Weak Identification: Causes, Consequences, and Solutions

Isaiah Andrews

Harvard

March 5, 2019

Introduction

What is weak identification?

- In many economic contexts, data contains little information about structural or causal parameters
- When this happens, standard approaches to estimation and inference can break down
  - Estimators may be biased with highly non-normal distributions
  - Tests may have size far from the desired level
  - Non-robust procedures can lead to highly misleading inferences

Introduction

Examples of weak identification

- Best-studied form of weak identification weak instruments. Arises in a number of empirical contexts, e.g. estimating
  - Returns to schooling (Bound, Jaeger, and Baker, 1995)
  - The elasticity of inter-temporal substitution (Yogo, 2004)
  - Taylor rule parameters (Mavroeidis, 2010)
- Weak identification also issue in wide range of nonlinear models, e.g.
  - DSGE models (Canova and Sala, 2009)
  - New Keynesian Phillips curve models (Mavroeidis, Plagbourg-Moller and Stock, 2013)
  - Nonlinear regression models (D. Andrews and Cheng, 2013)
  - BLP (Armstrong, 2016)

Introduction

Plan for the talk

- I'll start by talking about three examples
  - Linear IV
  - BLP using product characteristics as instruments
  - DSGE model
- I'll then talk about
  - Why the usual approximations break down
  - How we can assess if conventional procedures are reliable in a given application
  - Robust inference procedures

Outline

Three Examples

Weak IV model

- Consider the linear model

$$Y_t = X_t\theta + \varepsilon_t$$

$$X_t = Z_t'\pi + v_t$$

  with Z_t a vector of instruments, $E\left[Z_t\varepsilon_t\right] = E\left[Z_t v_t\right] = 0$

- Can estimate θ by two-stage least squares or two-step GMM, test $H_0 : \theta = \theta_0$ with a t-test

Three Examples

Weak IV model

t-tests may over-reject when π too small

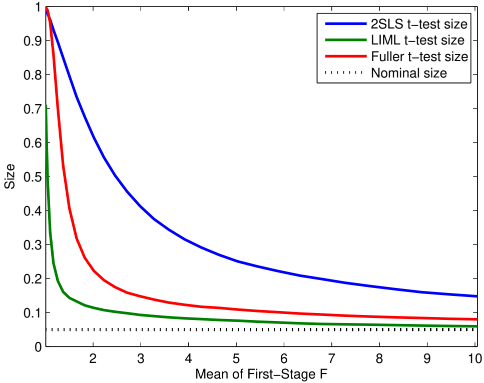

Figure: Size of nominal 5% t-tests for $H_0 : \beta = 5$ in homoskedastic linear IV model with 500 observations and 5 instruments. Correlation of (ε_t, v_t) is $-.98$.

Three Examples

BLP with product characteristic instruments

- Armstrong (2016)
- Suppose we observe aggregate data on a market of single product firms engaged in Bertrand competition
  - Then if we consider large-market asymptotics, for the logit or random-coefficient logit models, strategies using the characteristics of other products as instruments have asymptotically declining identifying power
  - Estimates based on these instruments will be inconsistent, even if we know the distribution of random coefficients
- For data from many markets, shows a similar result for logit model
  - Characteristic instruments asymptotically irrelevant if number of firms per market increases faster than number of markets

Three Examples

DSGE model

Consider the toy DSGE model

$$
\begin{gathered}
bE_t\pi_{t+1} + \kappa x_t - \pi_t = 0 \\
r_t - E_t\pi_{t+1} - \rho\Delta a_t = E_t x_{t+1} - x_t \\
\tfrac{1}{b}\pi_t + u_t = r_t
\end{gathered}
$$

where x_t is output, π_t is inflation, and r_t is an unobserved interest rate

$$
\begin{gathered}
\Delta a_t = \rho\Delta a_{t-1} + \varepsilon_{a,t};\ \ u_t = \delta u_{t-1} + \varepsilon_{u,t} \\
(\varepsilon_{a,t}, \varepsilon_{u,t})' \sim N\left(0, \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_u^2 \end{bmatrix}\right).
\end{gathered}
$$

Three Examples

DSGE model

- In joint work with Anna Mikusheva, show model is unidentified when $\rho = \delta$
- Classical statistical approaches perform poorly when $|\rho - \delta|$ small

$\rho - \delta$	0.05	0.1	0.2	0.3	0.5	0.7
Size of 5% Test	88.9%	79.8%	52.5%	28.1%	12.1%	9.8%

Table: Size of nominal 5% Wald tests based on 200 observations. From Andrews and Mikusheva (2014)

Outline

Why do Conventional Techniques Fail?

