$ indicating $\mathbb{E}[\epsilon_i v_i] = 0$.
2. Using interactions, quadratic, and cubic powers of 10 initial instruments generate total of 75 instruments.
   - Initial instruments generated multivariate normal with toeplitz covariance structure.
3. Model intermediate identification by dividing first stage signal by $n^{1/3}$, for $n = 500$.

I compare performance of Jackknife K-test, Combination test, Anderson-Rubin test, and Jackknife LM test.

Intro

Simulation Study

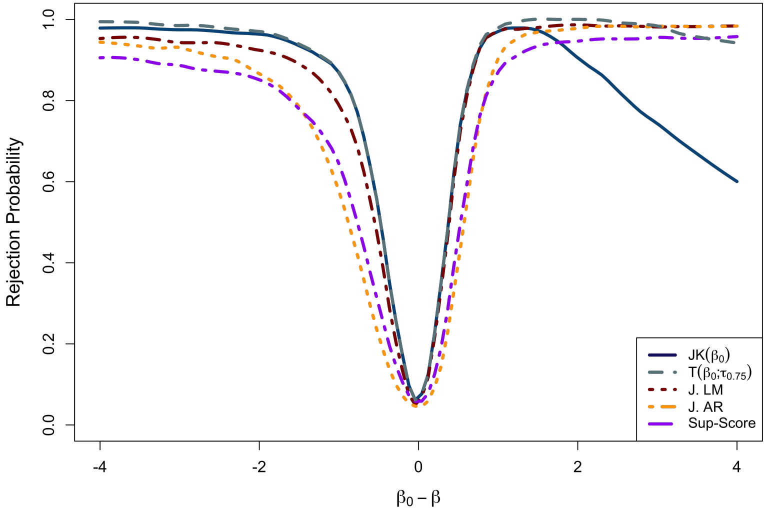


Figure 1: Calibrated Power Curves under intermediate identification strength with $d_z = 75$, $\varrho = 0.3$, and $n = 500$

Simulation Study

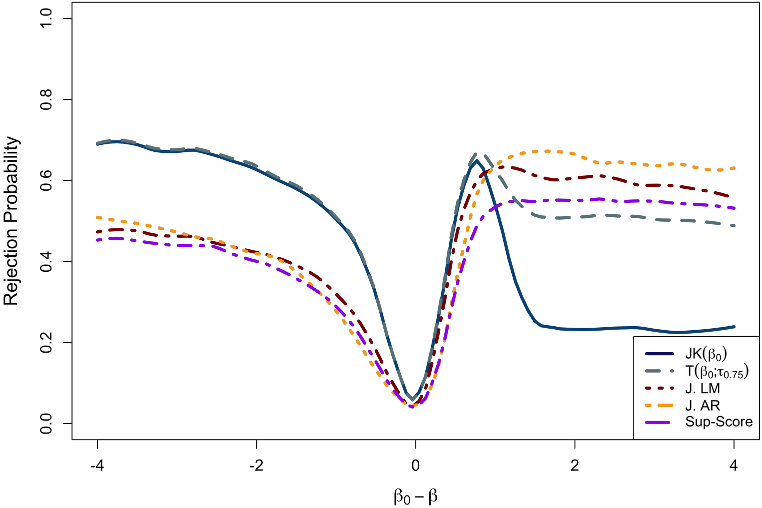


Figure 2: Calibrated Power Curves under intermediate identification strength with $d_z = 75$, $\varrho = 0.5$, and $n = 500$

Table of Contents

Setting

I apply the proposed testing procedures to the data of Gilchrist and Sands (2016). The data consists of 1671 opening weekend days from 2002 to 2012. For each weekend day, i, we observe

- The total sales of wide-released movies $7w$ days after opening weekend day i, for $w = 0, \ldots, 5$.
- A vector of 52 weather related instrumental variables consisting of, for each Saturday and Sunday of the corresponding opening weekend:
  - The proportion of national movie theaters experiencing a maximum temperature in one of sixteen 5° temperature bins from $[10^\circ, 100^\circ]$.
  - The proportion of national movie theaters experiencing maximal hourly precipitation in one of six $0.25''$ precipitation bins from $[0'', 1.5'']$.
  - The proportion of national movie theaters experiencing any sort of rainfall.
  - The proportion of national movie theaters experiencing any sort of snowfall.
- A vector of date controls to control for seasonality in movie viewership.

Setting

Interested in spillover effects on sales in later weeks from a strong opening weekend.

