Detecting Identification Failure in Moment Condition Models

Jean-Jacques Forneron, Boston University

October 25, 2019

Introduction: Identification in GMM

- Interested in moment condition models: GMM, MD, SMM,...

$$g_n(\theta_0) \stackrel{def}{=} \mathbb{E}\left(\bar{g}_n(\theta_0)\right) = 0, \quad \theta_0 \in \Theta \subset \mathbb{R}^{d_\theta}$$

Introduction: Identification in GMM

- Interested in moment condition models: GMM, MD, SMM,...

$$g_n(\theta_0) \stackrel{def}{=} \mathbb{E}\left(\bar{g}_n(\theta_0)\right) = 0, \quad \theta_0 \in \Theta \subset \mathbb{R}^{d_\theta}$$

- Usual regularity conditions imply:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, V)$$

where $\hat{\theta}_n = \text{argmin}_{\theta \in \Theta} \|\bar{g}_n(\theta)\|_W$, $W = W_n(\theta)$ weight matrix

Introduction: Identification in GMM

- Interested in moment condition models: GMM, MD, SMM,...

$$g_n(\theta_0) \overset{def}{=} \mathbb{E}\left(\bar{g}_n(\theta_0)\right) = 0, \quad \theta_0 \in \Theta \subset \mathbb{R}^{d_\theta}$$

- Usual regularity conditions imply:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, V)$$

where $\hat{\theta}_n = \text{argmin}_{\theta \in \Theta} \|\bar{g}_n(\theta)\|_W$, $W = W_n(\theta)$ weight matrix

- Important assumption: global identification

$$\forall \varepsilon > 0, \exists \delta(\varepsilon) > 0 : \inf_{\|\theta - \theta_0\| > \varepsilon} \|\mathbb{E}(\bar{g}_n(\theta))\|_W \geq \delta(\varepsilon)$$

Introduction: Identification in GMM

- Interested in moment condition models: GMM, MD, SMM,...

$$g_n(\theta_0) \stackrel{def}{=} \mathbb{E}\left(\bar{g}_n(\theta_0)\right) = 0, \quad \theta_0 \in \Theta \subset \mathbb{R}^{d_\theta}$$

- Usual regularity conditions imply:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, V)$$

where $\hat{\theta}_n = \text{argmin}_{\theta \in \Theta} \|\bar{g}_n(\theta)\|_W$, $W = W_n(\theta)$ weight matrix

- Important assumption: global identification

$$\forall \varepsilon > 0, \exists \delta(\varepsilon) > 0 : \inf_{\|\theta - \theta_0\| > \varepsilon} \|\mathbb{E}(\bar{g}_n(\theta))\|_W \geq \delta(\varepsilon)$$

- However, typically non-Gaussian under:
  - i. **set/weak identification**: $\nexists \delta(\varepsilon) > 0$ or $\delta(\varepsilon) \asymp n^{-1/2}$
  - ii. **local identification failure**: $\partial_\theta g_n(\theta_0)$ (close to) singular

Dealing with Identification Failure

- Requires identification robust inference:

  e.g. Anderson and Rubin (1949); Stock and Wright (2000); Moreira (2003); Kleibergen (2005); Andrews and Cheng (2012); Andrews and Mikusheva (2016); Chen et al. (2018),...

- More computationally demanding than standard inference

Dealing with Identification Failure

- Requires identification robust inference:

  e.g. Anderson and Rubin (1949); Stock and Wright (2000); Moreira (2003); Kleibergen (2005); Andrews and Cheng (2012); Andrews and Mikusheva (2016); Chen et al. (2018),...

- More computationally demanding than standard inference
- Typically for the full vector θ; subvector inference:
  - i. *Projection* (Dufour and Taamouti, 2005)
  - ii. *Bonferroni* (McCloskey, 2017)
  - $\Rightarrow$ Conservative
  - $\nearrow$ power: concentrate out identified nuisance parameters

Dealing with Identification Failure

- Requires identification robust inference:

  e.g. Anderson and Rubin (1949); Stock and Wright (2000); Moreira (2003); Kleibergen (2005); Andrews and Cheng (2012); Andrews and Mikusheva (2016); Chen et al. (2018),...
- More computationally demanding than standard inference
- Typically for the full vector θ; subvector inference:
  i. *Projection* (Dufour and Taamouti, 2005)
  ii. *Bonferroni* (McCloskey, 2017)

  $\Rightarrow$ Conservative

  $\nearrow$ power: concentrate out identified nuisance parameters
- **This paper:** is about answering two questions:
  i. is θ strongly globally identified?
  ii. which components of θ are not identified?

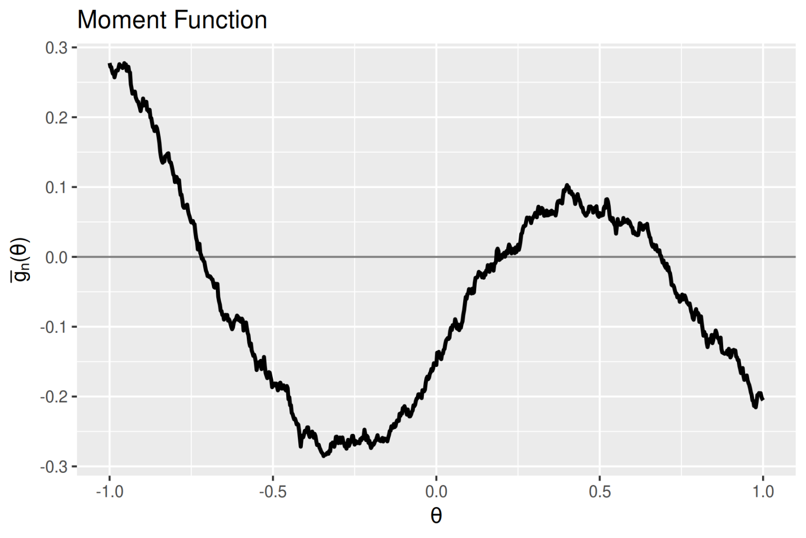
Contributions of the Paper

- Two main contributions:
  i. A generic approach to detecting both weak/set identification and local identification failure
  ii. A two-step procedure for robust subvector inference

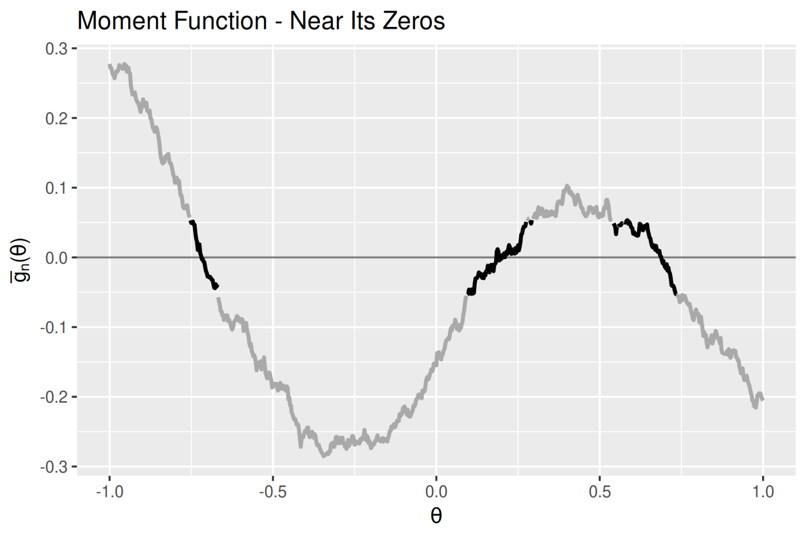
Contributions of the Paper

- Two main contributions:
  i. A generic approach to detecting both weak/set identification and local identification failure
  ii. A two-step procedure for robust subvector inference
- Introduce a quasi-Jacobian matrix:
  - Jacobian $\simeq$ local linear approx. in a $\|\cdot\|$ neighborhood of θ_0
    - informative about *local identification only*
  - quasi-Jacobian local linear approx. in a $\|g_n(\cdot)\|_W$ neighborhood
    - informative about *local + global identification*

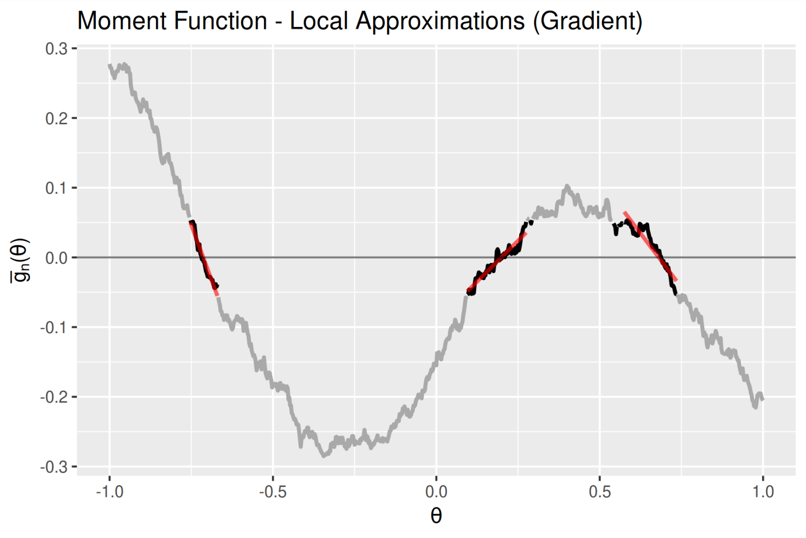
The quasi-Jacobian: Intuition



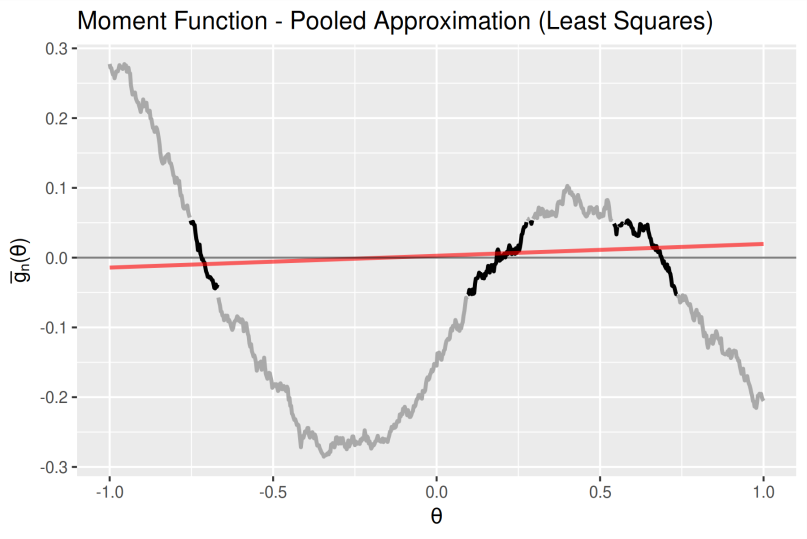
The quasi-Jacobian: Intuition



The quasi-Jacobian: Intuition



The quasi-Jacobian: Intuition



How to Build the Approximation in Practice

- Two estimators

  i. **Sup-norm (∞):** $(A_{n,\infty}, B_{n,\infty})$

$$\operatorname{argmin}_{A,B}\left\{\sup_{\theta\in\Theta}\left[\|\bar{g}_n(\theta)-A-B\theta\|\times K\left(\left\|\frac{\bar{g}_n(\theta)}{\kappa_n}\right\|_W\right)\right]\right\}$$

  ii. **Least-squares (LS):** $(A_{n,LS}, B_{n,LS})$

$$\operatorname{argmin}_{A,B}\int_\Theta\left[\|\bar{g}_n(\theta)-A-B\theta\|^2\times K\left(\left\|\frac{\bar{g}_n(\theta)}{\kappa_n}\right\|_W-\left\|\frac{\bar{g}_n(\hat{\theta}_n)}{\kappa_n}\right\|_W\right)\right]d\theta$$

- where

  i. K is a kernel, either:

     - Lipschitz-continuous, strictly positive on the support $[-1,1]$
     - Exponential $K(x)=C_1\exp(-C_2|x|^a)$, $C_1, C_2, a>0$ (LS only)

  ii. κ_n is a bandwidth

     - $\sqrt{n}\kappa_n\to\infty, \kappa_n^2=o(n^{-1/2})$ (e.g. $\kappa_n=\sqrt{2\log\log(n)/n}$)
     - $\tilde{\kappa}_n=\kappa_n\log(n)^{1/a}$ $\sqrt{n}\tilde{\kappa}_n\to\infty, \tilde{\kappa}_n^2=o(n^{-1/2})$ (exp. kernel)

Remarks

- **Sup-Norm** $(A_{n,\infty}, B_{n,\infty})$**:**
    i. strong theoretical predictions about $B_{n,\infty}$
    ii. convex optimization problem but more challenging to compute
- **Least-squares** $(A_{n,LS}, B_{n,LS})$**:**
    i. very easy to compute:
    $(A_{n,LS}, B'_{n,LS}) = \left(\int_\Theta X(\theta)X(\theta)'\hat{K}_n(\theta)d\theta\right)^{-1}\int_\Theta X(\theta)\bar{g}_n(\theta)'\hat{K}_n(\theta)d\theta$
    where $\hat{K}_n(\theta) = K\left(\|\bar{g}_n(\theta)\|_W/\kappa_n - \|\bar{g}_n(\hat{\theta}_n)\|_W/\kappa_n\right), X(\theta) = (1, \theta')$
    ii. theoretical predictions depend on the topology of Θ_0

What happens under weak or set identification?

