An Identification and Dimensionality Robust Test for Linear IV

Manu Navjeevan

TEXAS A&M UNIVERSITY

December, 2024

Consider a heteroskedastic linear IV model;

$$y_i = x'_i \beta + z'_{1i} \Gamma + \epsilon_i, \quad \mathbb{E}[\epsilon_i | z_i] = 0$$

Researcher observes $(y_i, x_i, z_i)'$ and is interested in testing $H_0: \beta = \beta_0 \text{ vs } H_1: \beta \neq \beta_0$.

- 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.
 - Potentially high-dimensional, $d_z \gg n$, and weakly related to x_i .

When weak instruments are suspected, want to use identification robust tests. Validity of these tests rely on alternate assumptions about number of instruments.

- Low Dimensional: Staiger and Stock (1997), Kleibergen (2002), Moreira (2003).
 - Analyses treat *d_z* as fixed or growing slowly with sample size. Tests control size under heteroskedasticity when *d²_z*/*n* → 0 (Andrews and Stock, 2007).

When weak instruments are suspected, want to use identification robust tests. Validity of these tests rely on alternate assumptions about number of instruments.

- Low Dimensional: Staiger and Stock (1997), Kleibergen (2002), Moreira (2003).
 - Analyses treat *d_z* as fixed or growing slowly with sample size. Tests control size under heteroskedasticity when *d²_z*/*n* → 0 (Andrews and Stock, 2007).
- Many Instruments: Mikusheva and Sun (2021), Matsushita and Otsu (2022).
 - Allow $d_z/n \to \varrho \in [0, 1)$ but use Chao et al. (2012) CLT that requires $d_z \to \infty$.
 - If $d_z \rightarrow \infty$ slowly, finite sample size control can be poor Lim et al. (2024) .

When weak instruments are suspected, want to use identification robust tests. Validity of these tests rely on alternate assumptions about number of instruments.

- Low Dimensional: Staiger and Stock (1997), Kleibergen (2002), Moreira (2003).
 - Analyses treat *d_z* as fixed or growing slowly with sample size. Tests control size under heteroskedasticity when *d²_z*/*n* → 0 (Andrews and Stock, 2007).
- Many Instruments: Mikusheva and Sun (2021), Matsushita and Otsu (2022).
 - Allow $d_z/n \to \varrho \in [0, 1)$ but use Chao et al. (2012) CLT that requires $d_z \to \infty$.
 - If $d_z \rightarrow \infty$ slowly, finite sample size control can be poor Lim et al. (2024).
- High Dimensional: Belloni et al. (2012), Mikusheva (2023).
 - $d_z \gg n$ allowed under strong identification but limited work when identification is weak.
 - Unclear how to pre-test for weak-IV when using Lasso/ML first stage.

Motivation

Even when $d_z < n$, these assumptions can be difficult to interpret.

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

•
$$d_z^3 = 1,000,000 \gg n = 45,659;$$

• $d_z = 100 \xrightarrow{?} \infty.$

- 2. Gilchrist and Sands (2016). Something to Talk About, JPE.
 - $d_z^3 \approx 140,000 \gg n = 1,671;$ • $d_z = 52 \xrightarrow{?} \infty.$
- 3. Paravisini et al. (2014). Dissecting the Effect of Credit on Trade, ReStud.
 - $d_z^3 = 1,000 \propto n = 5,996;$ • $d_z = 10 \xrightarrow{?} \infty.$
- 4. Derenoncourt (2022). Can You Move to Opportunity? AER.
 - $d_z^3 = 729 > n = 239;$ • $d_z = 9 \xrightarrow{?} \infty.$

- 1. Propose a new test that 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 $d_z \rightarrow \infty$.
- 3. To improve power against certain alternatives, I propose a combination with sup-score test of Belloni et al. (2012).

Contribution

п	d_z	Q	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

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

- Weak Identification: Nelson and Startz (1990), Bound et al. (1995), Stock and Wright (2000), Ahmad et al. (2001), Andrews et al. (2006), Hansen et al. (2008), Chaudari and Zivot (2009), Andrews and Cheng (2012), Cheng (2008, 2015), Chaudari et al. (2014), Andrews and Guggenberger (2019), 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), Gold et al. (2020), Lim et al. (2022), Fan et al. (2023).
- Gaussian Approximation: Lindeberg (1922), Chatterjee (2006), Pouzo (2015), Chernozhukov et al. (2013, 2017), Graham (2017).