Formally, interested in parameters β_w for $w = 1, \ldots, 6$ from the linear model

$$\text{Sales}^{\perp}_{wi} = \beta_w \text{Sales}^{\perp}_{0i} + \epsilon_{wi} \tag{3}$$

where

- $\text{Sales}^{\perp}_{0i}$ represents the sales of newly-released movies on opening weekend day i, after partialling out date controls and a constant.
- For $w = 1, \ldots, 5$, $\text{Sales}^{\perp}_{wi}$ represents the sales of the same movies $7w$ days after opening day i, after partialling out date controls and a constant.
- $\text{Sales}^{\perp}_{6i} = \sum_{w=1}^{5} \text{Sales}^{\perp}_{wi}$.

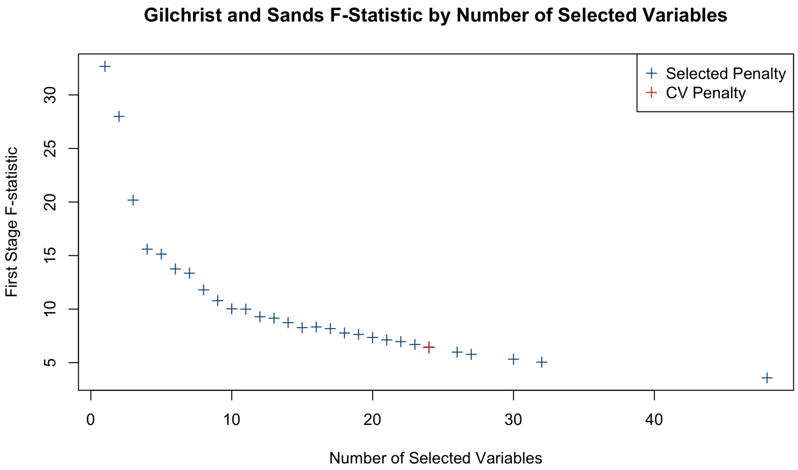
Initial Instrument Selection

In their main analysis, the authors set LASSO penalty to select either one, two, or three instruments. Then run two stage least square on selected instruments.

Instruments Selected	First Stage F-stat.
One Instrument	38.30
Two Instruments	25.86
Three Instruments	20.95
All Instruments	**3.804**

Identification seems strong when using selected instruments, but weak when using all instruments.

Initial Instrument Selection



Initial Instrument Selection

Moreover, it is unclear whether F-statistic on selected instruments is interpretable. To demonstrate, I show results from a simple simulated exercise.

- Start with 10 independent weakly relevant instruments
- Generate additional 55 irrelevant instruments by taking all square terms and interactions.
- Set LASSO penalty to select only a certain number of instruments.

Run this simulation 1000 times with $n = 1000$ and report results from using 2SLS on selected vs. relevant instruments.

Initial Instrument Selection

Compare average F-statistic and 95% Confidence Interval coverage probability from using selected instruments to using oracle estimator which already knows the relevant instruments.

	Selected Instruments		*Oracle Estimator*	
Number of Instruments	F-stat.	Coverage Prob.	F-stat.	Coverage Prob.
One Instrument	12.539	0.302	4.911	0.904
Two Instruments	11.185	0.150	5.040	0.830
Three Instruments	10.060	0.070	4.820	0.810

Coverage with LASSO selected instruments is much worse despite F-statistics being significantly higher.

Identification Robust Results

Given lack of clarity on identification strength, I revisit the analysis of Gilchrist and Sands (2016) using the identification robust tests proposed above. I invert the proposed tests to generate confidence intervals for parameters $\beta_w, w = 1, \ldots, 6$

$$\text{Sales}_{wi}^{\perp} = \beta_w \text{Sales}_{0i}^{\perp} + \epsilon_{wi}$$

Identification Robust Results

		95% Confidence Interval		
Param.	Estimate	Original	$JK(\beta_0)$	$S(\beta_0)$
β_1	0.475	$[0.428, 0.522]$	$[0.436, 0.557]$	$\emptyset$
β_2	0.269	$[0.223, 0.314]$	$[0.227, 0.334]$	$[0.294, 0.334]$
β_3	0.164	$[0.131, 0.197]$	$[0.134, 0.214]$	$[0.087, 0.094]$
β_4	0.121	$[0.096, 0.146]$	$[0.100, 0.167]$	$\emptyset$
β_5	0.093	$[0.073, 0.113]$	$[0.080, 0.134]$	$\emptyset$
β_6	1.222	$[1.077, 1.367]$	$[1.003, 1.391]$	$[0.990, 1.518]$

Table 2: Confidence Intervals Using 48 Linearly Independent Instruments

In main specification, difference between size of $S(\beta_0)$ CI and $JK(\beta_0)$ CI is almost 1.5x difference between length of $JK(\beta_0)$ CI and original.

Identification Robust Results

Repeat analysis including interactions between temperature instruments and all other instruments. Resulting confidence intervals are wider than before.