$\Rightarrow$ $\text{plim}_{n\to\infty} B_n$ singular and

i. **eigenvalues** (rank) of B_n informative about identifiability of θ

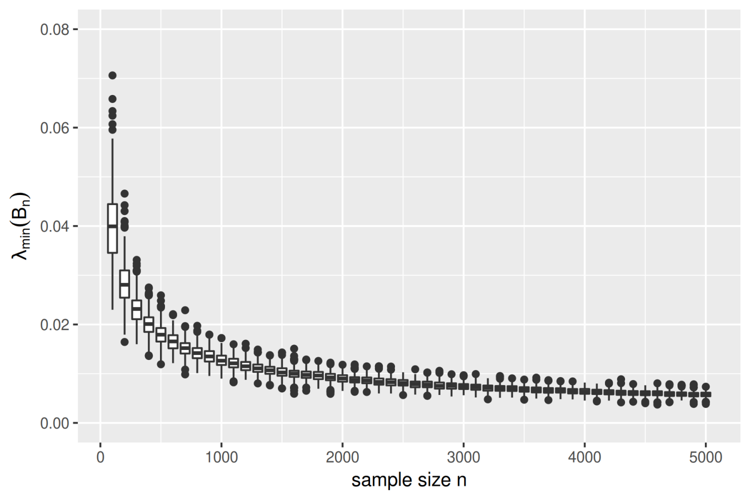
ii. **eigenvectors** informative about span of identification failure

- Remark: $g_n(\theta)$ linear $\Rightarrow$ approximation is exact - IV:

$$A_n = Z'y/n, B_n = -Z'X/n$$

- Stock and Yogo (2005) use Cragg-Donald rank test

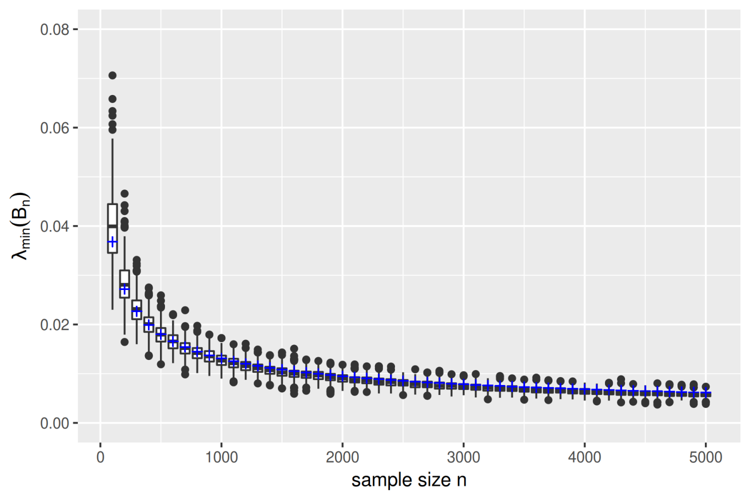
Illustration: Smallest Eigenvalue and Sample Size



Boxplot: $\lambda_{\min}(B_n)$.

Setting: $y_i = \theta_1 x_{1,i} + \theta_1\theta_2 x_{2,i} + e_i$, $\theta_1 = 2/\sqrt{n}$, 500 replications

Illustration: Smallest Eigenvalue and Sample Size



Boxplot: $\lambda_{\min}(B_n)$. Blue Cross: fitted predicted rate

Setting: $y_i = \theta_1 x_{1,i} + \theta_1\theta_2 x_{2,i} + e_i$, $\theta_1 = 2/\sqrt{n}$, 500 replications

Related Literatures (non-exhaustive)

- For an identification safari: Lewbel (2018)
- **Local/global identification in the population**:
  - Koopmans and Reiersol (1950); Fisher (1967); Rothenberg (1971); Brown (1983); Komunjer (2012), ...
  - hard to check for many models (e.g. SMM/Indirect Inference)
- **Detecting identification failure in finite samples**:
  - Stock and Yogo (2005); Olea and Pflueger (2013); Wright (2003); Inoue and Rossi (2011); Arellano et al. (2012); Bravo et al. (2012); Antoine and Renault (2017), ...
- **Distribution under weak identification**:
  - Nelson and Startz (1990); Choi and Phillips (1992); Dufour (1997); Staiger and Stock (1997); Stock and Wright (2000)...
- **Identification robust inference**: (prev. slide)

Definitions and Main Assumptions

Identification Regimes

Definition (Identification Regimes - $g_n(\theta) = \mathbb{E}(\bar{g}_n(\theta))$)

1. **strong identification**
   i. $\forall \varepsilon > 0, \exists \delta(\varepsilon) > 0 : \inf_{\|\theta-\theta_0\|>\varepsilon} \|g_n(\theta)\|_W \geq \delta(\varepsilon)$
   ii. $\exists \varepsilon > 0$ and $\underline{c} > 0$, $\|\theta - \theta_0\| \leq \varepsilon \Rightarrow \|g_n(\theta)\|_W \geq \underline{c}\|\theta - \theta_0\|$
2. **semi-strong identification**
   i. 1.i. holds;
   ii. $n \times \lambda_{\min}(\partial_\theta g_n(\theta_0)'\partial_\theta g_n(\theta_0)) \to \infty$; cond. on hod $+ \exists \varepsilon > 0$, $\|\theta - \theta_0\| \leq \varepsilon \Rightarrow \|g_n(\theta)\|_W \asymp \|\partial_\theta g_n(\theta_0)[\theta - \theta_0]\|$
3. **higher-order local identification**
   i. 1.i. holds
   ii. $\exists \varepsilon > 0$, $P_1, \dots, P_r$ projection matrices with $P_r \neq 0$, $\|\theta - \theta_0\| \leq \varepsilon \Rightarrow \|g_n(\theta)\|_W \asymp \sum_{j=1}^r \|P_j(\theta - \theta_0)\|^j$
4. **weak/set identification**:
   i. $\exists \theta_0 \neq \theta_1 \in \Theta_0 = \{\theta, \lim_{n\to\infty} \sqrt{n}\|g_n(\theta)\|_W < \infty\}$

Identification Regimes and Asymptotic Properties of $\hat{\theta}_n$

Identification Regime	$\hat{\theta}_n$ consistent?	Rate of convergence	Limiting distribution
Strong	Yes	$\sqrt{n}$	Gaussian
Semi-Strong	Yes	slower than $\sqrt{n}$	Gaussian
Higher-Order	Yes	$n^{1/4}$ or slower	non-Gaussian
Weak or Set	No	-	non-Gaussian

- Goal: characterize the behaviour of $B_{n,\infty}$, $B_{n,LS}$ in each regime
- assume $\exists \theta_0$ st. $\|g_n(\theta_0)\|_W = 0$
- for (semi)-strong, assume $\theta_0 \in int(\Theta)$ - boundary not covered

References

Simple Examples

1. (Cheng, 2015) **Nonlinear Least-Squares:**

$$y_t = \theta_{1,n} x_{1,t} + \theta_{1,n}\theta_2 x_{2,t} + e_t$$

   - semi/strong $\sqrt{n} \times |\theta_{1,n}| \to +\infty$
   - weak/set $\sqrt{n} \times |\theta_{1,n}| = O(1)$

Some Details

2. (Gospodinov and Ng, 2015) **Possibly non-invertible MA(1):**

$$y_t = \sigma[e_t - \vartheta e_{t-1}], \quad e_t \sim (0, 1, \tau_n) \quad \tau_n = \text{skewness}$$

   - strong $\sqrt{n} \times |\tau_n| \to \infty$; $\vartheta \in \mathbb{R}/\{-1, 0, 1\}$
   - weak/set $\sqrt{n} \times |\tau_n| = O(1)$; $\vartheta \in \mathbb{R}/\{-1, 0, 1\}$
   - second-order $\vartheta \in \{-1, 1\}; \tau \in \mathbb{R}$

Assumptions on $\bar{g}_n$ (Common)

Assumption (Common)

1. **Uniform CLT**
2. **Stochastic Equicontinuity**

+ conditions on weighting matrix $W_n(\theta)$ (invertible, ...)

allows for:

1. non-smooth, discontinuous moments
   e.g. quantile IV, SMM with indicator function, etc.
2. profile moments $\bar{g}_n\big[\theta, \hat{\gamma}(\theta)\big]$ - $\hat{\gamma}(\theta)$ (semi)-strongly identified

Preview: Asymptotic Properties of $\hat{\theta}_n$ and $B_{n,LS/\infty}$

Identification Regime	Asymptotics for $\hat{\theta}_n$	Asymptotics for $B_{n,LS/\infty}$
Strong	Gaussian	$B_{n,LS/\infty} \simeq$ Jacobian
Semi-Strong	Gaussian	$B_{n,LS/\infty} \simeq$ Jacobian
Higher-Order	Non-Gaussian	$B_{n,LS/\infty} v_j \asymp \text{bandwidth}^{1-1/j}$
Weak or Set	Non-Gaussian	$B_{n,LS/\infty} v \asymp$ bandwidth

- for all directions:
    - v_j in which $\|g_n(\theta_0 + v_j)\| \asymp \|v_j\|^j$
    - $v = \theta_0 - \theta_1$ with $\theta_0, \theta_1 \in \Theta_0$, weakly identified set

Asymptotic Behaviour of $B_{n,LS/\infty}$

$A_{n,LS/\infty}, B_{n,LS/\infty}$ - (Semi)-Strong Identification

Theorem (Semi-Strong Identification)

Suppose the model is (semi)-strongly identified, compact kernel conditions + assumptions above hold and

$$\kappa_n^2 = o\Big[\lambda_{\min}\Big(\partial_\theta g_n(\theta_0)'\partial_\theta g_n(\theta_0)\Big)\Big]$$

then:

i. $A_{n,LS/\infty} = \bar{g}_n(\theta_0) - B_{n,LS/\infty}\theta_0 + o_p(n^{-1/2})$

ii. $B_{n,LS/\infty}H_n = \partial_\theta g_n(\theta_0)H_n + o_p(n^{-1/2}\kappa_n^{-1})$

where $H_n = [\partial_\theta g_n(\theta_0)'\partial_\theta g_n(\theta_0)]^{-1/2}$

Implication: the estimator based on $A_{n,LS/\infty}, B_{n,LS/\infty}$

$$H_n^{-1}[\hat{\theta}_{n,LS/\infty} - \hat{\theta}_{n,GMM}] = o_p(n^{-1/2})$$

$B_{n,LS/\infty}$ is a smoothed estimate of the Jacobian $\partial_\theta g_n(\theta_0)$

Theorem (Weak and Set Identification)

Suppose the model is weakly and/or set identified, compact kernel conditions + assumptions above hold, then:

i. $|\lambda_{\min}(B_{n,\infty})| = O_p(\kappa_n)$

ii. $\forall v \in V = Span(\{\theta_0 - \theta_1, \theta_0, \theta_1 \in \Theta_0\})$: $B_{n,\infty} v = O_p(\kappa_n)$

V is the span of the identification failure

Theorem 2: sketch of the proof

- By construction, $\|\bar{g}_n(\theta)\|\hat{K}_n(\theta) \leq \|K\|_\infty \kappa_n / \lambda_{\min}(W)$

$\Rightarrow$ By minimization, we have

$$\sup_{\theta\in\Theta}\left[\|A_{n,\infty} + B_{n,\infty}\theta - \bar{g}_n(\theta)\|_W \hat{K}_n(\theta)\right] \leq \sup_{\theta\in\Theta}\left[\|\bar{g}_n(\theta)\|_W \hat{K}_n(\theta)\right] \leq O(\kappa_n)$$

$+$ reverse triangle inequality:

$$O(\kappa_n) \geq \|A_{n,\infty} + B_{n,\infty}\theta - \bar{g}_n(\theta)\|\hat{K}_n(\theta) \geq \|A_{n,\infty} + B_{n,\infty}\theta\|\hat{K}_n(\theta) - O(\kappa_n)$$
$$\Rightarrow O(\kappa_n) \geq \|A_{n,\infty} + B_{n,\infty}\theta\|\hat{K}_n(\theta) \geq 0$$