Test Statistic

Power Properties

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

$$y_i = x_i \beta + \epsilon_i$$

$$x_i = \underbrace{\mathbb{E}[x_i \mid z_i]}_{\Pi_i} + v_i \qquad \mathbb{E}[(\epsilon_i, v_i)' \mid z_i] = 0$$

- Controls *z*_{1*i*} 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 $\epsilon_i(\beta_0) \coloneqq y_i - x_i\beta_0$.

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

$$y_i = x_i \beta + \epsilon_i$$

$$x_i = \underbrace{\mathbb{E}[x_i \mid z_i]}_{\Pi_i} + v_i \qquad \mathbb{E}[(\epsilon_i, v_i)' \mid z_i] = 0$$

- Controls *z*_{1*i*} 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.

Ideal test would use first stage to test null hypothesis (Chamberlain, 1987).

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

Unfortunately, plugging in a naive estimate of the first-stage is not viable when identification is weak.

• Distribution of first-stage estimate becomes relevant to distribution of test statistic; cannot approximate limiting behavior without knowledge of underlying DGP.

Instead, borrow an idea from Kleibergen (2002, 2005) and construct $\widehat{\Pi}_i$ to be uncorrelated with structural errors.

- 1. First partial out null errors, $\epsilon_i(\beta_0)$, from the endogenous variable, x_i , using a nuisance function.
 - Nuisance function is simple to estimate. Consistency achievable with standard methods when $d_z \ll n$ or using ML methods when $d_z \gg n$.
 - If errors are homoskedastic, nuisance function is a constant.

Nuisance Parameter

- 2. Construct $\widehat{\Pi}_i$ via a jackknife-ridge regression of "partialled-out" endogenous variable on instruments.
 - Combining jackknife-and partialling out approaches leaves Π
 _i independent of ε_i(β₀) and uncorrelated with ε_j(β₀).
 - First-stage estimate not required to be consistent so in principle other options are available. Ridge allows $d_z \gg n$.

Test statistic is then constructed

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

Key Idea: If all variables were normally distributed, could condition on $\{\widehat{\Pi}_i\}_{i=1}^n$ and see that $JK(\beta_0) \approx \chi^2(1)$ under H_0 .

Typically justify treating variables as if they were normally distributed through CLT(s) and CMT. However, these standard tools are not applicable here:

- Main Issue. Allow $d_z \gg n$ so central limit theorems are not applicable.
 - Even if $d_z \ll n$, denominator looks like a randomly weighted quadratic form, so unclear how to proceed.
- Secondary. First-stage estimates may not be consistent, behavior of numerator and denominator can both "jump around" in limit.
 - Prevents application of CMT. Also an issue when controlling estimation error.

Instead, apply modifications of Lindeberg's interpolation argument (Lindeberg, 1922).

 Basic Idea: One-by-one replace terms in expression of *JK*(β₀) with gaussian analogs and bound resulting one-step distributional changes.

Interpolation argument needs to be modified to accomodate for a "divide-by-zero" problem.

- Standard argument require derivatives w.r.t individual observations to have bounded moments.
- This will not be satisfied under weak identification since denominator of test statistic can be arbitrarily close to zero.
- When *d_x* = 1, problem can be simplified, but when *d_x* > 1 more involved argument is developed.

Let $JK_G(\beta_0)$ be the version of the test statistic constructed with jointly Gaussian variables.

Theorem 1

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

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

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

Test Statistic

Power Properties

Process of partialling out structural errors introduces bias into first stage estimates. Against certain alternatives first stage signal is "erased", $\mathbb{E}[\widehat{\Pi}_i] = 0 \ \forall i \in [n]$, and test has trivial power.

- 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 conditioning and mixing statistics.

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

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

is larger than a bootstrap critical value $c_{1-\alpha}^{S}$.

- Sup-score test does not inorporate first-stage information but does not face a power decline against particular alternatives.
- Has correct asymptotic size even when $d_z \gg n$.

Conditioning statistic, *C*, attempts to detect whether first stage signal is erased.

$$C = \sup_{i \in [n]} \left| \frac{1}{\sqrt{n}} \widehat{\Pi}_i \right|$$

Combination test decides which test to run based on value of conditioning statistic.

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

In practice, take τ to be the 75th quantile of conditioning statistic under assumption that first-stage signal is erased.

Conditioning Statistic Quantiles

Theorem 2

Suppose the conditions of Theorem 1 hold along with strengthened moment and balanced design \square conditions. Further, assume $\log^{M}(d_{z}n)/n \rightarrow 0$ for a defined constant *M*. Then, the test $T(\beta_{0}; \tau)$ has asymptotic size α for any choice of cutoff τ .