		95% Confidence Interval		
Param.	Estimate	Original	$JK(\beta_0)$	$S(\beta_0)$
β_1	0.475	$[0.428, 0.522]$	$[0.443, 0.604]$	$[0.416, 0.477]$
β_2	0.269	$[0.223, 0.314]$	$[0.215, 0.342]$	$\emptyset$
β_3	0.164	$[0.131, 0.197]$	$[0.094, 0.228]$	$\emptyset$
β_4	0.121	$[0.096, 0.146]$	$[0.087, 0.154]$	$[0.034, 0.121]$
β_5	0.093	$[0.073, 0.113]$	$[0.054, 0.121]$	$[0.121, 0.208]$
β_6	1.222	$[1.077, 1.367]$	$[0.916, 1.435]$	$[0.918, 1.562]$

Table 3: Confidence Intervals Using 524 Instruments

Table of Contents

General Case

To this point, have considered a single endogenous variable. Now consider the general case where $x_i = (x_{1i}, \ldots, x_{d_x i})' \in \mathbb{R}^{d_x}$.

$$\begin{aligned} y_i &= x_i'\beta + \epsilon_i \\ x_{1i} &= \Pi_{1i} + v_{1i} \\ &\vdots \\ x_{d_x i} &= \Pi_{d_x i} + v_{d_x i} \end{aligned} \qquad \mathbb{E}[\epsilon_i | z_i] = \mathbb{E}[v_i | z_i] = 0$$

For each $\ell \in [d_x]$, construct $\epsilon_i(\beta_0) := y_i - x_i'\beta_0$ and

$$r_{\ell i} := x_{\ell i} - \epsilon_i(\beta_0)\rho_\ell(z_i), \quad \rho_\ell(z_i) := \frac{\mathrm{Cov}(\epsilon_i(\beta_0), x_{\ell i} | z_i)}{\mathrm{Var}(\epsilon_i(\beta_0) | z_i)}.$$

Condition analysis on realization of instruments $\mathbf{z} := (z_1', \ldots, z_n')'$.

Test Statistic

For each $i \in [n]$ and $\ell \in [d_x]$, estimate $\hat{r}_{\ell i}$ and use hat matrix H to construct

$$\widehat{\Pi}_{\ell i} = \sum_{j \neq i} h_{ij} \hat{r}_{\ell j}$$

Collect $\widehat{\Pi}_i = (\widehat{\Pi}_{1i}, \ldots, \widehat{\Pi}_{d_x i}) \in \mathbb{R}^{d_x}$ and define the vectors

$$\begin{aligned}
\epsilon(\beta_0) &= \left(\epsilon_1(\beta_0), \ldots, \epsilon_n(\beta_0)\right)' \in \mathbb{R}^n \\
\widehat{\Pi} &= \left(\widehat{\Pi}_1', \ldots, \widehat{\Pi}_n'\right)' \in \mathbb{R}^{n \times d_x} \\
\widehat{\Pi}_\epsilon &= \left(\epsilon_1(\beta_0)\widehat{\Pi}_1', \ldots, \epsilon_n(\beta_0)\widehat{\Pi}_n'\right)' \in \mathbb{R}^{n \times d_x}
\end{aligned}$$

Test statistic can then be expressed

$$JK(\beta_0) := \epsilon(\beta_0)'\widehat{\Pi}\left(\widehat{\Pi}_\epsilon'\widehat{\Pi}_\epsilon\right)^{-1}\widehat{\Pi}'\epsilon(\beta_0)$$

Limiting Distribution

Want to apply same argument as before, showing that quantiles of $JK(\beta_0)$ can be approximated by quantiles of a gaussian analog statistic

$$JK_G(\beta_0) := \tilde{\epsilon}(\beta_0)'\tilde{\Pi}\left(\tilde{\Pi}_\epsilon'\tilde{\Pi}_\epsilon\right)^{-1}\tilde{\Pi}'\tilde{\epsilon}(\beta_0)$$

where $\tilde{\Pi}, \tilde{\Pi}_\epsilon$, and $\tilde{\epsilon}(\beta_0)$ are constructed using $(\tilde{\epsilon}_i(\beta_0), \tilde{r}_{1i}, \ldots, \tilde{r}_{d_x i})'$ generated

- Independently of data and across indices
- According to a gaussian distribution with the same mean and covariance matrix as $(\epsilon_i(\beta_0), r_{1i}, \ldots, r_{d_x i})'$

Under H_0, $JK_G(\beta_0) \sim \chi^2(d_x)$.

Limiting Distribution

Showing approximation when $d_x > 1$ requires a modification of original argument

- When $d_x = 1$, relied on observation that

$$JK(\beta_0) = \frac{N}{D} \leq a \iff JK^a = N - aD \leq 0$$

  In general, have to work directly with form of test statistic.