- Also, wp $\nearrow 1$, both $\hat{K}_n(\theta_0)$ and $\hat{K}_n(\theta_1) \geq \underline{K} > 0$
- Apply inequality above for θ_0, then for θ_1 and we get:

$$0 \leq \|B_{n,\infty}(\theta_1 - \theta_0)\| \leq O_p(\kappa_n)$$

- For any pair $(\theta_0, \theta_1) \in \Theta_0$, the results follow

$A_{n,\infty}, B_{n,\infty}$ - Higher-Order Local Identification

Theorem (Higher-Order Local Identification)

Suppose the model is locally higher-order identified at order $r \geq 2$, compact kernel conditions + assumptions above hold, then:

i. $|\lambda_{\min}(B_{n,\infty})| = O_p(\kappa_n^{1-1/r})$

ii. $\forall v_j \in Span(P_j)$: $B_{n,\infty} v_j = O_p(\kappa_n^{1-1/j})$

Recall: P_j is the direction in which $\|g_n\|_W$ goes to 0 no faster than a polynomial of order j

Least-Squares Approximation: Notation

- Let $\hat{\pi}_n$ be the density implied by $\hat{K}_n$:

$$\hat{\pi}_n(\theta) = \frac{K(\|\bar{g}_n(\theta)\|_W/\kappa_n - \|\bar{g}_n(\hat{\theta}_n)\|_W/\kappa_n)}{\int_\Theta K(\|\bar{g}_n(\theta)\|_W/\kappa_n - \|\bar{g}_n(\hat{\theta}_n)\|_W/\kappa_n)d\theta}$$

- K = Gaussian density, $\kappa_n = n^{-1/2}$: quasi-Bayesian estimation (Chernozhukov and Hong, 2003; Creel et al., 2015)
- Quasi-posterior mean, variance:

$$\bar{\theta}_n = \int_\Theta \theta \hat{\pi}_n(\theta) d\theta, \quad \Sigma_n = \int_\Theta (\theta - \bar{\theta}_n)(\theta - \bar{\theta}_n)' \hat{\pi}_n(\theta) d\theta$$

- Moon and Schorfheide (2012); Chen et al. (2018): posterior concentrates on the identified set

$B_{n,LS}$ and the Quasi-Posterior Variance

Lemma (Relation between $B_{n,LS}$ and Σ_n)

Under any identification regime, compact/exponential kernel conditions + assumptions above hold + technical cond. for exponential kernel, then:

i. $\mathit{trace}\left(B_{n,LS}\Sigma_n B'_{n,LS}\right) = O_p(\tilde{\kappa}_n^2)$

ii. $\lambda_{\min}\left(B'_{n,LS}B_{n,LS}\right)\lambda_{\max}(\Sigma_n) = O_p(\tilde{\kappa}_n^2)$

where $\tilde{\kappa}_n = \kappa_n$ for compact kernel and $\kappa_n \log(n)^{1/a}$ for exp. kernel

Lemma 1: sketch of the proof

- Least-squares formula: $A_{n,LS} = \int \bar{g}_n(\theta)\hat{\pi}_n(\theta)d\theta - B_{n,LS}\bar{\theta}_n$
- Objective becomes

$$\int \|B_{n,LS}(\theta - \bar{\theta}_n) - [\bar{g}_n(\theta) - \int \bar{g}_n(\tilde{\theta})\hat{\pi}_n(\tilde{\theta})d\tilde{\theta}]\|^2 \hat{\pi}_n(\theta)d\theta \leq O_p(\tilde{\kappa}_n^2)$$

- Similar strategy as before implies:

$$\int \|B_{n,LS}(\theta - \bar{\theta}_n)\|^2 \hat{\pi}_n(\theta)d\theta \leq O_p(\tilde{\kappa}_n^2)$$

- By definition of the Frobenius norm, it implies:

$$\underbrace{\int \text{trace}\left(B_{n,LS}(\theta - \bar{\theta}_n)(\theta - \bar{\theta}_n)' B'_{n,LS}\right) \hat{\pi}_n(\theta)d\theta}_{=\text{trace}(B_{n,LS}\Sigma_n B'_{n,LS})} \leq O_p(\tilde{\kappa}_n^2)$$

- Which implies the results

Proposition (Weak and Set Identification)

Suppose the model is weakly and/or set identified, compact/exponential kernel conditions + assumptions above hold and $\exists \theta_0 \neq \theta_1 \in \Theta_0$ *with:*

a. $0 < \varepsilon \leq \|\theta_0 - \theta_1\|$
b. $\exists \eta > 0$, *for* $j \in \{0, 1\}$*:*

$$\hat{\pi}_n \left(\mathcal{B}_{\varepsilon/3}\left(\theta_j\right)\right) \geq \eta + o_p(1)$$

then:

i. $\lambda_{\max}(\Sigma_n) \geq \eta\varepsilon^2/[36 d_\theta] + o_p(1)$
ii. $\lambda_{\min}(B_{n,LS}) \leq O_p(\tilde{\kappa}_n)$

$A_{n,LS}, B_{n,LS}$ - Anatomy of an Identification Failure

Theorem (Topology of the Weakly Identified Set)

Suppose one the following holds

a. $int(\Theta_0) \neq \emptyset$ *(omni-directional failure)*
b. $\Theta_0 = \cup_{j=1}^{k}\{\theta_j\}$ + *same local behaviour*
c. $\Theta_0 = \cup_{j=1}^{k} S_j$, k_j *dimensional manifolds* + *local behaviour*

then the previous Theorem holds

Remark: sets S_j with largest k_j dominate the (quasi)-posterior

Corollary (Global Re-Parameterization)

Suppose that $\theta = \varphi(\alpha, \gamma)$; $int(\mathcal{A}_0) \neq \emptyset$ *($\cup$ manifolds)*, $\Gamma_0 = \{\gamma_0\}$ + *conds. on* φ, *local behaviour,... then Proposition c. above holds.*

Counter-Example

Detecting Identification Failure and Two-Step Subvector Inference

Subvector Inference: General Idea

- Focus on weak/set vs. (semi)-strong identification
- Linear hypothesis:

$$H_0 : \; R\theta_0 = c \text{ vs. } R\theta_0 \neq c$$

- Main idea $\theta = (\theta_1, \theta_2)$
  - θ_1 weak/set/higher-order identified: needs to be fixed
  - θ_2 (semi)-strongly identified, estimable for θ_1 fixed
- Simple case: $\text{span}(R) = \text{span}(P_{\theta_1})$, i.e.

$$\Theta_0 \cap \{\theta \in \Theta, \; R\theta = c\} = \{\theta_{0,c}\} \text{ singleton}$$

  or empty depending on c
- If not, can add restrictions $\tilde{R}$ until θ is point identified
- We'll use this to do **two-step inference**

Two-Step Subvector Inference: Second Step

- Suppose θ weak/set identified on Θ, (semi)-strongly identified on $\Theta \cap \{\theta \in \Theta, R\theta = c, \tilde{R}\theta = \tilde{c}\}$ for each $\tilde{c}$
- Projection Inference:
  i. construct $\tilde{CS}_{1-\alpha}$ for $(R', \tilde{R}')'\theta$ assuming the remaining coefficients are (semi)-strongly identified
  ii. the confidence set for $R\theta$ collects all values of $R\theta$ in $\tilde{CS}_{1-\alpha}$
- Remarks:
  - a lower rank for $\tilde{R} \Rightarrow$ less conservative, more power
  - full projection when $\text{rank}(R', \tilde{R}') = d_\theta$

First Step: Collapsing the Identified Set into a Singleton

- Consider a deterministic sequence of constraint matrices

$$R_1 = R, R_2 = (R_1', \tilde{R}_2')', \dots, R_{\mathcal{L}} = (R_{\mathcal{L}-1}', \tilde{R}_{\mathcal{L}}')'$$
$$1 \leq \text{rank}(R_1) < \cdots < \text{rank}(R_{\mathcal{L}}) = d_\theta$$

- By construction $\exists \ell^\star \leq \mathcal{L}$ (smallest) such that $\forall \ell \geq \ell^\star$:

$$\Theta_0 \cap \{\theta \in \Theta,\ R_\ell \theta = c_\ell\}$$

  is either a singleton or the empty set depending only on c_ℓ
- Assume remaining parameters are (semi)-strongly identified
  - we could re-compute $B_{n,LS}$ with the restrictions to check
- Want an algorithm that finds $\hat{\ell}_n \geq \ell^\star$ wp $\nearrow 1$

Which parameters to fix?

Lemma (Collapsing the Weakly Identified Set)

Let $\underline{\lambda}_n > 0$ st. $\kappa_n = o(\underline{\lambda}_n)$; suppose B_n is a $O_p(\kappa_n)$ on V. If we use $\underline{\lambda}_n$ as a cutoff to pick $\hat{\ell}_n$ st:

i. $\hat{d}_V = \#\{j \leq d_\theta, \lambda_j(B_n) \leq \underline{\lambda}_n\}$

ii. $rank(R_{\hat{\ell}_n}) \geq \hat{d}_V$, $\#\{j \leq d_\theta - rank(R_{\hat{\ell}_n}), \lambda_j(B_n P^{\perp}_{R_{\hat{\ell}_n}}) \leq \underline{\lambda}_n\} = 0$

then, wp $\nearrow 1$, $rank(P_{R_{\hat{\ell}_n}} P_V) = rank(P_V)$, i.e. $\hat{\ell}_n \geq \ell^\star$

Remarks:

- rule-of-thumb for $\underline{\lambda}_n$ in a few slides
- prev. results $\Rightarrow \#\{j \leq d_\theta - \text{rank}(R_{\hat{\ell}_n}), \lambda_j(B_n P^{\perp}_{R_{\hat{\ell}_n}}) \leq \underline{\lambda}_n\} \geq 1$ wp $\nearrow 1$ for each $1 \leq \ell < \ell^\star$
- use a family-wise error rate argument for the group $1 \leq \ell < \ell^\star$

Implications for Subvector Inference

Theorem (Two-Step Subvector Inference)

Under weak or set identification: suppose test statistics $S_{\ell,n}$ for $R_\ell\theta = c_\ell$ satisfies

$$\inf_{\ell^\star \leq \ell \leq \mathcal{L}} \mathbb{P}(S_{\ell,n} \leq c_{1-\alpha,\ell}) \geq 1 - \alpha + o(1)$$

then $\hat{\ell}_n \geq \ell^\star$ wp $\nearrow$ 1 implies

$$\mathbb{P}(S_{\hat{\ell}_n,n} \leq c_{1-\alpha,\hat{\ell}_n}) \geq 1 - \alpha + o(1)$$

Under (semi)-strong identification: conditions on eigenvalues & $\underline{\lambda}_n$ imply $\hat{\ell}_n = 1$ wp $\nearrow 1$ and

$$\mathbb{P}(S_{\hat{\ell}_n,n} \leq c_{1-\alpha,\hat{\ell}_n}) = \mathbb{P}(S_{1,n} \leq c_{1-\alpha,1}) + o(1)$$

Remarks

Designing a Cutoff $\underline{\lambda}_n$ - 1/2

- Use a simple asymptotic framework
  - (semi)-strong local asymptotics
  - look at simple linear t-test over all directions
  - bound worst-case size distortion in terms of $\lambda_{\min}(B_n)$
  - function of n, the signal - i.e. $\lambda_{\min}(B_n)$ - and the noise
- A given level of size distortion requires:
  - $\lambda_{\min}(B_n) \leq$ quantities(n,co-variances)
  - use this as a cutoff $\underline{\lambda}_n$ to detect identification failure

Designing a Cutoff $\underline{\lambda}_n$ - 2/2

- Non-local asymptotics (MA model)
  - partition the parameter space into clusters
  - within each cluster use rule of thumb above
  - distance between clusters also implies size distortion
- Higher-order asymptotics
  - check residual curvature
  - non-linearities $\Rightarrow$ size distortion

Details

Monte-Carlo Illustrations

Example 1: NLS

- Simple example:

$$y_t = \theta_1 x_{1,t} + \theta_1 \theta_2 x_{2,t} + e_t$$

- Identification failure $\theta_{1,0} = 0$, weak identification $\theta_{1,0} \simeq 0$
- Two cutoffs $\underline{\lambda_n}$: $\sqrt{\log(n)/n}$, rule-of-thumb
- Null hypothesis:

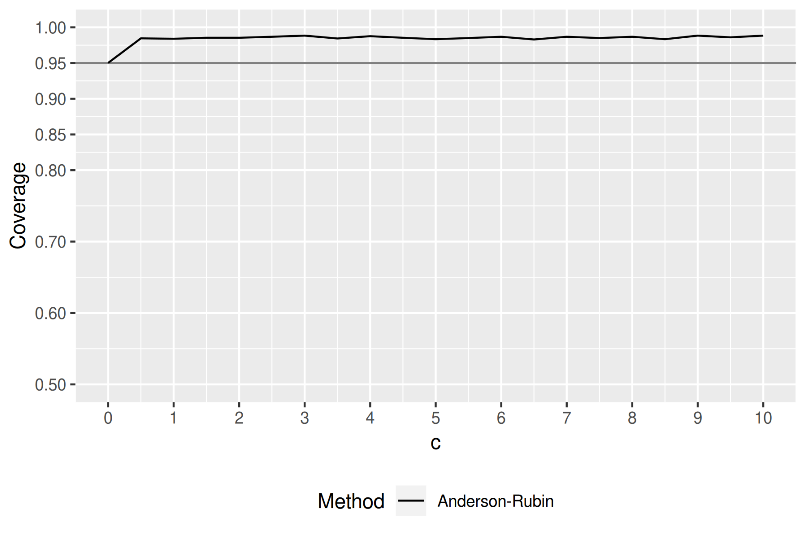
$$H_0 : \theta_1 = \theta_{1,0} = c/\sqrt{n}$$

- Pretend like we don't know the identification structure
- $\lambda_{\min}(B_{n,LS/\infty}) \leq \lambda_n$ suggests weak identification

Example 1: NLS

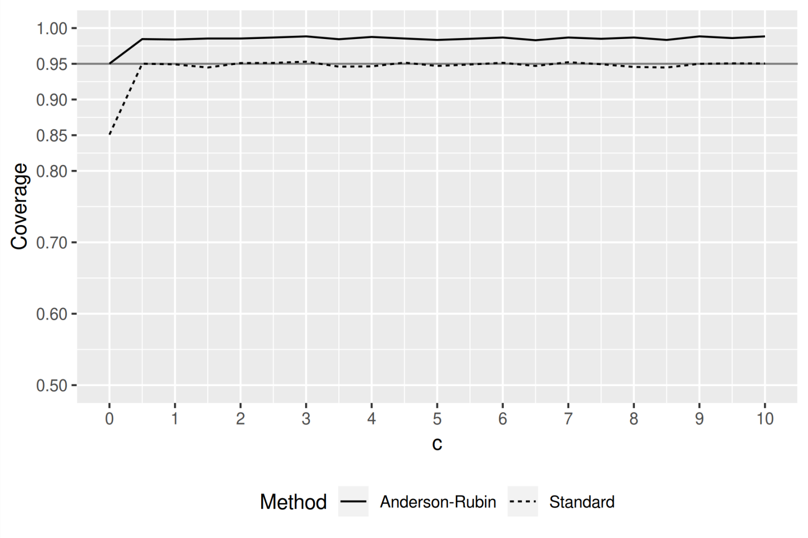
- Use projected S-statistic (two-step/Anderson-Rubin) and Wald/QLR test (standard)
- Main difference between two-step & Anderson-Rubin is critical value: data-driven (χ^2_1 or χ^2_2) vs. fixed (χ^2_2)
- For $H_0 : \theta_1 = 0$, projection inference is not conservative; it has exact asymptotic coverage
  - AR/S-statistic does not depend on $\theta_2 \Rightarrow \chi^2_2$ distribution

Identification Robust Projection Inference



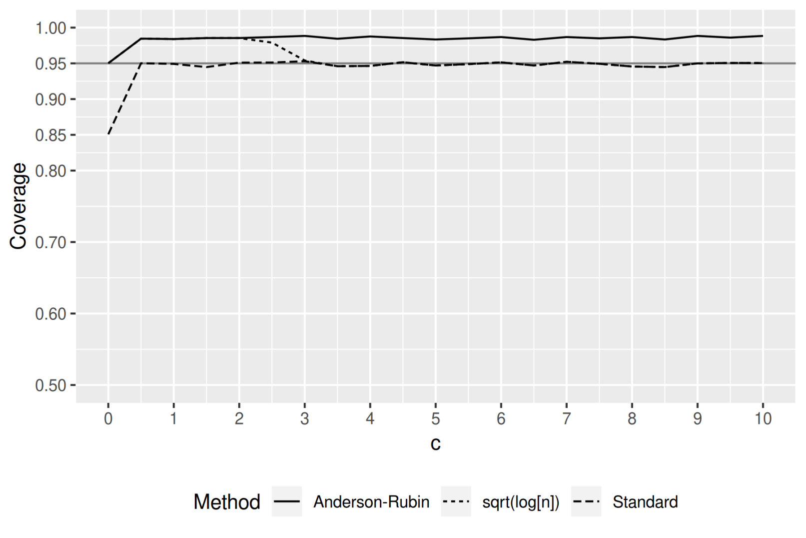
Note: $n = 1,000$, $\theta_1 = c/\sqrt{n}$, $\theta_2 = 5$, $x_1, x_2, e \sim \mathcal{N}(0,1)$

Standard QLR Inference



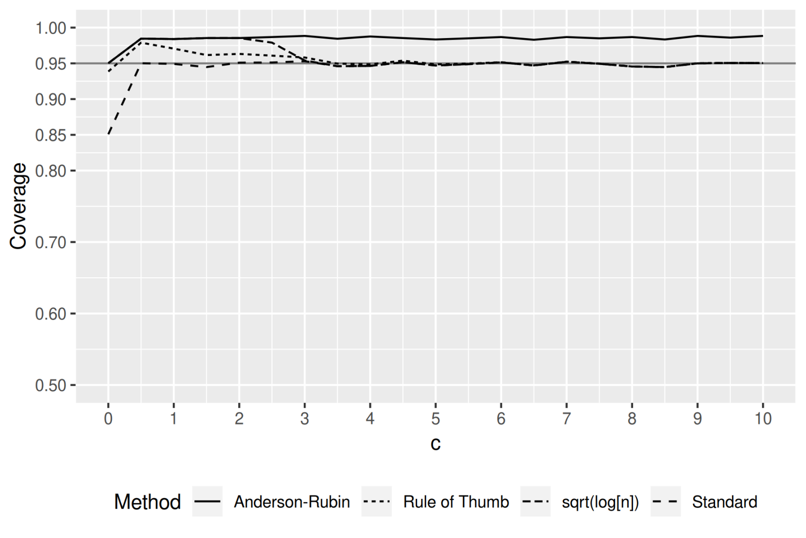
Note: $n = 1,000$, $\theta_1 = c/\sqrt{n}$, $\theta_2 = 5$, $x_1, x_2, e \sim \mathcal{N}(0,1)$

Two-Step Approach $\lambda_n = \sqrt{\log n / n}$



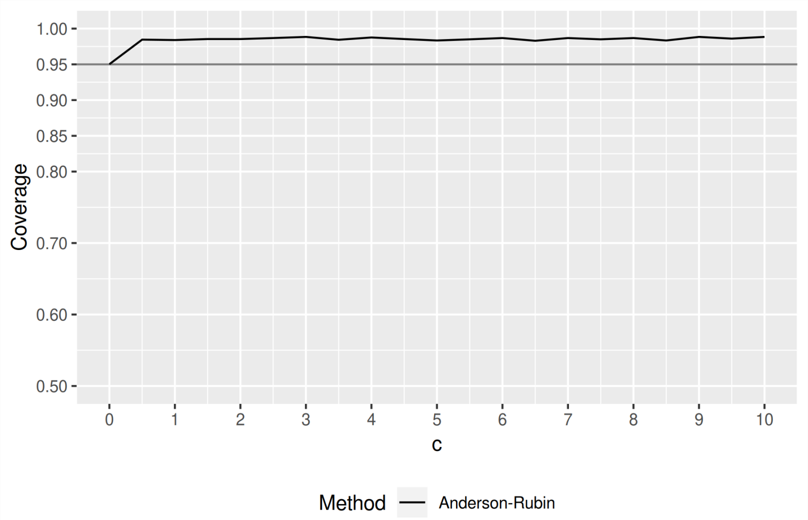
Note: $n = 1,000$, $\theta_1 = c/\sqrt{n}$, $\theta_2 = 5$, $x_1, x_2, e \sim \mathcal{N}(0,1)$

Two-Step Approach λ_n = Rule-of-Thumb



Note: $n = 1,000$, $\theta_1 = c/\sqrt{n}$, $\theta_2 = 5$, $x_1, x_2, e \sim \mathcal{N}(0,1)$

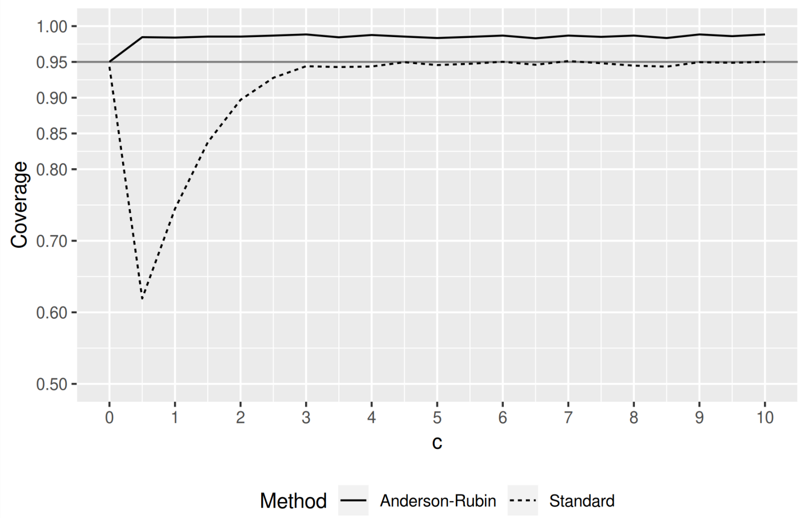
Identification Robust Projection Inference



Note: $n = 1{,}000$, $\theta_1 = c/\sqrt{n}$, $\theta_2 = 5$, $x_1, x_2, e \sim \mathcal{N}(0, 1)$

Example 2 | Example 3 | Empirical Application

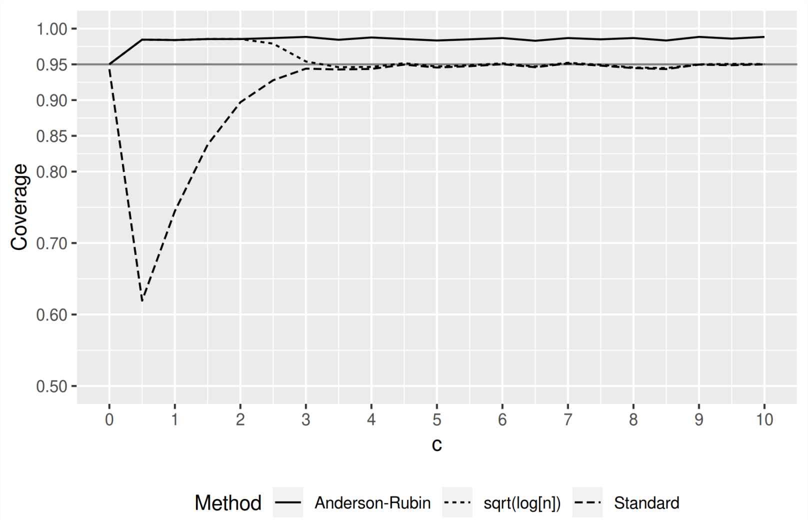
Standard Wald Inference



Note: $n = 1,000$, $\theta_1 = c/\sqrt{n}$, $\theta_2 = 5$, $x_1, x_2, e \sim \mathcal{N}(0, 1)$

Example 2 Example 3 Empirical Application

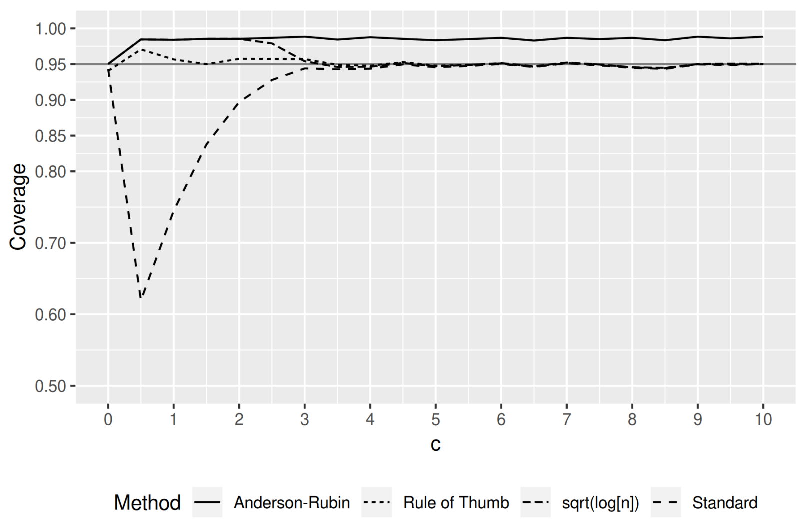
Two-Step Approach $\lambda_n = \sqrt{\log n / n}$



Note: $n = 1,000$, $\theta_1 = c/\sqrt{n}$, $\theta_2 = 5$, $x_1, x_2, e \sim \mathcal{N}(0,1)$

Example 2 Example 3 Empirical Application

Two-Step Approach $\lambda_n =$ Rule-of-Thumb



Note: $n = 1,000,\ \theta_1 = c/\sqrt{n},\ \theta_2 = 5,\ x_1, x_2, e \sim \mathcal{N}(0,1)$

Example 2 Example 3 Empirical Application

Conclusion

Conclusion: A Simple Solution to a Complex Problem

- Covers a wide range of moments and identification failures
- Computationally attractive: massively parallel
- Open questions
    i. Beyond GMM: general M-estimation problems
    ii. From type I to uniform type II inferences?
    iii. Identification failure in semi-nonparametric models?