Proof establishes joint Gaussian approximation of test and conditioning statistics. Then uses that gaussian test statistics are marginally independent of conditioning statistic under H_0 .

I apply the proposed testing procedures to the data of Gilchrist and Sands (2016). The data consists of 1,671 opening weekend days 🕫 from 2002 to 2012. For each weekend day, *i*, we observe

- The total sales of wide-released \square movies 7w days after opening weekend day *i*, for w = 0, ..., 5.
- A vector of 52 weather related instrumental variables
- A vector of date controls to control for seasonality in movie viewership.

Interested in spillover effects on sales in later weeks from a strong opening weekend. Formally, interested in parameters β_w for w = 1, ..., 6 from the linear model

$$Sales_{wi} = \beta_w Sales_{0i} + \epsilon_{wi} \tag{1}$$

where

- Sales_{0i} represents the sales of newly-released movies on opening weekend day i
- For w = 1,..., 5, Sales[⊥]_{wi} represents the sales of the same movies 7w days after opening day i
- Sales_{6i} = $\sum_{w=1}^{5}$ Sales_{wi}.

To hand large d_z , authors employ a post-Lasso estimate of the first stage. In paper, I point out that the reported F-statistic can be misleading.

Empirical Application: Results

Parameter	β2	β3	β_4	β_5	β6	β7		
Estimate (s.e.)	0.475 (0.024)	0.269 (0.023)	0.164 (0.017)	0.121 (0.013)	0.093 (0.010)	1.222 (0.074)		
	Initial instrument set, $d_z = 52$							
JK(β ₀)	$ \stackrel{\leftarrow 0.114}{\longleftrightarrow} [0.441, 0.555] $	$ \stackrel{\leftarrow 0.114}{\longleftrightarrow} [0.234, 0.348] $	$ \stackrel{\leftarrow 0.074}{\leftarrow 0.127, 0.201} $	$ \stackrel{\longleftarrow 0.074}{\longleftrightarrow} 0.936, 0.167]$	$ \stackrel{\leftarrow 0.046}{\longrightarrow} [0.0736, 0.120] $	$ \stackrel{\longleftarrow 0.375}{\longleftarrow} 0.989, 1.365]$		
$S(\beta_0)$	Ø	$ \stackrel{\leftarrow 0.033}{\longleftrightarrow} 0.294, 0.328]$	Ø	Ø	Ø	$ \stackrel{\leftarrow 0.561}{\leftarrow} 0.989, 1.551]$		
JLM	$ \stackrel{\leftarrow 0.140}{\longrightarrow} [0.428, 0.569] $	$ \stackrel{\leftarrow 0.127 \longrightarrow}{\leftarrow 0.221, 0.348} $	$ \stackrel{\longleftarrow 0.087}{\longleftrightarrow} 0.134, 0.221] $	$ \stackrel{\longleftarrow 0.074 \longrightarrow}{[0.100, 0.174]} $	$ \stackrel{\longleftarrow 0.060}{\longleftrightarrow} \\ [0.080, 0.140] $	$ \stackrel{\leftarrow 0.441 \longrightarrow}{[0.989, 1.384]} $		
	Initial instruments plus all interactions with temp. instruments, $d_z = 524$							
JK(β ₀)	$\stackrel{\leftarrow 0.040 \longrightarrow}{[0.462, 0.502]}$	$ \stackrel{\leftarrow 0.033}{\longleftrightarrow} 0.268, 0.301] $	$ \stackrel{\leftarrow 0.020 \longrightarrow}{[0.154, 0.174]} $	$ \stackrel{\leftarrow 0.013 \longrightarrow}{[0.094, 0.107]} $	← 0.007 → [0.067, 0.074]	$ \stackrel{\leftarrow 0.120 \longrightarrow}{[1.043, 1.164]} $		
$S(\beta_0)$	$ \stackrel{\longleftarrow 0.047}{\longleftrightarrow} 0.415, 0.462]$	Ø	Ø	$ \stackrel{\leftarrow 0.207}{\longleftrightarrow} 0.040, 0.247]$	$ \stackrel{\leftarrow 0.060}{\longrightarrow} [0.161, 0.221] $	Ø		
JLM	$ \stackrel{\longleftarrow 0.080}{\longleftrightarrow} 0.441, 0.522]$	$ \stackrel{\leftarrow 0.060}{\longleftrightarrow} \\ [0.247, 0.308]$	$ \stackrel{\longleftarrow 0.040}{\longleftrightarrow} 0.147, 0.187]$	$ \stackrel{\longleftarrow 0.027}{\longleftrightarrow} 0.027 \stackrel{\longrightarrow}{\longrightarrow} 0.094, 0.120] $	$ \stackrel{\longleftarrow 0.013}{\longrightarrow} \\ [0.067, 0.080]$	$ \stackrel{\longleftarrow 0.227}{\longleftrightarrow} 0.990, 1.217] $		

Table 2: 95% Confidence Intervals and Interval Lengths in the data of Gilchrist and Sands (2016).