Limiting Distribution

Showing approximation when $d_x > 1$ requires a modification of original argument

- When $d_x = 1$, relied on observation that

$$JK(\beta_0) = \frac{N}{D} \leq a \iff JK^a = N - aD \leq 0$$

  In general, have to work directly with form of test statistic.

- When $\lambda_{\min}(\widehat{\Pi}_\epsilon' \widehat{\Pi}_\epsilon) \approx 0$, derivatives of $JK(\beta_0)$ may become arbitrarily large. This poses a problem for interpolation argument.

Limiting Distribution

Showing approximation when $d_x > 1$ requires a modification of original argument

- When $d_x = 1$, relied on observation that

$$JK(\beta_0) = \frac{N}{D} \leq a \iff JK^a = N - aD \leq 0$$

  In general, have to work directly with form of test statistic.

- When $\lambda_{\min}(\widehat{\Pi}_\epsilon' \widehat{\Pi}_\epsilon) \approx 0$, derivatives of $JK(\beta_0)$ may become arbitrarily large. This poses a problem for interpolation argument.
- To get around this, use a data-dependent choice of functions $\varphi(\cdot)$ to approximate $\mathbf{1}\{\cdot \leq a\}$
  - Data dependent $\varphi(\cdot)$ has small derivatives in directions where $\lambda_{\min}(\widehat{\Pi}_\epsilon' \widehat{\Pi}_\epsilon) \approx 0$.

Limiting Distribution

Showing approximation when $d_x > 1$ requires a modification of original argument

- When $d_x = 1$, relied on observation that

$$JK(\beta_0) = \frac{N}{D} \leq a \iff JK^a = N - aD \leq 0$$

  In general, have to work directly with form of test statistic.

- When $\lambda_{\min}(\widehat{\Pi}'_\epsilon \widehat{\Pi}_\epsilon) \approx 0$, derivatives of $JK(\beta_0)$ may become arbitrarily large. This poses a problem for interpolation argument.
- To get around this, use a data-dependent choice of functions $\varphi(\cdot)$ to approximate $\mathbf{1}\{\cdot \leq a\}$
- Establish anticoncentration bound; $\Pr(\lambda_{\min}(\tilde{\Pi}'_\epsilon \tilde{\Pi}_\epsilon) \leq \delta_n) \to 0$ whenever $\delta_n \searrow 0$.
  - Allows quality of approximation to improve with sample size.

Limiting Distribution

Modified interpolation argument relies on stronger moment conditions.

Theorem 4

Suppose that strengthened moment conditions hold along with generalizations of the balanced design and consistent estimation conditions. Then in local neighborhoods of the null:

$$\sup_{a \in \mathbb{R}} \left| \Pr(JK(\beta_0) \leq a) - \Pr(JK_G(\beta_0) \leq a) \right| \to 0$$

In particular, under H_0, $JK(\beta_0) \rightsquigarrow \chi^2(d_x)$.

Combination Test

Also propose a combination test based on the generalized conditioning statistic

$$C := \min_{1 \le \ell \le d_x} \max_{1 \le i \le n} \left| \frac{\sum_{j \neq i} h_{ij} \hat{r}_j}{\left(\sum_{j \neq i} h_{ij}^2 \right)^{1/2}} \right|$$

After looking at value of C, researcher can decide whether to use sup-score test or test based on $JK(\beta_0)$;

$$T(\beta_0; \tau) = \begin{cases} \mathbf{1}\{S(\beta_0) > c^S_{1-\alpha}\} & \text{if } C \le \tau \\ \mathbf{1}\{JK(\beta_0) > \chi^2_{1-\alpha}(d_x)\} & \text{otherwise} \end{cases}$$

Combination Test

Theorem 4

Suppose that strengthened moment conditions hold along with generalizations of the balanced design and consistent estimation conditions. Then the combination test has asymptotic size α for any choice of cutoff τ.

Conclusion

This paper proposes a new test for the structural parameter in a linear IV model. This proposed test

1. Has exact asymptotic size so long as a nuisance parameter can be consistently estimated. This is possible under approximate sparsity even when $d_z \gg n$ but does not require $d_z \to \infty$.
2. Can be combined with the sup-score test to improve power against certain alternatives.
3. Is shown to perform well in an empirical application and simulation study.

Conclusion

This paper proposes a new test for the structural parameter in a linear IV model. This proposed test

1. Has exact asymptotic size so long as a nuisance parameter can be consistently estimated. This is possible under approximate sparsity even when $d_z \gg n$ but does not require $d_z \to \infty$.
2. Can be combined with the sup-score test to improve power against certain alternatives.
3. Is shown to perform well in an empirical application and simulation study.