THANK YOU!

References

Anderson, T. W. and Rubin, H. (1949). Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *The Annals of Mathematical Statistics*, 20(1):46–63.

Andrews, D. W. and Cheng, X. (2012). Estimation and Inference With Weak, Semi-Strong, and Strong Identification. *Econometrica*, 80(5):2153–2211.

Andrews, D. W. K. and Guggenberger, P. (2019). Identification- and Singularity-Robust Inference for Moment Condition Models. *Forthcoming in Quantitative Economics*.

Andrews, I. and Mikusheva, A. (2016). Conditional Inference With a Functional Nuisance Parameter. *Econometrica*, 84(4):1571–1612.

Antoine, B. and Renault, E. (2012). Efficient minimum distance estimation with multiple rates of convergence. *Journal of Econometrics*, 170(2):350–367.

Antoine, B. and Renault, E. (2017). Testing Identification Strength. Discussion papers, Department of Economics, Simon Fraser University.

References ii

Arellano, M., Hansen, L. P., and Sentana, E. (2012). Underidentification? *Journal of Econometrics*, 170(2):256–280.

Bravo, F., Carlos Escanciano, J., and Otsu, T. (2012). A Simple Test for Identification in GMM under Conditional Moment Restrictions. In *Badi H. Baltagi, R. Carter Hill, Whitney K. Newey, Halbert L. White (ed.) Essays in Honor of Jerry Hausman (Advances in Econometrics, Volume 29) Emerald Group Publishing Limited*, pages 455–477.

Brown, B. W. (1983). The Identification Problem in Systems Nonlinear in the Variables. *Econometrica*, 51(1):175.

Chen, X., Christensen, T. M., and Tamer, E. (2018). Monte Carlo Confidence Sets for Identified Sets. *Econometrica*, 86(6):1965–2018.

Cheng, X. (2015). Robust inference in nonlinear models with mixed identification strength. *Journal of Econometrics*, 189(1):207–228.

Chernozhukov, V. and Hong, H. (2003). An MCMC approach to classical estimation. *Journal of Econometrics*, 115(2):293–346.

Choi, I. and Phillips, P. C. (1992). Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations. *Journal of Econometrics*, 51(1-2):113–150.

Creel, M., Gao, J., Hong, H., and Kristensen, D. (2015). Bayesian Indirect Inference and the ABC of GMM. (Ses 1459975).

Dovonon, P. and Hall, A. R. (2018). The asymptotic properties of GMM and indirect inference under second-order identification. *Journal of Econometrics*, 205(1):76–111.

Dufour, J.-M. (1997). Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models. *Econometrica*, 65(6):1365.

Dufour, J.-M. and Taamouti, M. (2005). Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments. *Econometrica*, 73(4):1351–1365.

Fisher, F. M. (1967). The Identification Problem in Econometrics. *Economica*, 34(135):344.

Gospodinov, N. and Ng, S. (2015). Minimum Distance Estimation of Possibly Noninvertible Moving Average Models. *Journal of Business & Economic Statistics*, 33(3):403–417.

Inoue, A. and Rossi, B. (2011). Testing for weak identification in possibly nonlinear models. *Journal of Econometrics*, 161(2):246–261.

References iv

Kleibergen, F. (2005). Testing Parameters in GMM Without Assuming that They Are Identified. *Econometrica*, 73(4):1103–1123.

Komunjer, I. (2012). GLOBAL IDENTIFICATION IN NONLINEAR MODELS WITH MOMENT RESTRICTIONS. *Econometric Theory*, 28(04):719–729.

Koopmans, T. C. and Reiersol, O. (1950). The Identification of Structural Characteristics. *The Annals of Mathematical Statistics*, 21(2):165–181.

Lewbel, A. (2018). The Identification Zoo - Meanings of Identification in Econometrics. *Journal of Economic Literature, forthcoming*.

McCloskey, A. (2017). Bonferroni-based size-correction for nonstandard testing problems. *Journal of Econometrics*, 200(1):17–35.

Moon, H. R. H. and Schorfheide, F. (2012). Bayesian and Frequentist Inference in Partially Identified Models. *Econometrica*, 80(2):755–782.

Moreira, M. J. (2003). A Conditional Likelihood Ratio Test for Structural Models. *Econometrica*, 71(4):1027–1048.

Nelson, C. R. and Startz, R. (1990). Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator. *Econometrica*, 58(4):967.

References v

Newey, W. and McFadden, D. (1994). Large Sample Estimation and Hypothesis Testing. *Handbook of Econometrics*, 36(4):2111–2234.

Olea, J. L. M. and Pflueger, C. (2013). A Robust Test for Weak Instruments. *Journal of Business & Economic Statistics*, 31(3):358–369.

Owen, A. B. (2003). Quasi-monte carlo sampling. *Monte Carlo Ray Tracing: Siggraph*, 1:69–88.

Rothenberg, T. J. (1971). Identification in Parametric Models. *Econometrica*, 39(3):577.

Rotnitzky, A., Cox, D. R., Bottai, M., and Robins, J. (2000). Likelihood-Based Inference with Singular Information Matrix. *Bernoulli*, 6(2):243.

Staiger, D. and Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3):557.

Stock, J. H. and Wright, J. H. (2000). GMM with Weak Identification. *Econometrica*, 68(5):1055–1096.

Stock, J. H. and Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In Andrews, D. W. K. and Stock, J. H., editors, *Identification and Inference for Econometric Models*, pages 80–108. Cambridge University Press, Cambridge.

References vi

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wright, J. H. (2003). DETECTING LACK OF IDENTIFICATION IN GMM. *Econometric Theory*, 19(02).

Illustration: NLS

$$y_t = \theta_{1,n} x_{1,t} + \theta_{1,n}\theta_2 x_{2,t} + e_t$$

- Suppose $x_{1,t}, x_{2,t} \sim \mathcal{N}(0,1)$ uncorrelated
- Moments:

$$\bar{g}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} y_t (x_{1,t}, x_{2,t})' - \theta_1 (1, \theta_2)'$$

- Suppose $\theta_{1,n} = c_0/\sqrt{n}$ then for any c, θ_2

$$\sqrt{n} \times \bar{g}_n(c/\sqrt{n}, \theta_2) \xrightarrow{d} \mathcal{N}\left(\underbrace{c_0(1,\theta_{2,0})' - c(1,\theta_2)'}_{\text{Information}}, \underbrace{V}_{\text{Noise}}\right)$$

Main

$A_{n,LS}, B_{n,LS}$ - (Semi)-Strong Identification

Theorem (Semi-Strong Identification)

Suppose the model is (semi)-strongly identified, compact or exponential kernel conditions + assumptions above hold and

$$\kappa_n^2 = o\left[\lambda_{\min}(\partial_\theta g_n(\theta_0)'\partial_\theta g_n(\theta_0))\right]$$

then:

i. $A_{n,LS} = \bar{g}_n(\hat{\theta}_{n,GMM}) - B_{n,LS}\hat{\theta}_{n,GMM} + o_p(n^{-1/2})$

ii. $B_{n,LS}H_n = \partial_\theta g_n(\hat{\theta}_{n,GMM})H_n + o_p(1)$ *(full rank)*

iii. $H_n^{-1}[\hat{\theta}_{n,LS} - \hat{\theta}_{n,GMM}] = o_p(n^{-1/2})$

iv. $H_n^{-1}\Sigma_n H_n^{-1} = O_p(\tilde{\kappa}_n^2)$

$A_{n,LS}, B_{n,LS}$ - Higher-Order Local Identification

Theorem (Higher-Order Local Identification)

Suppose the model is higher-order locally identified at an order $r \geq 2$, compact/exponential kernel conditions + assumptions above hold then:

$$\Sigma_n = \sum_{j=1}^{r} P_j O_p(\tilde{\kappa}_n^{2/j}) P_j'$$

using the Lemma, this implies that:

a. $v_j \in Span(P_j) \Rightarrow B_{n,LS} v_j = O_p(\tilde{\kappa}_n^{1-1/j})$
b. $|\lambda_{\min}(B_{n,LS})| = O_p(\tilde{\kappa}_n^{1-1/r})$

Summary of the Results for $B_{n,LS/\infty}$ Main

References for Indentification Regimes

- Each regime has asymptotic implications for $\hat{\theta}_{n,GMM}$:
  1. $\Rightarrow$ consistent + $\sqrt{n}$ asymptotically normal (1.ii. $\Rightarrow \partial_\theta g_n$ full rank) (Newey and McFadden, 1994; van der Vaart, 1998)
  2. $\Rightarrow$ consistent + slower than $\sqrt{n}$ asymptotically normal (Antoine and Renault, 2012; Andrews and Cheng, 2012)
  3. $\Rightarrow$ consistent + slower than $\sqrt{n}$ convergent, not asymptotically normal (Rotnitzky et al., 2000; Dovonon and Hall, 2018)
  4. $\Rightarrow$ not consistent, not asymptotically normal (Staiger and Stock, 1997; Stock and Wright, 2000)

Main

Illustration

- We know that $B_{n,LS/\infty}$ is $O_p(\kappa_n)$ on the span of the identification failure V
- Example: $\theta = (\theta_1, \theta_2)$; $\theta_1 - \theta_2$ point identified, $\theta_1 + \theta_2$ set identified, the model is linear and:

$$B_n = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

- We could use $B_n(1,1) = (0,0)$ and fix $\theta_1 + \theta_2$ (eigenvector)...
- if we **want to fix something interpretable** i.e. θ_1 or θ_2:

$$B_{n,LS/\infty}(1,0) = (1,-1); \quad B_{n,LS/\infty}(0,1) = (-1,1)$$

- fixing either θ_1/θ_2 makes the other point identified as seen by:

$$B_n P^{\perp}_{(1,0)} \text{ and } B_n P^{\perp}_{(0,1)} \text{ have rank} = 1, B_n \text{ has rank } 1$$

Remarks on the assumptions

- The critical assumption is that the free parameters are (semi)-strongly identified so that the $T_{j,n}$s (e.g. S, K stat) have asymptotically correct size
- To be sure, we could compute $\tilde{B}_{n,LS/\infty}$ for $\bar{g}_n(\theta)$ on $\{R_{\hat{j}_n}\theta - c_{\hat{j}_n}\}$ and compare its eigenvalues with $\underline{\lambda}_n$
- Can also check for first-order identification failure after collapsing the identified set

Main

Empirical Applications

Assumptions a., b. Counter-Example

- Let
$$g_n(\theta) = \theta^4 \sin(1/\theta)$$
- g is smooth - but not analytical - has infinitely many zeros
- $\nexists \underline{c}, k$ such that $\underline{c} d(\theta, \Theta_0)^k \leq |g(\theta)|$
- For any $\varepsilon > 0$, $\hat{\pi}_n \left(\mathcal{B}_\varepsilon(0)\right) \to 1$
- Nb: if g_n were analytic with ∞-many zeros around $\theta_0 \Rightarrow g_n$ identically zero around θ_0

Main

Further Asymptotic Results

Proposition (Quasi-Central Limit Theorem)

Suppose that $\bar{g}_n$ is smooth, $\partial_\theta g_n$ satisfies stoch. equicont., CLT then under (semi)-strong identification:

$$\sqrt{n}\begin{pmatrix} A_{n,LS} - B_{n,LS}\theta_0 \\ vec(B_{n,LS} - \bar{B}_{nLS}) \end{pmatrix} = \sqrt{n}\begin{pmatrix} \bar{g}_n(\theta_0) \\ vec(\partial_\theta \bar{g}_n(\theta_0) - g_n(\theta_0)) \end{pmatrix} + o_p(1)$$

$$\xrightarrow{d} \mathcal{N}(0, V)$$

where

$$\bar{B}_{n,LS} = \Sigma_n^{-1}\int_\Theta (\theta - \bar{\theta}_n)\int_\Theta \{g_n(\theta) - g_n(\tilde{\theta})\}'\hat{\pi}_n(\tilde{\theta})d\tilde{\theta}\hat{\pi}_n(\theta)d\theta$$

Remark on (Quasi)-CLT for $A_{n,LS}, B_{n,LS}$

- After some re-centering, we always have

$$\begin{aligned}&\sqrt{n}[B_{n,LS} - \bar{B}_{n,LS}]\\&= \Sigma_n^{-1} \int_\Theta (\theta - \bar{\theta}_n) \int_\Theta [\mathbb{G}_n(\theta) - \mathbb{G}_n(\tilde{\theta})]' \hat{\pi}_n(\tilde{\theta}) d\tilde{\theta} \hat{\pi}_n(\theta) d\theta\end{aligned}$$

- $\text{int}(\Theta_0) \neq \emptyset$ implies it is a sequence of bounded linear operators applied to an empirical process; which can be used to prove a CLT
- Higher-order local identification and manifold valued identified set are more difficult...