Tighter CIs also seen in second application to data of Angrist and Krueger (1991).

Number of Instruments	JAR	JLM	$JK(\beta_0)$	$S(\beta_0)$
Initial Instrument Set (dz=180)	$ \stackrel{\longleftarrow 0.193 \longrightarrow}{[0.008, 0.201]} $	$ \stackrel{\longleftarrow 0.066}{\longrightarrow} \\ [0.067, 0.133]$	$ \stackrel{\longleftarrow 0.034}{\longrightarrow} [0.067, 0.101] $	Ø
All Interactions $(d_z=1,530)$	$ \stackrel{\leftarrow 0.249}{\longrightarrow} \\ [-0.047, 0.202] $	$ \stackrel{\leftarrow 0.098}{\longrightarrow} \\ [0.025, 0.123]$	$ \stackrel{\leftarrow}{\longleftarrow} 0.034 \longrightarrow \\ [0.008, 0.042] $	Ø

Table 3: 95% Confidence Intervals and Interval Lengths in the data of Angrist and Krueger (1991).

Paper provides explanations for improvements in power:

- 1. $JK(\beta_0)$ statistic makes use of higher quality first stage estimators.
- Individual scores in numerator of *JK*(β₀) statistic are uncorrelated. Variance of score does not need to account for covariance between terms.

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 $d_z \rightarrow \infty$.
- 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.

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 $d_z \rightarrow \infty$.
- 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

Power decline occurs when $\mathbb{E}[\widehat{\Pi}_i] \approx 0$. Conditioning statistic attempts to detect this event

$$C = \max_{i \in [n]} \left| \frac{1}{\sqrt{n}} \widehat{\Pi}_i \right|$$

Quantiles of *C* can be calculated under the assumption that $\mathbb{E}[\widehat{\Pi}_i] = 0$ via multiplier bootstrap procedure.

$$(1 - \theta)$$
 quantile $\approx (1 - \theta)$ quantile of $\max_{i \in n} \left| \frac{1}{\sqrt{n}} \sum_{j \neq i} e_i h_{ij} r_j \right|$ conditional on data

where e_1, \ldots, e_n are iid N(0, 1) bootstrap weights, h_{ij} are the linear weights used to construct $\widehat{\Pi}_i$, and r_1, \ldots, r_n represent the "partialled-out" versions of x_1, \ldots, x_n .

🔁 Back

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

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



For any v > 0 and random variable *X*, define the Orlicz quasi-norm

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

- 1. Moment Assumptions There is a constant c > 1 and $v \in (0, 1] \cup \{2\}$ such that $\|\epsilon_i\|_{\psi_v} \le c$ and $c^{-1} \le \mathbb{E}[|\epsilon_i|^l |r_i|^k] \le c$ for any $i \in [n]$ and $0 \le l + k \le 6$.
- 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 $\hat{\gamma}$ that satisfies $\|\hat{\gamma} - \gamma\|_1 \rightarrow_p 0$.



For any v > 0 and random variable *X*, define the Orlicz quasi-norm

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

- 1. Moment Assumptions There is a constant c > 1 such that $\mathbb{E}[|\epsilon_i|^l |r_i|^k] \le c$ for any $i \in [n]$ and $0 \le l + k \le 6$.
- 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 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 $|\mathbb{E}[\epsilon_i(\beta_0)]| \leq |\mathbb{E}[r_i]|$ for all $i \in [n]$.



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 $|\mathbb{E}[b_{\ell}(z_i)\epsilon_i(\beta_0)]| \leq |\mathbb{E}[r_i]|$ for all $i \in [n]$ and $\ell \in [d_b]$.

🕄 Back

In practice, I take τ to be the 75th 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, \ldots, e_n are i.i.d standard normal generated independently of the data.

