An Identification and Dimensionality Robust Test for Linear IV

Manu Navjeevan

Texas A&M University

September, 2024

Setup

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 $H_0 : \beta = \beta_0$ 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.

Existing Weak IV Robust Tests

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 $d_z^3/n \to 0$ (Andrews and Stock, 2007).

Existing Weak IV Robust Tests

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 $d_z^3/n \to 0$ (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 $d_z \to \infty$ slowly, finite sample size control can be poor Lim et al. (2024) .

Existing Weak IV Robust Tests

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 $d_z^3/n \to 0$ (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 $d_z \to \infty$ 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

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

Contribution

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 $d_z \to \infty$.
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), 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).

Table of Contents

Model

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

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

- 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 $\epsilon_i(\beta_0) := y_i - x_i\beta_0$.

Model

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

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

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

- 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 $\epsilon_i(\beta_0) := y_i - x_i\beta_0$.

Remark.

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

Ideal Test Statistic

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

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

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.

Test Statistic

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 $\widehat{\Pi}_i$ independent of $\epsilon_i(\beta_0)$ and uncorrelated with $\epsilon_j(\beta_0)$.
   - First-stage estimate not required to be consistent so in principle other options are available. Ridge allows $d_z \gg n$.

Test Statistic

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.

Gaussian Approximation

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 approximation error.

Gaussian Approximation

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

- Basic Idea: One-by-one replace terms in expression of $JK(\beta_0)$ 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.

Managing Estimation Error

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 hold. Then, in local neighborhoods of H_0,

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

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

Table of Contents

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.

Combination Test

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

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

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

- 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 $d_z \gg n$.

Combination Test

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.

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

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

Conditioning Statistic Quantiles

Combination Test

Theorem 2

Suppose the conditions of Theorem 1 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.

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.

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

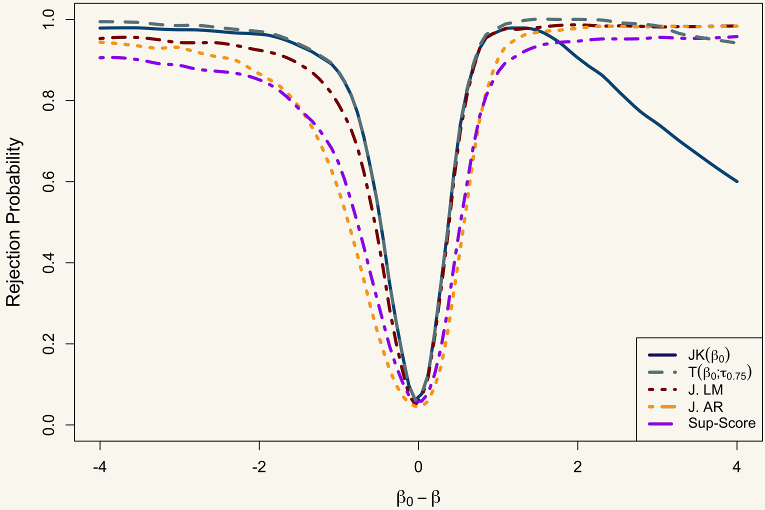

Figure 1: Calibrated Power Curves under intermediate identification strength with $d_z = 75$, $\varrho = 0.3$, and $n = 500$

Simulation Study

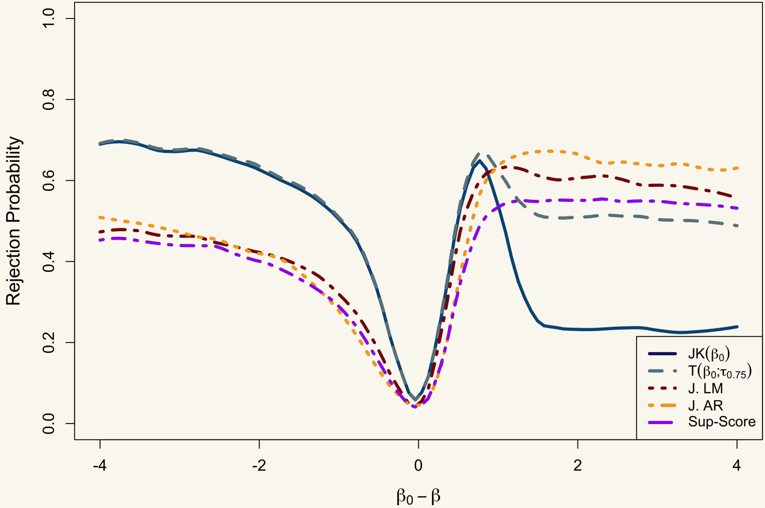

Figure 2: Calibrated Power Curves under intermediate identification strength with $d_z = 75$, $\varrho = 0.5$, and $n = 500$

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

Conditioning Statistic

Power decline occurs when $\mathbb{E}[\widehat{\Pi}_i] = 0$ for all $i \in [n]$. 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.

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

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

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

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

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 $\|\widehat{\gamma} - \gamma\|_1 \rightarrow_p 0$.

Theorem 1

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

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 75^{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 $e_1, \ldots, e_n$ 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

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] 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).

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

[18] 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).

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

[20] D. Gold, J. Lederer, and J. Tao, “Inference for high-dimensional instrumental variables regression”, Journal of Econometrics **217**, 79–111 (2020).

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

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

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

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

References

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

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

[27] D. Lim, W. Wang, and Y. Zhang, *A valid anderson-rubin test under both fixed and diverging number of weak instruments*, 2024.

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

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

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

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

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

[33] 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).

References

[34] 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).

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

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

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