Thank you all very much

Ridge Penalty

Following recommendations in Harrell (2015), Wieringen (2023), set ridge penalty parameter so effective degrees if freedom is no more than a fraction of sample size:

$$\lambda^\star := \inf\{\lambda \geq 0 : \text{trace}(\mathbf{z}(\mathbf{z}'\mathbf{z} + \lambda I_{d_z})^{-1}\mathbf{z}' \leq n/5\}$$

Back

Alternate Hat Matrices

Alternate choices of hat matrix could include

1. A "true" jackknife OLS / Ridge,

$$\widehat{\Pi}_i = z_i' \widehat{\phi}_{(-i)}$$

   where $\widehat{\phi}_{(-i)}$ is the OLS / Ridge regression parameter from regressing $r_{(-i)}$ on $\mathbf{z}_{(-i)}$

2. The deleted diagonal projection matrix of Chao et al. (2012) used in Crudu et al. (2021), Mikusheva and Sun (2021), Matsushita and Otsu (2022);

$$[H]_{ij} = [\mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}]_{ij}\mathbf{1}\{i \neq j\}$$

3. Any hat matrix resulting from a preliminary unsupervised learning to reduce the dimensionality of $\mathbf{z}$, such as PCA.

Back

Verifying Balanced Design

A sufficient condition for the balanced design requirement is that there is a fixed quantile $q \in (0, 100)$ such that

$$\frac{q^{\text{th}} \text{ quantile of } \mathbb{E}[(\sum_{j \neq i} h_{ij} r_j)^2]}{\max_i \mathbb{E}[(\sum_{j \neq i} h_{ij} r_j)^2]} \quad \text{is bounded away from zero}$$

Definitions

1. **Approximate Sparsity** Function $\rho(z_i)$ has an approximate sparse representation in basis $b(z_i) \in \mathbb{R}^{d_b}$; there exists a $\gamma \in \mathbb{R}^{d_b}$ such that $\rho(z_i) = b(z_i)'\gamma + \xi_i$ and
   (a) $s = \{j : \gamma_j \neq 0\}$ satisfies $s^2 \log^M(d_b n)n \to 0$
   (b) $(\frac{1}{n}\sum_{i=1}^n \xi_i^2)^{1/2} = o(n^{-1/2})$

   Back

2. **Balanced Design** Let $\widehat{\Pi}_i^I := \sum_{j\neq i} h_{ij} r_j$. Assume that there is a constant $c > 1$ such that

$$\frac{\max_i \mathbb{E}[(\widehat{\Pi}_i^I)^2]}{\frac{1}{n}\sum_{i=1}^n \mathbb{E}[(\widehat{\Pi}_i^I)^2]} \leq c$$

   Plus a technical condition requiring that the hat matrix H is contracted using > 1 effective instrument.

   Back

Limiting Distribution Assumptions

For any $\nu > 0$ and random variable X, define the Orlicz quasi-norm

$$\|X\|_{\psi_\nu} = \inf\{t > 0 : \mathbb{E}\exp(|X|^\nu / t^\nu) \leq 2\}$$

1. **Moment Assumptions** There is a constant $c > 1$ and $\nu \in (0, 1] \cup \{2\}$ such that $\|\epsilon_i\|_{\psi_\nu} \leq c$ and $c^{-1} \leq \mathbb{E}[|\epsilon_i|^l |r_i|^k] \leq c$ for any $i \in [n]$ and $0 \leq l + k \leq 6$.
2. **Balanced Design** Let $\widehat{\Pi}_i^I := \sum_{j \neq i} h_{ij} r_j$. Assume that there is a constant $c > 1$ such that

$$\frac{\max_i \mathbb{E}[(\widehat{\Pi}_i^I)^2]}{\frac{1}{n}\sum_{i=1}^n \mathbb{E}[(\widehat{\Pi}_i^I)^2]} \leq c$$

   Plus a technical condition requiring that the hat matrix H is constructed using > 1 effective instrument.
3. **Consistency** The function $\rho(z_i)$ has an approximately sparse representation in basis $b(z_i)$ and researcher has access to an estimator $\widehat{\gamma}$ that satisfies $\|\widehat{\gamma} - \gamma\|_1 \to_p 0$.