🕄 Back

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

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

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

- ¹ K. Adusumilli, Treatment effect estimation in high dimensions without sparsity or collinearity conditions, tech. rep. (University of Pennsylvania, Department of Economics, 2017).
- ² I. Ahmad, X. Chen, and Q. Li, "Model check by kernel methods under weak moment conditions", Computational Statistics & Data Analysis 36, 403–409 (2001).
- ³ S. Anatolyev, "Inference in regression models with many regressors", Journal of Econometrics 170, Thirtieth Anniversary of Generalized Method of Moments, 368–382 (2012).
- 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).

- ⁹ J. D. Angrist and A. B. Krueger, "Does compulsory school attendance affect schooling and earnings?", The Quarterly Journal of Economics **106**, 979–1014 (1991).
- ¹⁰ P. Bastos, J. Silva, and E. Verhoogen, "Export destinations and input prices", American Economic Review 108, 353–92 (2018).
- ¹¹ P. A. Bekker, "Alternative approximations to the distributions of instrumental variable estimators", Econometrica 62, 657–681 (1994).
- ¹² 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).
- ¹³ 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).
- ¹⁴G. Chamberlain, "Asymptotic efficiency in estimation with conditional moment restrictions", Journal of Econometrics 34, 305–334 (1987).
- ¹⁵ 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).

- ¹⁶ J. C. Chao and N. R. Swanson, "Consistent estimation with a large number of weak instruments", Econometrica 73, 1673–1692 (2005).
- ¹⁷ S. Chatterjee, "A generalization of the Lindeberg principle", The Annals of Probability 34, 2061–2076 (2006).
- ¹⁸ E. Derenoncourt, "Can you move to opportunity? evidence from the great migration", American Economic Review **112**, 369–408 (2022).
- ¹⁹ Y. Fan, F. Han, and H. Park, "Estimation and inference in a high-dimensional semiparametric gaussian copula vector autoregressive model", Journal of Econometrics 237, 105513 (2023).
- ²⁰ D. S. Gilchrist and E. G. Sands, "Something to talk about: social spillovers in movie consumption", Journal of Political Economy **124**, 1339–1382 (2016).
- ²¹ D. Gold, J. Lederer, and J. Tao, "Inference for high-dimensional instrumental variables regression", Journal of Econometrics 217, 79–111 (2020).
- ²² B. S. Graham, "An econometric model of network formation with degree heterogeneity", Econometrica 85, 1033–1063 (2017).
- ²³J. Hahn, "Optimal inference with many instruments", Econometric Theory 18, 140–168 (2002).
- ²⁴C. Han and P. C. B. Phillips, "Gmm with many moment conditions", Econometrica 74, 147–192 (2006).

- ²⁵ C. Hansen, J. Hausman, and W. Newey, "Estimation with many instrumental variables", Journal of Business & Economic Statistics 26, 398–422 (2008).
- ²⁶ F. Kleibergen, "Testing parameters in gmm without assuming that they are identified", Econometrica 73, 1103–1123 (2005).
- ²⁷ D. Lim, W. Wang, and Y. Zhang, A conditional linear combination test with many weak instruments, 2022.
- ²⁸ D. Lim, W. Wang, and Y. Zhang, A valid anderson-rubin test under both fixed and diverging number of weak instruments, 2024.
- ²⁹ J. W. Lindeberg, "Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung", Mathematische Zeitschrift 15, 211–225 (1922).
- ³⁰ Y. Matsushita and T. Otsu, "A jackknife lagrange multiplier test with many weak instruments", Econometric Theory, 1–24 (2022).
- ³¹ A. Mikusheva, "Many weak instruments in time series econometrics", Working Paper (2023).
- ³² A. Mikusheva and L. Sun, "Inference with many weak instruments", The Review of Economic Studies 89, 2663–2686 (2021).
- ³³ M. J. Moreira, "A conditional likelihood ratio test for structural models", Econometrica 71, 1027–1048 (2003).

- ³⁴ 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).
- ³⁵ D. Paravisini, V. Rappoport, P. Schnabl, and D. Wolfenzon, "Dissecting the Effect of Credit Supply on Trade: Evidence from Matched Credit-Export Data", The Review of Economic Studies 82, 333–359 (2014).
- ³⁶ D. Pouzo, "Bootstrap consistency for quadratic forms of sample averages with increasing dimension", Electronic Journal of Statistics 9, 3046–3097 (2015).
- ³⁷ D. Staiger and J. H. Stock, "Instrumental variables regression with weak instruments", Econometrica 65, 557–586 (1997).
- ³⁸ J. H. Stock and J. H. Wright, "Gmm with weak identification", Econometrica 68, 1055–1096 (2000).