Source of usual asymptotic approximations

- Consider a GMM model with moment condition $g\left(X_t,\theta\right)$, identifying assumption

$$E\left[g\left(X_t,\theta_0\right)\right]=0$$

  - Define $g_T\left(\theta\right)=\frac{1}{T}\sum g\left(X_t,\theta\right)$

- Usual GMM approximations are exact in finite samples under three assumptions
  1. $g_T\left(\theta\right)=g_T\left(\theta_0\right)+\frac{\partial}{\partial\theta}g_T\left(\theta_0\right)\left(\theta-\theta_0\right)$
  2. $\frac{\partial}{\partial\theta}g_T\left(\theta_0\right)$ full rank and non-random
  3. $\sqrt{T}g_T\left(\theta_0\right)\sim N\left(0,\Sigma\left(\theta_0\right)\right)$ for $\Sigma\left(\theta_0\right)$ known

- Breakdown of usual approximations means one (or more) of these must be a poor approximation to finite-sample behavior

Why do Conventional Techniques Fail?

Source of usual asymptotic approximations

- Assumption (3) justified by the central limit theorem
  - Holds regardless of model identification status
- Assumptions (1) and (2) justified using identification assumptions
  - Derived by assuming $E\left[g\left(X_t,\theta\right)\right]=0$ only if $\theta=\theta_0$, departs rapidly from zero for $\theta\neq\theta_0$
  - Since assumed $E\left[g\left(X_t,\theta\right)\right]$ departs rapidly from zero, $E\left[\frac{\partial}{\partial\theta}g\left(X_t,\theta_0\right)\right]$ must be large
    - In particular larger than variance of $\frac{\partial}{\partial\theta}g_T\left(\theta_0\right)$, justifying (2)
  - Thus, can reject values θ outside a small neighborhood of θ_0
  - Over small neighborhood of θ_0, first order Taylor approximations are reasonable, justifying (1)

Why do Conventional Techniques Fail?

Example: Linear IV

- In linear IV model, (1) holds exactly, and (3) is a reasonable approximation
- When instruments are weak, leads to a breakdown of (2)
  - IV moment condition $g_T(\theta) = \frac{1}{T}\sum (Y_t - X_t\theta) Z_t$

$$\frac{\partial}{\partial\theta} g_T(\theta) = -\frac{1}{T}\sum X_t Z_t = -\frac{1}{T}\sum (Z_t Z_t')\pi + \frac{1}{T}\sum Z_t v_t$$

  - When π small, these terms are of the same order
    - $\frac{\partial}{\partial\theta} g_T(\theta)$ approximately normal with non-degenerate variance

Why do Conventional Techniques Fail?

Example: DSGE

- Suppose we estimate a DSGE model by matching moments
- Have some function $h(X_t)$, take

$$g(X_t, \theta) = h(X_t) - \tilde{h}(\theta)$$

- Here, assumption (2) holds exactly, and (3) is again a reasonable approximation
- In joint work with Anna Mikusheva, show that $\tilde{h}(\theta)$ can be highly nonlinear relative to the sample size
  - (1) is a very poor approximation

Outline

Detecting Weak Identification

How can we tell when the usual approximations fail?

- We can try to check whether (1)-(3) are reasonable approximations
  - What defines "reasonable"?
- The first stage F-statistic can be viewed as checking (2)
  - With a single endogenous regressor and k instruments, is of the form

$$F = \frac{1}{k} \cdot \frac{\partial}{\partial \theta} g_T(\theta)' \widehat{Var}\left(\frac{\partial}{\partial \theta} g_T(\theta)\right)^{-1} \frac{\partial}{\partial \theta} g_T(\theta)$$

  - Will be large when the mean of $\frac{\partial}{\partial \theta} g_T(\theta)$ is large relative to its variance
  - Stock and Yogo (2005) show that in IV with homoskedastic errors can use F to assess bias of estimators and size distortion in tests
    - Critical values for size distortion larger than "rule of thumb"

Detecting Weak Identification

The trouble with the first stage F-statistic

- Unfortunately, in overidentifed models Stock and Yogo (2005) results rely heavily on the assumption of iid homoskedastic data
- Olea and Pflueger (2013) show that conventional F statistic not a reliable guide to bias when used with heteroskedastic, serially correlated, or clustered data
- In Andrews (2018), I show in simulation that, even when using robust formulations of F statistic, can get large values of F together with distortions for conventional procedures
  - Have an example where mean of first stage F is 100,000, t-tests have 10% size distortion

Detecting Weak Identification

Alternative approaches

- In joint work with Anna Mikusheva, have given approach to checking (1) in moment-matching and minimum-distance models
  - $g(X_t, \theta) = h(X_t) - \tilde{h}(\theta)$
  - Based on measuring the curvature of $\tilde{h}$
  - Gives bounds on behavior for minimum-distance statistics
- Both this and the first stage F-statistic are based on measuring "inputs"
  - What about comparing "outputs"?