Main

Practical Implications: which parameters to fix?

- We can re-write each $\theta \in \Theta_0$ as:

$$\theta = \theta_0 + v,\; v \in V = \mathrm{Span}(\{\theta_1 - \theta_0,\, \theta_0, \theta_1 \in \Theta_0\})$$

- For the projection matrix P_V and the orthogonal $P_V^{\perp}$:

$$P_V\theta = P_V\theta_0 + v, \quad P_V^{\perp}\theta = P_V^{\perp}\theta_0 + 0$$

- The first one is not unique: v can vary
- The second one is unique $\Rightarrow$ identified

Practical Implications: which parameters to fix?

- Suppose $(u, v^\star)$ forms a basis with $\text{rank}(P_V^\perp P_{v^*}^\perp) = \text{rank}(P_{v^*}^\perp)$
- Pick $\theta_1 \in \Theta_0$ with $P_{v^\star}\theta_1 = c$ fixed

$$P_V^\perp \theta_1 = P_V^\perp (P_{v^\star}^\perp \theta_1 + P_{v^\star}\theta_1) = P_V^\perp P_{v^\star}^\perp \theta_1 + P_V^\perp \underbrace{P_{v^\star}\theta_1}_{=c \text{ fixed}}$$

- Since $P_V^\perp \theta_1 = P_V^\perp \theta_0$, we have the system:

$$P_V^\perp P_{v^\star}^\perp \theta_1 = P_V^\perp P_{v^\star}^\perp \theta_0 - P_V^\perp c$$
$$P_{v^\star}\theta_1 = c$$

- Rk: $\text{rank}(P_V^\perp P_{v^*}^\perp) = \text{rank}(P_{v^*}^\perp) \Rightarrow$ the system has full rank

$\Rightarrow$ The solution is unique: θ is identified up to $P_{v^\star}\theta$ (fixed)

Main

Remarks

- Weak/set: if free (nuisance) parameters (semi)-strongly identified when $\ell \geq \ell^\star$ and $S_{n,\ell} =$ S/K/cQLR statistic:

$$\mathbb{P}(S_{\hat{\ell}_n,n} \leq c_{1-\alpha,\hat{\ell}_n}) \geq 1 - \alpha + o(1)$$

- Semi-strong: $\lambda_{\min}[\partial_\theta g_n(\theta_0)]$ slightly larger than $O(n^{-1/2}) \Rightarrow$ false positives... better than false negatives?

On the Cutoff $\underline{\lambda}_n$ for the Eigenvalues (Just-Identified)

- Which cutoff $\lambda_{min}(B_{n,LS}) \leq \underline{\lambda}_n$ to detect identification failure?
- Similar to (Stock and Yogo, 2005): just-identified + gaussian

$$\bar{g}_n(\theta) = A_n + B_n(\theta - \theta_0)$$

$$A_n = \bar{g}_n(\theta_0) - B_n\theta_0,\ B_n - \overline{B}_n = \Delta_n = O_p(1/\sqrt{n})$$

quasi-CLT for $A_{n,LS}, B_{n,LS}$

- Using the Woodbury identity recursively:

$$\hat{\theta}_n - \theta_0 = \underbrace{-\overline{B}_n^{-1}\bar{g}_n(\theta_0)}_{\text{CLT term}} + \underbrace{\overline{B}_n^{-2}\Delta_n\bar{g}_n(\theta_0)}_{\text{Non-Standard Term}} - \overline{B}_n^{-3}\Delta_n^2 \dots$$

$$\Rightarrow \text{bias} \simeq \overline{B}_n^{-2}\underbrace{\mathbb{E}[\Delta_n\bar{g}_n(\theta_0)]}_{=O(1/n)} \quad \text{variance} \simeq \overline{B}_n^{-1}\underbrace{\mathbb{V}[\bar{g}_n(\theta_0)]}_{=O(1/n)}\overline{B}_n^{-1\prime}$$

A Rule-of-Thumb for $\underline{\lambda}_n$

- Rate of convergence depends on $\lambda(\overline{B}_n^{-1})$
- Pick $v_{j,n}$ (complex) left-eigenvector of $\bar{B}_{n,LS}$:

$$v_{j,n}(\hat{\theta}_n - \theta_0) = \lambda_j^{-1} v_{j,n} \bar{g}_n(\theta_0) + \lambda_j^{-2} v_{j,n} \Delta_n \bar{g}_n(\theta_0) + \dots$$

- Size distortion in that direction depends on (*bias*^{2}/*variance*):

$$\frac{1}{n|\lambda_j|^2} \frac{v_{j,n}^\star V_{12} V_{21} v_{j,n}}{v_{j,n}^\star V_1 v_{j,n}} \leq \frac{1}{n|\lambda_{\min}(\bar{B}_{n,LS})|^2} \frac{|V_{12} V_{21}|}{\lambda_{\min}(V_1)}$$

- Design cutoff $\underline{\lambda}_n$ based on a sequence of size distortions $\searrow 0$
- Over-identified: involves W as well
- Higher-Order: residual curvature matters

Additional Remarks

Remarks

- Rule-of-thumb designed for problems with $\bar{g}_n$ flat around θ_0
  - Counter-example: MA(1) locally identified but not globally
- Alternative Representation:
  - Think of $\Theta_0 = \cup_{j=1}^{k} S_j$ - disjoint sets S_j - then $\hat{\theta}_n \in \mathcal{N}(S_j)$ for some $j \in \{1, \ldots, k\}$ wp $\nearrow 1$
  - Compute a Wald statistic for $H_0 : \theta = \theta_{0,j^\star} \in S_{j^\star}$
  - Size distortions: within ($j = j^\star$) and between sets ($j \neq j^\star$)
- Simple idea: partition $\hat{\Theta}_{0,n} = \{\theta, \|\bar{g}_n(\theta)\|_W \leq \kappa_n\}$ using cluster algorithm (e.g. k-means), then
  - Compute rule-of-thumb within cluster (as prev. slides)
  - Compute rule-of-thumb between clusters (distance)

Example 2: MA(1)

- Simple example:

$$y_t = \sigma[e_t + \vartheta e_{t-1}], e_t \sim (0, 1, \tau)$$

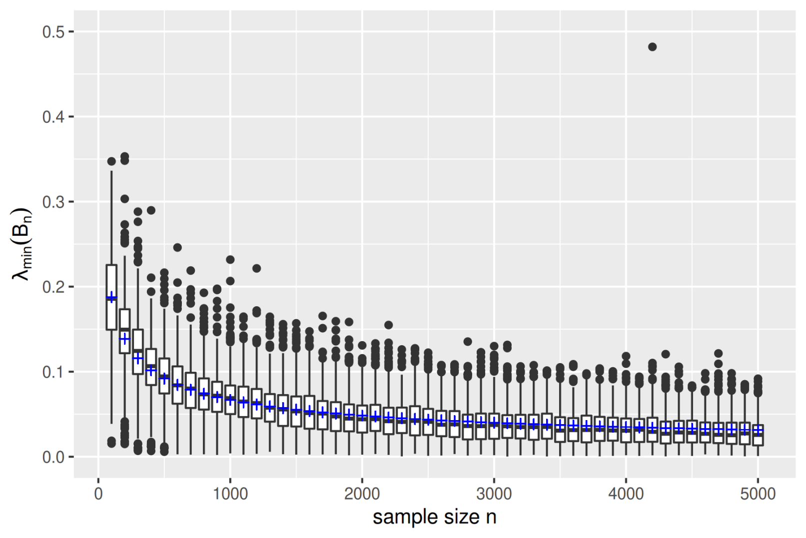
- Identification failure $\tau = 0$, weak identification $\tau \simeq 0$
- Two cutoffs $\underline{\lambda_n}$: $\sqrt{\log(n)/n}$, rule-of-thumb
- Null hypothesis:

$$H_0 : \theta = \theta_0 = 2$$

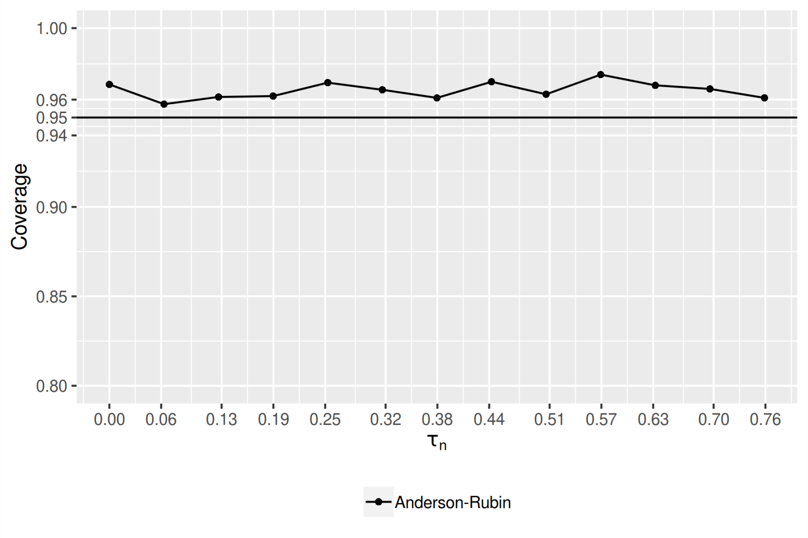
- With 4 estimating moments and
  - $\tau_n = 2 \times n^{-1/2}$, $e_t \sim GEV(0, 1, \tau_n)$
  - $\kappa_n = \max(q_{0.99}(\chi^2_4), \sqrt{2\log(\log[n])/n})$
- Compare AR (χ^2_2 critical value: oracle), Wald/QLR and Two-Step

Main

Example 2: MA(1) - Distribution of $\lambda_{\min}(B_{n,LS})$

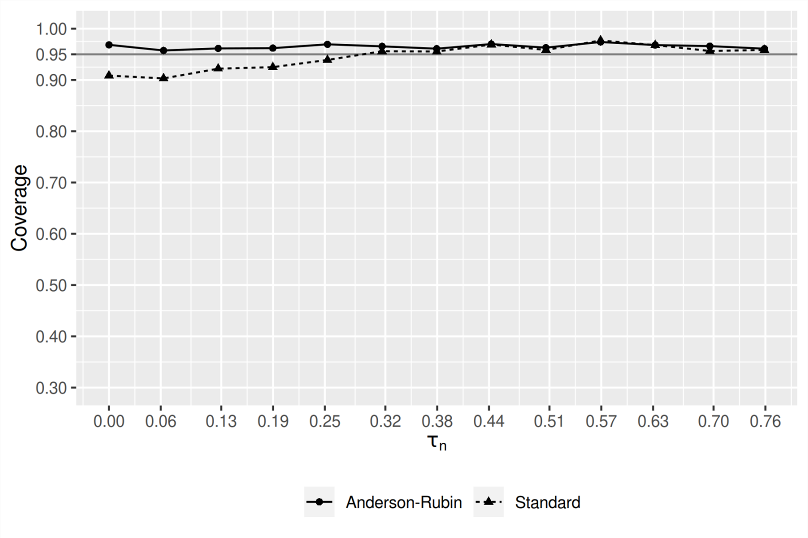


Identification Robust Projection Inference



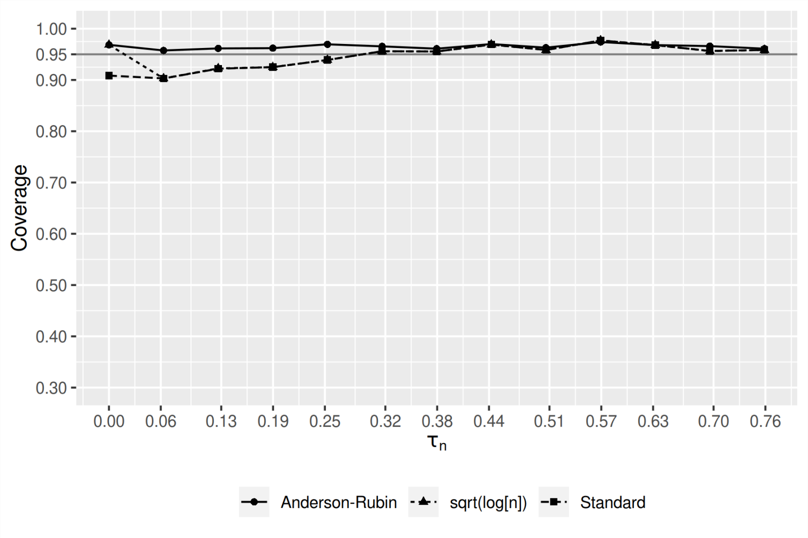
Note: $n = 1,000$, $\tau = c/\sqrt{n}$, $\vartheta_0 = 2$, $e_t \sim \mathcal{N}(0,1)$