Prop Theorem 2

Infeasible Local Power Assumptions

For any $\nu > 0$ and random variable X, define the Orlicz quasi-norm

$$\|X\|_{\psi_\nu} = \inf\{t > 0 : \mathbb{E}\exp(|X|^\nu/t^\nu) \leq 2\}$$

1. **Moment Assumptions** There is a constant $c > 1$ such that $\mathbb{E}[|\epsilon_i|^l|r_i|^k] \leq c$ for any $i \in [n]$ and $0 \leq l + k \leq 6$.
2. **Balanced Design** Let $\widehat{\Pi}_i^I := \sum_{j\neq i} h_{ij}r_j$. Assume that there is a constant $c > 1$ such that

$$\frac{\max_i \mathbb{E}[(\widehat{\Pi}_i^I)^2]}{\frac{1}{n}\sum_{i=1}^n \mathbb{E}[(\widehat{\Pi}_i^I)^2]} \leq c$$

   Plus a technical condition requiring that the hat matrix H is constructed using > 1 effective instrument.

Back

Local Neighborhoods and Consistency

Define $s_n := \max_i \mathbb{E}[(\widehat{\Pi}_i^I)^2]$ and the local power index P

$$P := \mathbb{E}\left[\left(\frac{s_n}{\sqrt{n}} \sum_{i=1}^{n} \Pi_i \widehat{\Pi}_i^I\right)^2\right]$$

Local neighborhoods are characterized by (i) P being bounded and (ii) a technical condition roughly requiring that $|\mathbb{E}[\epsilon_i(\beta_0)]| \lesssim |\mathbb{E}[r_i]|$.

Proposition 1

If the second condition is satisfied and $P \to \infty$, then the test based on $JK_I(\beta_0)$ is consistent.

Back Consistency Sketch

Consistency Sketch

The consistency result relies on showing that $\Pr(N_I^2 - aD_I \leq 0) \longrightarrow 0$ for any $a \in \mathbb{R}_+$, where N_I is the scaled numerator of $JK_I(\beta_0)$ and D_I is the scaled denominator of $JK_I(\beta_0)$.

1. Scaled denominator is bounded in probability, suffices to show that $\Pr(|N_I| \leq M) \longrightarrow 0$ for any fixed M.
2. Statement $\Pr(|N_I| \leq M) \longrightarrow 0$ for any fixed M follows if $\mathrm{Var}(N_I) = O(1)$ and $\mathbb{E}[N_I^2] \longrightarrow \infty$, since $\mathrm{Var}(|N_I|) = \mathbb{E}[N_I^2] - (\mathbb{E}[|N_I|])^2 \leq \mathrm{Var}(N_I)$.
3. The power index P represents the second moment of the scaled numerator, N_I. Under additional regularity condition, can show that $\mathrm{Var}(N_I) = O(1)$.

Back

Local Neighborhoods

Local Neighborhoods are defined by

1. The local power index P is bounded, $P \leq c$.

$$P := \mathbb{E}\left[\left(\frac{s_n}{\sqrt{n}}\sum_{i=1}^{n}\Pi_i\widehat{\Pi}_i^I\right)^2\right]$$

2. A technical condition roughly requiring that $|\mathbb{E}[\epsilon_i(\beta_0)]| \lesssim |\mathbb{E}[r_i]|$ for all $i \in [n]$.

Back

Strengthened Local Neighborhoods

Local Neighborhoods are defined by

1. The local power index P is bounded, $P \leq c$.

$$P := \mathbb{E}\left[\left(\frac{s_n}{\sqrt{n}}\sum_{i=1}^{n}\Pi_i\widehat{\Pi}_i^I\right)^2\right]$$

2. A technical condition roughly requiring that $|\mathbb{E}[b_\ell(z_i)\epsilon_i(\beta_0)]| \lesssim |\mathbb{E}[r_i]|$ for all $i \in [n]$ and $\ell \in [d_b]$.

Back

Combination Test

In practice, I take τ to be the 40^{th} quantile of conditioning statistic under assumption that $\mathbb{E}[\widehat{\Pi}_i^I] = 0$ for all $i \in [n]$. Simulated;

$$\tau = 40^{\text{th}} \text{ quantile of } \sup_{i \in [n]} \left| \frac{\sum_{j \neq i} e_j h_{ij} \hat{r}_j}{\left(\sum_{j \neq i} h_{ij}^2 \right)^{1/2}} \right| \text{ conditional on } \{y_i, x_i, z_i\}_{i=1}^n$$

where $e_1, \dots, e_n$ are i.i.d standard normal generated independently of the data.

Back

Combination Test Conditions

In addition to the conditions of Theorem 2, assume that there is a constant $c > 1$ such that

1. There is a $\nu \in (0,1] \cup \{2\}$ such that $\|r_i\|_{\psi_\nu} \leq c$;
2. The instruments and hat matrix are balanced in the sense that

$$\max_{\ell,i} \left| \frac{z_{\ell i}}{\left(\frac{1}{n}\sum_{i=1}^{n} z_{\ell i}^2\right)^{1/2}} \right| + \max_{i,j} \left| \frac{h_{ij}}{\left(\frac{1}{n}\sum_{i=1}^{n} h_{ij}^2\right)^{1/2}} \right| \leq c$$

3. $\log^{7+4/\nu}(d_z n) \to 0$.