Detecting Weak Identification

Evaluating conventional approximations

- We can check the implications of the usual approximations
  - Contours of GMM objective function should be approximately elliptical
  - If using Bayesian methods, posterior should be approximately normal
  - Bootstrap distribution of estimator should be approximately normal
- However, evaluating identification in this way will introduce size and coverage distortions
  - If I decide whether to use an identification-robust confidence set based on the data, I'm essentially using a pretest
  - Pretest bias can be severe

Detecting Weak Identification

Pretest with bounded distortions

- In Andrews (2018), I show that one can detect weak identification by comparing robust and non-robust confidence sets
  - For all the commonly-used non-robust confidence sets, we can create asymptotically equivalent robust confidence sets
  - Compare (e.g.) 95% non-robust confidence set with 90% robust
    - If usual approximations OK, non-robust should contain robust
  - By assessing identification in this way, limit pretest bias
    - i.e. at most 5% coverage distortion above

Outline

Robust Inference

Robust Test Statistics

- Even under weak identification, can still construct valid and (potentially) informative tests
- Based on asymptotic approximations which hold regardless of identification strength
    - Simplest is S statistic of Stock and Wright (2000):

$$S(\theta) = g_T(\theta)' \widehat{Var}(g_T(\theta))^{-1} g_T(\theta)$$

    - Can show that $S(\theta_0) \to_d \chi^2_k$ under very weak assumptions
    - In just-identified models, performs well in wide range of contexts
- Tests based on S asymptotically inefficient over-identified models under conventional asymptotics
    - Kleibergen (2005) suggests alternative statistic K
    - Also depends on $\frac{\partial}{\partial\theta} g_T(\theta)$

Robust Inference

Robust Test Statistics

- Tests based on K are (locally) asymptotically efficient when identification is strong
  - Can have very poor power when identification is weak
- Many papers have studied different ways of combining information in these statistics
  - Moreira (2003), Kleibergen (2005), Andrews (2016)
  - D. Andrews Moreira and Stock (2006) show that test of Moreira (2003) is "nearly" uniformly most powerful in linear IV models with one endogenous regressor and homoskedastic errors
    - Analogous result fails for non-homoskedastic models
- However, Moreira and Ridder (2018) show that all tests based on (S, K) have poor power under some DGPs
  - Though not clear that commonly arise in practice

- Several recent papers have addressed the question of weighted average power maximizing tests
  - Olea (2018), Moreira and Moreira (2015), Moreira and Ridder (2018)
  - Definition of optimality requires specifying a weight function: papers propose defaults for IV models
- For general GMM models, much less is known

Robust Inference

Robust Confidence Sets

- Dufour (1997): If parameter space is unbounded and identification can be arbitrarily weak, any robust confidence set must be unbounded with positive probability
  - Conventional "estimate $\pm$ standard error" confidence sets have zero coverage
- To form a robust confidence set, invert robust tests
  - For $\phi(\theta)$ a test of $\theta_0 = \theta$

$$CS = \{\theta : \phi(\theta) \text{ does not reject}\}$$

  - Easy to compute if θ is low-dimensional
  - For high-dimensional θ, curse of dimensionality

- All of the results above are for inference on the full vector of GMM parameters
- Often we are interested in some lower-dimensional $\beta = f(\theta)$
  - e.g. $\beta = \theta_i$, to report confidence sets for a single parameter
- Unfortunately this is a hard problem
  - Simplest option: projection method, popularized by Dufour
  - Given a confidence set CS_θ, let

$$CS_\beta = \{\beta : \beta = f(\theta), \theta \in CS_\theta\}$$

  - This will typically be inefficient under strong identification

Robust Inference

Improving on Projection

- Three main types of approach proposed so far to improve on this simple projection approach
- Methods for models with additional structure
  - D. Andrews and Cheng (2012): known scalar parameter which controls identification
  - Andrews and Mikusheva (2016): Minimum-distance models
  - Guggenberger, Kleibergen, Mavroeidis, and Chen (2012): Homoskedastic linear IV
- Choosing confidence sets to reduce conservativeness of projection under strong identification
  - Chaudhuri and Zivot (2011): construct confidence set for θ such that can come arbitrarily close to efficiency if model is strongly identified
  - Extended by D. Andrews (2017)