Standard QLR Inference



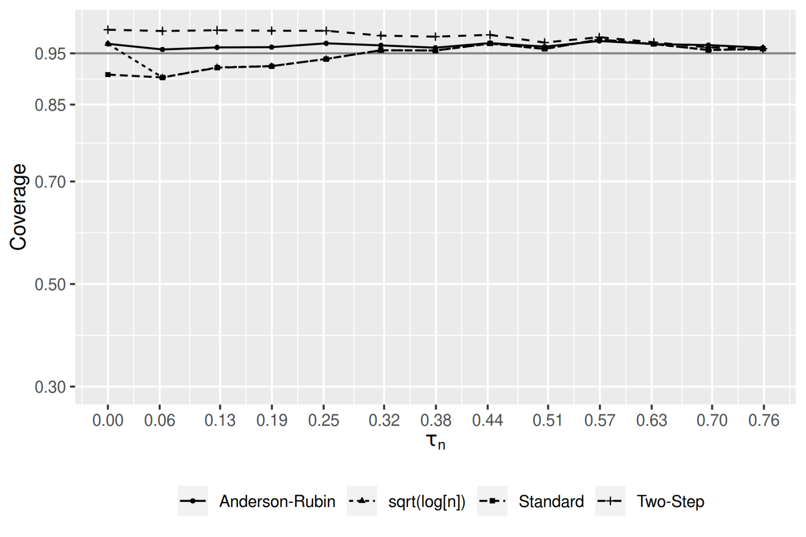
Note: $n = 1,000$, $\tau = c/\sqrt{n}$, $\vartheta_0 = 2$, $e_t \sim \mathcal{N}(0,1)$

Two-Step Approach $\lambda_n = \sqrt{\log n / n}$



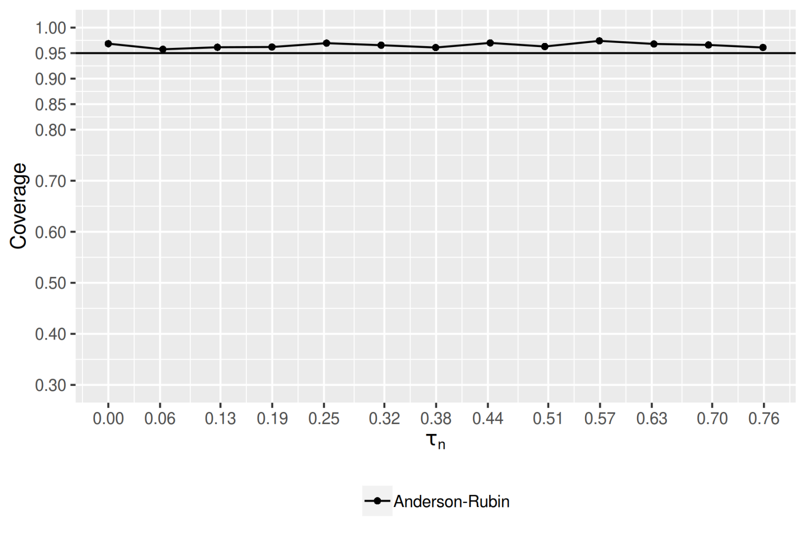
Note: $n = 1,000$, $\tau = c/\sqrt{n}$, $\vartheta_0 = 2$, $e_t \sim \mathcal{N}(0,1)$

Two-Step Approach λ_n = Rule-of-Thumb



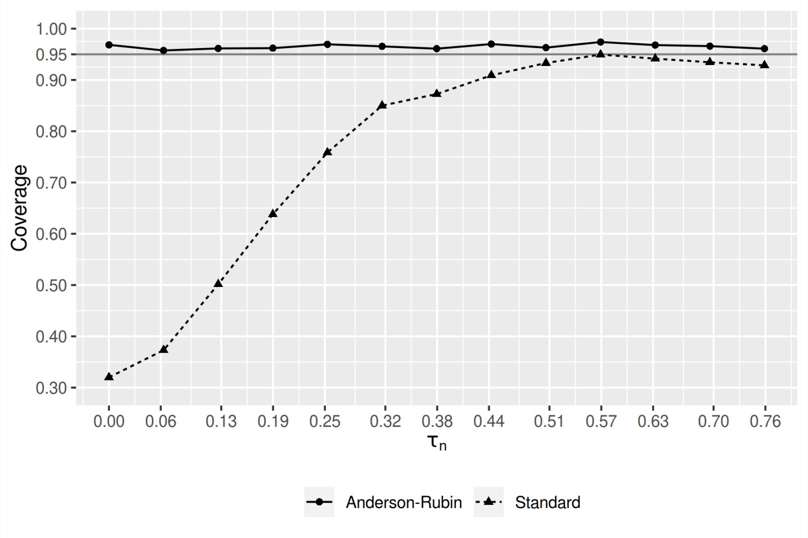
Note: $n = 1,000$, $\tau = c/\sqrt{n}$, $\vartheta_0 = 2$, $e_t \sim \mathcal{N}(0, 1)$

Identification Robust Projection Inference



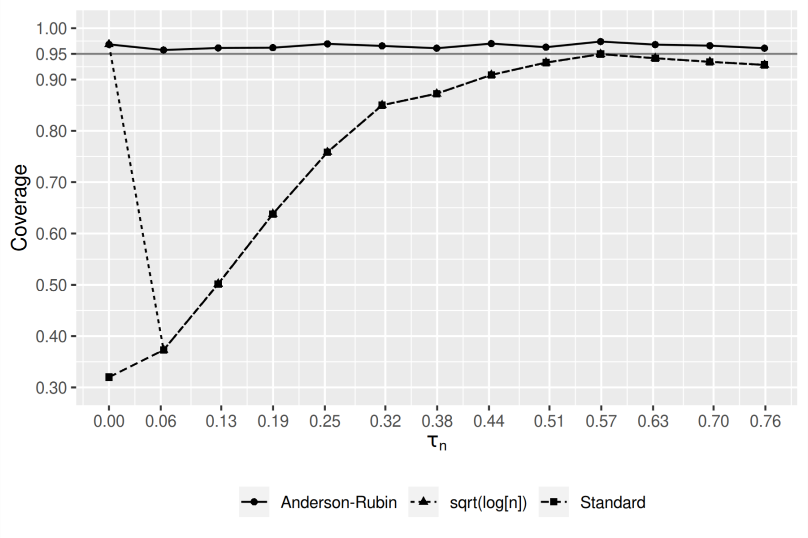
Note: $n = 1,000$, $\tau = c/\sqrt{n}$, $\vartheta_0 = 2$, $e_t \sim \mathcal{N}(0,1)$ Main

Standard Wald Inference



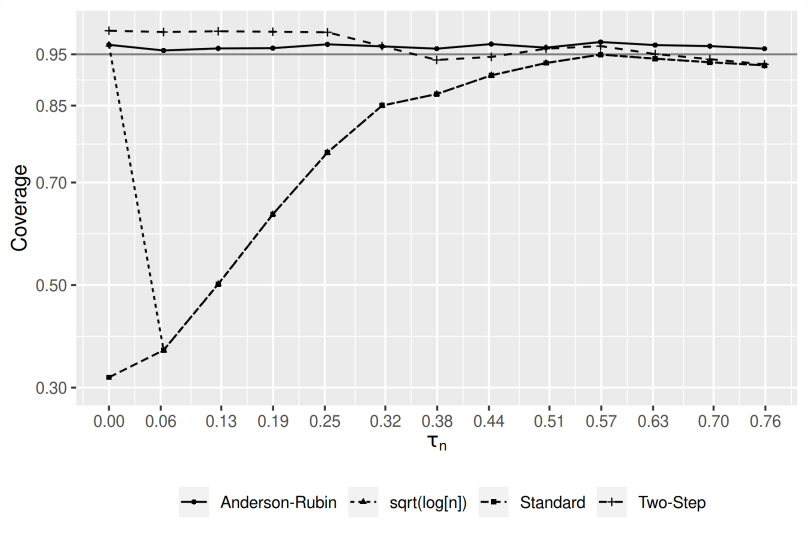
Note: $n = 1,000$, $\tau = c/\sqrt{n}$, $\vartheta_0 = 2$, $e_t \sim \mathcal{N}(0,1)$ Main

Two-Step Approach $\lambda_n = \sqrt{\log n / n}$



Note: $n = 1,000,\ \tau = c/\sqrt{n},\ \vartheta_0 = 2,\ e_t \sim \mathcal{N}(0,1)$ Main

Two-Step Approach λ_n = Rule-of-Thumb



Note: $n = 1,000$, $\tau = c/\sqrt{n}$, $\vartheta_0 = 2$, $e_t \sim \mathcal{N}(0,1)$ Main

Example 3: Higher-Order Identified NLS

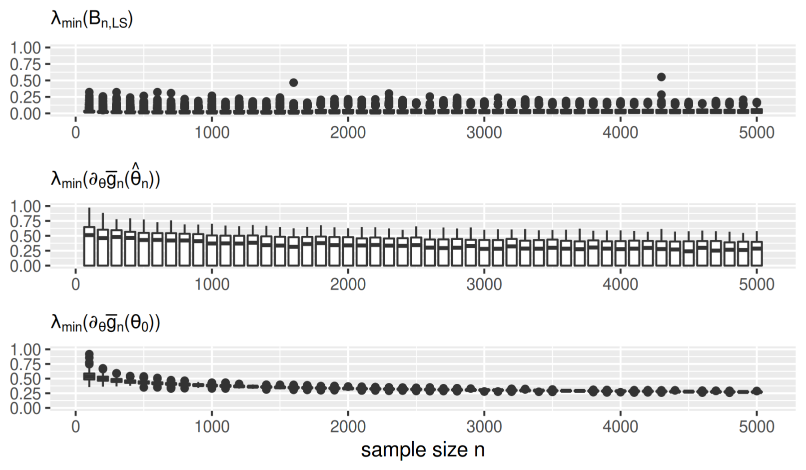
- Simple example:

$$y_t = \theta_1 x_{i,1} + \theta_{2,n}(\theta_{2,n} - \theta_1)^2 x_{2,i} + e_i, (x_{i,1}, x_{i,2}, e_i) \sim \mathcal{N}(0, I_3)$$

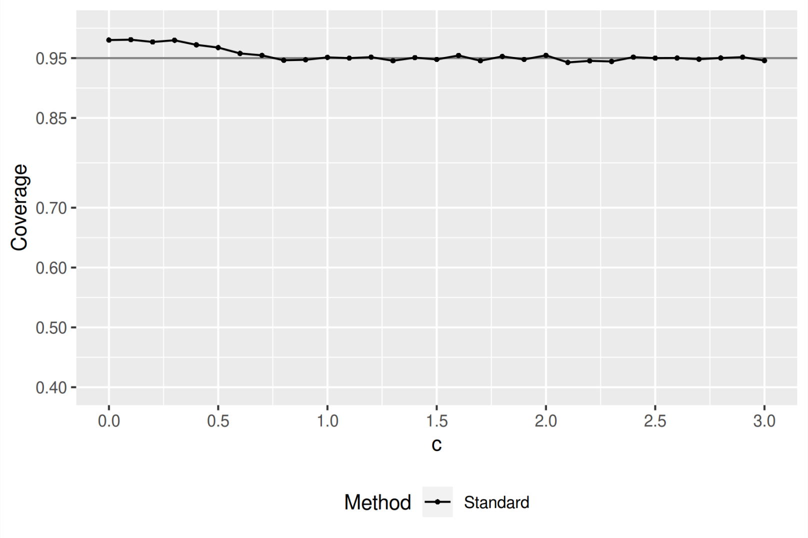
- Higher-order identification $\theta_{2,n} - \theta_1 = O(n^{-1/4})$
- Cutoff $\underline{\lambda_n}$: based on rule-of-thumb
- Estimating moments $\mathbb{E}(y_i(x_{i,1}, x_{i,2})) - (\theta_1, \theta_2(\theta_2 - \theta_1)^2)$
- Compare AR (χ^2_2 critical value: oracle), Wald/QLR and Two-Step