Back

Empirical Details

- **Wide Released** Displayed in over 600 theaters nationally during its run.
- **Opening Weekend Day** A Friday, Saturday, or Sunday of opening weekend.

Back

Modified Interpolation Conditions

Recall that for any $\nu > 0$ the Orlicz-norm is defined

$$\|X\|_{\psi_\nu} = \min t > 0 : \mathbb{E}\exp(|X|^\nu / t^\nu) \leq 2$$

1. **Moment Conditions** There is a $c > 1$ and $\nu \in (0, 1] \cup \{2\}$ such that $\|\epsilon_i(\beta_0)\|_\nu \leq c$ and $\|r_i\|_\nu \leq c$ and $c^{-1} \leq$ for all $i \in [n]$.
2. **Balanced Design** The general balanced design condition requires that for any $\alpha \in \mathbb{R}^{d_x}$

$$c^{-1} \leq \frac{\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{\ell=1}^{d_x}\alpha_j\widehat{\Pi}^I_{\ell i}\right)^2\right]}{\left(\sum_{\ell=1}^{d_x}\alpha_\ell \max_i \mathbb{E}[(\widehat{\Pi}^I_{\ell i})^2]\right)^2}$$

   as well as a condition requiring that the hat matrix is constructed using more than one "effective" instrument.
3. **Consistency** Each function $\rho_\ell(z_i)$ has an approximately sparse representation in the basis $b(z_i)$, $\rho_\ell(z_i) = b(z_i)'\gamma_\ell + \xi_{\ell i}$, and the researcher has access to an estimators $\widehat{\gamma}_\ell$, $\ell \in [d_x]$, that satisfy $\|\widehat{\gamma}_\ell - \gamma_\ell\|_1 \to_p 0$.

Back

Generalized Local Neighborhood

The multivariate local power index is given

$$P = \sum_{\ell=1}^{d_x} \mathbb{E}\left[\left(\frac{s_{\ell,n}}{\sqrt{n}} \sum_{i=1}^{n} \widehat{\Pi}^{I}_{\ell,i} \Pi_i'(\beta - \beta_0)\right)^2\right]$$

And generalized local neighborhoods are defined by

(a) P is bounded,

(b) A technical condition roughly requiring both $|\mathbb{E}[\epsilon_i(\beta_0)]| \lesssim \mathbb{E}[r_i]$ and $|\mathbb{E}[b(z_i)\epsilon_i(\beta_0)]| \lesssim \mathbb{E}[r_i]$.

Back

References

[1] K. Adusumilli, *Treatment effect estimation in high dimensions without sparsity or collinearity conditions*, tech. rep. (University of Pennsylvania, Department of Economics, 2017).

[2] S. Anatolyev, "Inference in regression models with many regressors", Journal of Econometrics **170**, Thirtieth Anniversary of Generalized Method of Moments, 368–382 (2012).

[3] T. W. Anderson and H. Rubin, "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations", The Annals of Mathematical Statistics **20**, 46–63 (1949).

[4] D. W. K. Andrews and X. Cheng, "Estimation and inference with weak, semi-strong, and strong identification", Econometrica **80**, 2153–2211 (2012).

[5] D. W. K. Andrews and P. Guggenberger, "Identification- and singularity-robust inference for moment condition models", Quantitative Economics **10**, 1703–1746 (2019).

[6] D. W. K. Andrews, M. J. Moreira, and J. H. Stock, "Optimal two-sided invariant similar tests for instrumental variables regression", Econometrica **74**, 715–752 (2006).

[7] D. W. Andrews and J. H. Stock, "Testing with many weak instruments", Journal of Econometrics **138**, 50th Anniversary Econometric Institute, 24–46 (2007).

[8] I. Andrews, "Conditional linear combination tests for weakly identified models", Econometrica **84**, 2155–2182 (2016).

References

[9] P. Bastos, J. Silva, and E. Verhoogen, "Export destinations and input prices", American Economic Review **108**, 353–92 (2018).

[10] P. A. Bekker, "Alternative approximations to the distributions of instrumental variable estimators", Econometrica **62**, 657–681 (1994).

[11] A. Belloni, D. Chen, V. Chernozhukov, and C. Hansen, "Sparse models and methods for optimal instruments with an application to eminent domain", Econometrica **80**, 2369–2429 (2012).

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

[13] G. Chamberlain, "Asymptotic efficiency in estimation with conditional moment restrictions", Journal of Econometrics **34**, 305–334 (1987).