Robust Inference

Improving on Projection: Numerically Intensive methods

- Numerically intensive methods
  - Elliott Mueller and Watson (2015), Moreira and Moreira (2013)
    - Derive weighted average power optimal tests in parametric models
    - Numerically daunting in cases with high-dimensional nuisance parameters unless have special structure, e.g. Moreira and Moreira (2013) can do linear IV with non-homoskedastic errors and multiple instruments
  - Bonferroni methods: McCloskey (2017)

Outline

Recent Projects

Unbiased Estimation in IV

- When instruments are weak, conventional estimators are biased
- Hirano and Porter (2015): in usual parameter space, there exists no mean, median, or quantile unbiased estimators in the linear IV model
- Result in joint work with Tim Armstrong: if sign of first stage is known, can construct an unbiased estimator for the coefficient θ on the endogenous regressor in a finite-sample normal model
    - Implies an asymptotically unbiased estimator under both weak and strong instrument asymptotics

Recent Projects

Unbiased Estimation in IV

- In just-identified model,

$$\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} (Z'Z)^{-1} Z'Y \\ (Z'Z)^{-1} Z'X \end{pmatrix} \sim N\left(\begin{pmatrix} \pi\theta \\ \pi \end{pmatrix}, \Sigma \right)$$

where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

- Unbiased estimator under $\pi > 0$ is

$$\hat{\theta}_U = \hat{\tau}\left(\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2 \right) + \frac{\sigma_{12}}{\sigma_2^2}$$

where $E[\hat{\tau}] = \frac{1}{\pi}$. By comparison

$$\hat{\theta}_{2SLS} = \frac{1}{\xi_2}\left(\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2 \right) + \frac{\sigma_{12}}{\sigma_2^2}$$

Recent Projects

Gaussian process approximation

- Joint work with Anna Mikusheva
- Under mild conditions,

$$\sqrt{T}\left(g_T(\cdot) - E\left[g_T(\cdot)\right]\right) \Rightarrow G(\cdot)$$

for G a mean-zero Gaussian process with consistently estimable covariance function

$$\Sigma\left(\theta, \tilde{\theta}\right) = Cov\left(G(\theta), G\left(\tilde{\theta}\right)\right)$$

Recent Projects

Gaussian process approximation

- Suggests the approximate model

$$g_T(\cdot) = E\left[g_T(\cdot)\right] + \frac{1}{\sqrt{T}} G(\cdot)$$

  where Σ is known
- Can think of $E\left[g_T(\cdot)\right]$ as an unknown parameter
- For testing $H_0 : \theta = \theta_0$, under null have $E\left[g_T(\theta)\right] = 0$
  - $E\left[g_T\left(\tilde{\theta}\right)\right]$ for $\tilde{\theta} \neq \theta$ is a functional nuisance parameter
  - Determines identification status of model

Recent Projects

Sufficient statistic

- One classical approach to dealing with nuisance parameters: condition on a sufficient statistic
  - Conditional on a sufficient statistic, the nuisance parameter has no effect on inference
- We show that in exact Gaussian model,

$$h_T(\cdot) = g_T(\cdot) - \Sigma(\cdot, \theta)\,\Sigma(\theta, \theta)^{-1} g_T(\theta)$$

  is a sufficient statistic for $E\left[g_T\left(\tilde{\theta}\right)\right]$, $\tilde{\theta} \neq \theta$ under the null
  - Conduct inference conditional on $h_T(\cdot)$
- We show that this gives asymptotically valid tests regardless of model identification

Conclusion

- Commonly-used approaches to inference can be unreliable when the data is uninformative
  - Arises in both linear and non-linear models
- Pretesting for identification can introduce distortions
  - Approaches based on F-statistic not in general valid in overidentified or nonlinear models
- To avoid drawing unreliable conclusions, can use robust inference procedures
  - Theory complete for homoskedastic linear IV
  - In progress for non-homoskedastic IV
  - Procedures available but much still unknown for nonlinear models and subset inference

The End

Thank you!