Main

Example 3: NLS - Distribution of $\lambda_{\min}(B_{n,LS})$

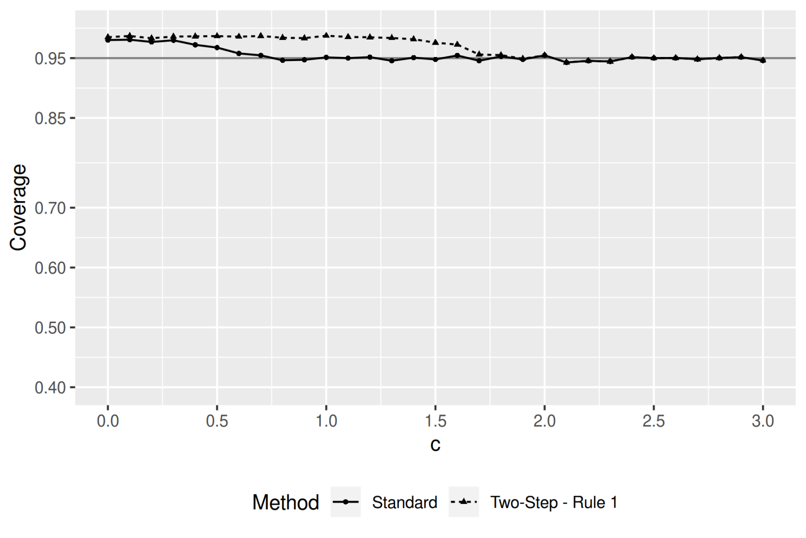


Standard QLR Inference



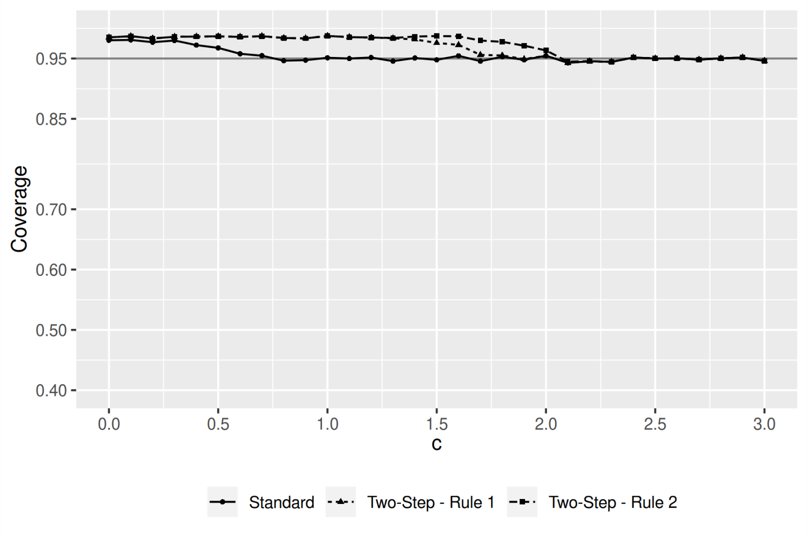
Note: $n = 1,000$

Two-Step: Rule-of-Thumb 1



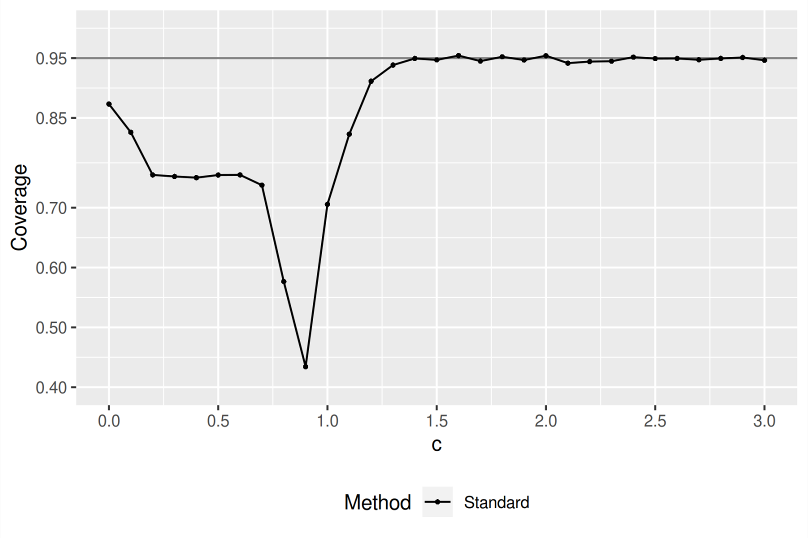
Note: $n = 1,000$

Two-Step: Rule-of-Thumb 2



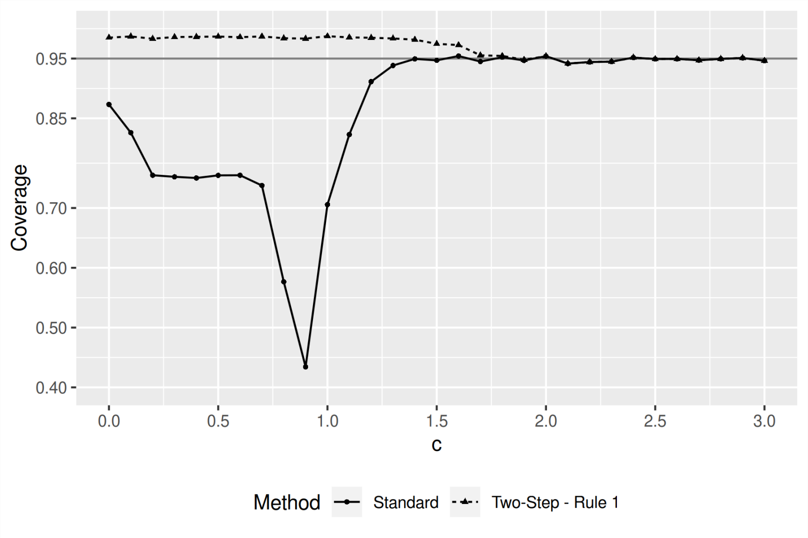
Note: $n = 1,000$

Standard Wald Inference



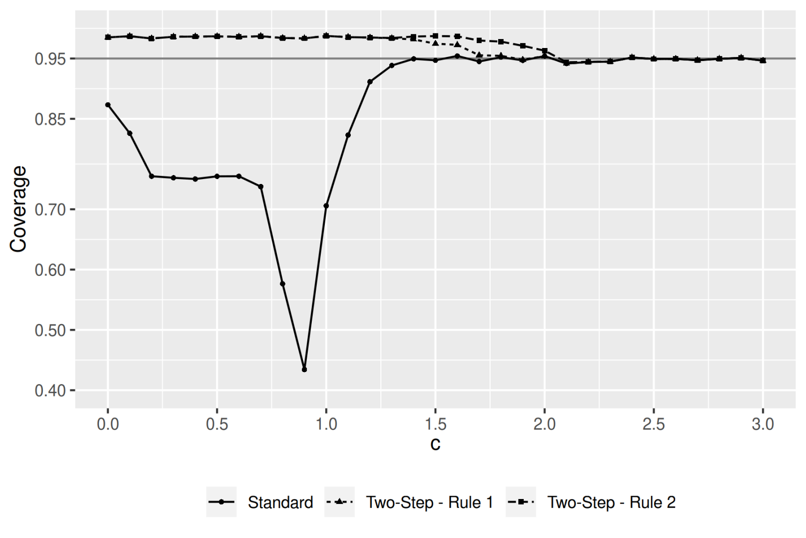
Note: $n = 1,000$ Main

Two-Step: Rule-of-Thumb 1



Note: $n = 1,000$ Main

Two-Step: Rule-of-Thumb 2



Note: $n = 1,000$ Main

Empirical Illustration

Illustration: Euler Equation

- Data: Stock and Wright (2000)
- Model:

$$\mathbb{E}\left(\left[\delta\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma} R_t - 1\right] Z_t\right) = 0$$

- $Z_t = (1, C_{t-1}, R_{t-1})$, $n = 103$ after taking lags
- W = Continuously-Updated Newey-West
- Bounds: $(\delta, \gamma) \in [0.7, 1.2] \times [0, 20]$
- Grid: 10^4 points from the Sobol sequence (quasi Monte-Carlo, see e.g. Owen, 2003, for an introduction)
- **Compute a quasi-Jacobian matrix $B_{n,LS}$ that summarizes the identifiability of (δ, γ)**

Illustration: $\hat{\Theta}_n = \left\{ \theta \in \Theta, \|\bar{g}_n(\theta)\|_W - \|\bar{g}_n(\hat{\theta}_n)\|_W \leq \kappa_n \right\}$

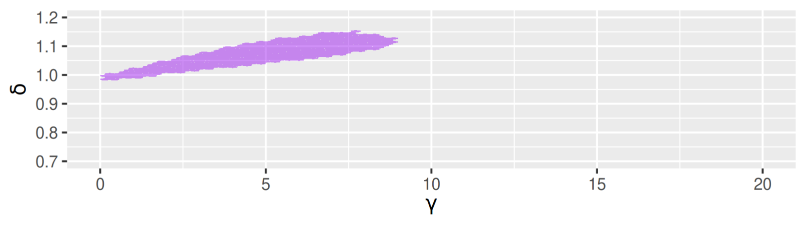


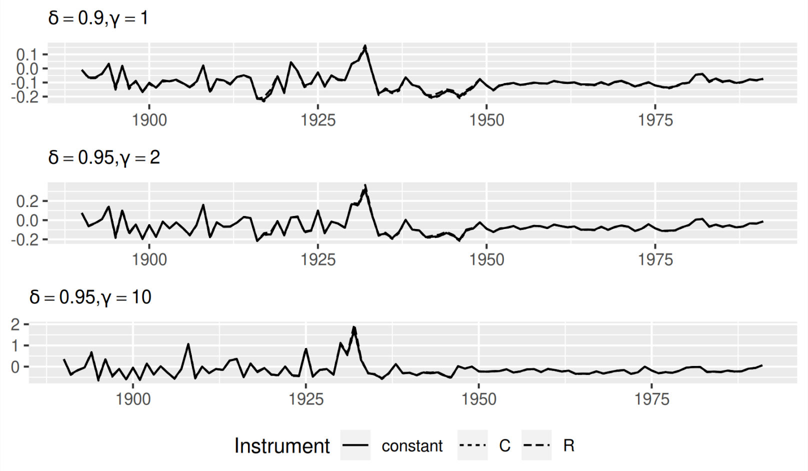
Illustration: Euler Equation - Linear Approximation

- Results $\theta = (\delta, \gamma)$

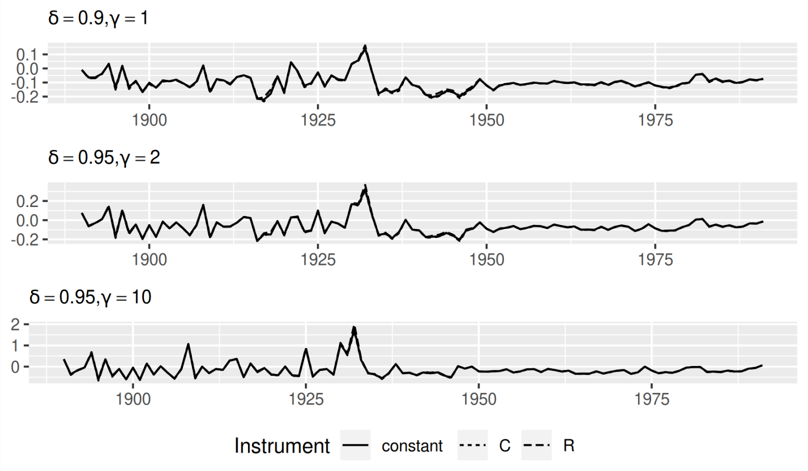
$$B_{n,LS} = \begin{pmatrix} 0.669 & -0.001 \\ 0.685 & -0.001 \\ 0.682 & 0.000 \end{pmatrix}$$

- Note that $\lambda(B_{n,LS}) \times \sqrt{n} = (11.929, 0.006)$

The Identification Problem in the Euler Equation



The Identification Problem in the Euler Equation



Moments are singular: amount to a single moment condition

Identification Robust Inference

- Require Singularity and Identification Robust Inference (Andrews and Guggenberger, 2019)
- Drop 2 moments, keep $Z_t = 1$; invert an AR test with χ^2_1 critical value:

$$CI_{95\%}(\delta) = [0.98, 1.17]; CI_{95\%}(\gamma) = [0.03, 20]$$

Main