[14] J. C. Chao, N. R. Swanson, J. A. Hausman, W. K. Newey, and T. Woutersen, "Asymptotic distribution of jive in a heteroskedastic iv regression with many instruments", Econometric Theory **28**, 42–86 (2012).

[15] J. C. Chao and N. R. Swanson, "Consistent estimation with a large number of weak instruments", Econometrica **73**, 1673–1692 (2005).

References

[16] S. Chatterjee, "A generalization of the Lindeberg principle", The Annals of Probability **34**, 2061–2076 (2006).

[17] Q. Chen and Z. Fang, "Improved inference on the rank of a matrix", Quantitative Economics **10**, 1787–1824 (2019).

[18] X. Cheng, W. W. Dou, and Z. Liao, "Macro-finance decoupling: robust evaluations of macro asset pricing models", Econometrica **90**, 685–713 (2022).

[19] F. Crudu, G. Mellace, and Z. Sándor, "Inference in instrumental variable models with heteroskedasticity and many instruments", Econometric Theory **37**, 281–310 (2021).

[20] E. Derenoncourt, "Can you move to opportunity? evidence from the great migration", American Economic Review **112**, 369–408 (2022).

[21] E. Gautier and C. Rose, *High-dimensional instrumental variables regression and confidence sets*, 2021.

[22] D. S. Gilchrist and E. G. Sands, "Something to talk about: social spillovers in movie consumption", Journal of Political Economy **124**, 1339–1382 (2016).

[23] F. Götze, A. Naumov, V. Spokoiny, and V. Ulyanov, "Large ball probabilities, Gaussian comparison and anti-concentration", Bernoulli **25**, 2538–2563 (2019).

References

[24] B. S. Graham, "An econometric model of network formation with degree heterogeneity", Econometrica **85**, 1033–1063 (2017).

[25] J. Hahn, "Optimal inference with many instruments", Econometric Theory **18**, 140–168 (2002).

[26] C. Han and P. C. B. Phillips, "Gmm with many moment conditions", Econometrica **74**, 147–192 (2006).

[27] C. Hansen, J. Hausman, and W. Newey, "Estimation with many instrumental variables", Journal of Business & Economic Statistics **26**, 398–422 (2008).

[28] H. Hao and T.-H. Lee, *Boosting gmm with many instruments when some are invalid or irrelevant*, tech. rep. (University of California at Riverside, Department of Economics, 2023).

[29] F. E. Harrell, *Regression modeling strategies, With applications to linear models, logistic and ordinal regression, and survival analysis*, Spring Series in Statistics (Springer Cham, 2015).

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

[31] F. Kleibergen, "Testing parameters in gmm without assuming that they are identified", Econometrica **73**, 1103–1123 (2005).

[32] D. Lim, W. Wang, and Y. Zhang, *A conditional linear combination test with many weak instruments*, 2022.

References

[33] J. W. Lindeberg, "Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung", Mathematische Zeitschrift **15**, 211–225 (1922).

[34] Y. Matsushita and T. Otsu, "A jackknife lagrange multiplier test with many weak instruments", Econometric Theory, 1–24 (2022).

[35] A. Mikusheva, "Many weak instruments in time series econometrics", Working Paper (2023).

[36] A. Mikusheva and L. Sun, "Inference with many weak instruments", The Review of Economic Studies **89**, 2663–2686 (2021).

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

[38] C. R. Nelson and R. Startz, "Some further results on the exact small sample properties of the instrumental variable estimator", Econometrica **58**, 967–976 (1990).

[39] W. K. Newey and F. Windmeijer, "Generalized method of moments with many weak moment conditions", Econometrica **77**, 687–719 (2009).

[40] P. C. Phillips and W. Y. Gao, "Structural inference from reduced forms with many instruments", Journal of Econometrics **199**, The Creative Mind in Econometrics: Studies in Celebration of Robert Basmann's 90th Year on Causation, Identification and Structural Equation Estimation, 96–116 (2017).

References

[41] D. Pouzo, "Bootstrap consistency for quadratic forms of sample averages with increasing dimension", Electronic Journal of Statistics **9**, 3046–3097 (2015).

[42] D. Staiger and J. H. Stock, "Instrumental variables regression with weak instruments", Econometrica **65**, 557–586 (1997).

[43] J. Stock and M. Yogo, *Testing for weak instruments in linear iv regression*, edited by D. W. Andrews (Cambridge University Press, New York, 2005), pp. 80–108.

[44] J. H. Stock and J. H. Wright, "Gmm with weak identification", Econometrica **68**, 1055–1096 (2000).

[45] J. Wang and E. Zivot, "Inference on structural parameters in instrumental variables regression with weak instruments", Econometrica, 1389–1404 (1998).

[46] W. N. van Wieringen, *Lecture notes on ridge regression*, 2023.