References

- Andrews, D. (2017), 'Identification-robust subvector inference', Unpublished Manuscript.
- Andrews, D. and Cheng, X. (2012), 'Estimation and inference with weak, semi-strong, and strong identification', *Econometrica* 80, 2153–2211.
- Andrews, D. and Cheng, X. (2013), 'Maximum likelihood estimation and uniform inference with sporadic identification failure', *Journal of Econometrics* 173, 36-56.
- Andrews, D., Moreira, M. and Stock, J. (2006), 'Optimal two-sided invariant similar tests of instrumental variables regression,' *Econometrica* 74, 715–752.
- Andrews, I. (2016), 'Conditional linear combination tests for weakly identified models,' *Econometrica* 84, 2155–2182.
- Andrews, I. (2018), 'Valid Two-Step Identification-Robust Confidence Sets for GMM'. *Review of Economics and Statistics*, 110(2), 337-348.

References

- Andrews, I. and Mikusheva, A. (2018), 'A geometric approach to weakly identified econometric models,' *Econometrica* 84, 1249-1264.
- Andrews, I. and Mikusheva, A. (2016), 'Conditional inference with a functional nuisance parameter', *Econometrica* 84, 1571-1612.
- Andrews, I. and Mikusheva A. (2014), 'Weak Identification in Maximum Likelihood: A Question of Information', *American Economic Review: Papers and Proceedings*, 104, 195-199
- Armstong, T. (2014), 'Large Market Asymptotics for Differentiated Product Demand Estimators with Economic Models of Supply', *Econometrica* 84, 1961-1980.
- Bound, J., Jaeger, D. & Baker, R. (1995), '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.

References

- Canova, F. and Sala, L. (2009), 'Back to square one: Identification issues in dsge models', *Journal of Monetary Economics* 56, 431–449.
- Chaudhuri, S. and E. Zivot (2011), 'A new method of projection-based inference in gmm with weakly identified nuisance parameters', Journal of Econometrics 164, 239–251.
- Dufour, J.M. (1997), 'Some impossibility theorems in Econometrics with Applications to Structural and Dynamic Models', *Econometrica* 65, 1365-1387.
- Elliott, G., Mueller, U. and Watson, M. (2015), 'Nearly optimal tests when a nuisance parameter is present under the null hypothesis', *Econometrica 83, 771-811.*
- Guggenberber P., F. Kleibergen, S. Mavroeidis, L. Chen (2012), 'On the Asymptotic Sizes of subset Anderson-Rubin and Lagrange Multiplier Tests in Linear Instrumental Variables Regression', *Econometrica* 80, 2649-2666.

References

- Hirano, K. & Porter, J. (2015), 'Location properties of point estimators in linear instrumental variables and related models', *Econometric Reviews*, 34 (6-10), 720-733.
- Kleibergen, F. (2005), 'Testing parameters in gmm without assuming they are identified', *Econometrica* 73, 1103–1123.
- Mavroeidis, S. (2010). 'Monetary Policy Rules and Macroeconomic Stability: Some New Evidence' *American Economic Review* 100, 491-503.
- Mavroeidis, S., M. Plagborg-Moller, and J. Stock (2014). 'Empirical evidence on inflation expectations in the new keynesian phillips curve'. *Journal of Economic Literature*, 52: 124-88.
- McCloskey, A. (2017). 'Bonferroni-Based Size-Correction for Nonstandard Testing Problems'. *Journal of Econometrics*, 200, 17-35.

References

- Moreira, M. (2003), 'A conditional likelihood ratio test for structural models', *Econometrica* 71, 1027–1048.
- Moreira, H. and M. Moreira, (2013), 'Contributions to the theory of optimal tests'. Unpublished Manuscript.
- Moreira, H. and M. Moreira, (2015), 'Optimal Two-Sided Tests for Instrumental Variables Regression with Heteroskedastic and Autocorrelated Errors'. Unpublished Manuscript.
- Moreira, M. and G. Ridder, (2018), 'Optimal Invariant Tests in an Instrumental Variables Regression with Heteroskedastic and Autocorrelated Errors'. Unpublished Manuscript.
- Olea, J. (2018), 'Efficient conditionally similar tests: Finite-sample theory and larges sample applications'. Unpublished Manuscript.
- Olea, J.L.M. and C. Pfleuger (2013), 'A robust test for weak instruments', *Journal of Business and Economic Statistics* 31, 358–369.

References

- Stock, J. and Wright, J. (2000), 'Gmm with weak identification', *Econometrica* 68, 1055–1096.
- Stock, J. and M. Yogo (2005), Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, Cambridge University Press, chapter Testing for Weak Instruments in Linear IV Regression, 80–108.
- Yogo, M. (2004), 'Estimating the elasticity of intertemporal substitution when instruments are weak', *Review of Economics and Statistics* 86, 797–810.