Inference under First-Order Degeneracy

Xinyue Bei — *UT Austin*
Manu Navjeevan — *Texas A &M University*

CIREQ-HEC Women in Econometrics Conference

June 13, 2026

Introduction

Classical Setup: Researcher observes estimate $\hat{\theta}$ of parameter $\theta \in \Theta \subseteq \mathbb{R}^d$

$$r_n(\hat{\theta} - \theta) \rightsquigarrow \mathcal{W}$$

- Object of interest is $g(\theta) : \Theta \to \mathbb{R}$
- The delta method relies on a non-zero gradient $\nabla g(\theta)$

This paper: inference on $g(\theta)$ near a point $\theta_\star$ such that

$$\nabla g(\theta_\star) = 0$$

Examples: Mediation Analysis

In causal mediation analysis, see e.g. Baron and Kenny (1986)

$$M = \theta_1 X + V$$
$$Y = \gamma X + \theta_2 M + U$$

Indirect effect can be written as

$$g(\theta) = \theta_1 \theta_2$$

- θ_1: Effect of treatment on mediator
- θ_2: Effect of mediator on outcome
- At $\theta_\star = 0$, $\nabla g(\theta_\star) = (\theta_{\star,2}, \theta_{\star,1})' = 0$.

Examples: Mediation Analysis

In causal mediation analysis, see e.g. Baron and Kenny (1986)

$$M = \theta_1 X + V$$
$$Y = \gamma X + \theta_2 M + U$$

Indirect effect can be written as

$$g(\theta) = \theta_1 \theta_2$$

- θ_1: Effect of treatment on mediator
- θ_2: Effect of mediator on outcome
- At $\theta_\star = 0$, $\nabla g(\theta_\star) = (\theta_{\star,2}, \theta_{\star,1})' = 0$.

Recent works have proposed tests for the specific null alternate pair

$$H_0 : g(\theta) = 0 \text{ v.s. } H_1 : g(\theta) \neq 0$$

- van Garderen & van Giersbergen (2024, REStat), Dufour, Renault, Zinde-Walsh (2025, Ann. Stat.), etc
- However, these works do not consider CIs in local regions of the origin

Problems with Delta-Method

Suppose that $\theta = \theta_\star + h/r_n$.

Under this asymptotic regime

$$r_n^2(g(\hat{\theta}) - g(\theta)) \rightsquigarrow 2h'\nabla^2 g(\theta_\star)\mathcal{W} + \mathcal{W}'\nabla^2 g(\theta_\star)\mathcal{W}$$

- The limiting distribution depends on the local parameter h which is not consistently estimable.
- Implication for inference: Standard Wald-type approaches may not work

Questions and Preview of Results

This leaves some open questions:

1. Is there some alternative to the plug-in estimator whose behavior can be approximated around $\theta_\star$?

Questions and Preview of Results

This leaves some open questions:

1. Is there some alternative to the plug-in estimator whose behavior can be approximated around $\theta_\star$?
   - Answer: No
   - Establish that regular estimation of $g(\theta)$ is impossible in local regions of $\theta_\star$

Questions and Preview of Results

This leaves some open questions:

1. Is there some alternative to the plug-in estimator whose behavior can be approximated around $\theta_\star$?
   - Answer: No
   - Establish that regular estimation of $g(\theta)$ is impossible in local regions of $\theta_\star$
2. Are there alternatives to Wald-type inference that may still perform well?

Questions and Preview of Results

This leaves some open questions:

1. Is there some alternative to the plug-in estimator whose behavior can be approximated around $\theta_\star$?
   - Answer: No
   - Establish that regular estimation of $g(\theta)$ is impossible in local regions of $\theta_\star$
2. Are there alternatives to Wald-type inference that may still perform well?
   - Answer: Yes
   - We propose a minimum distance based methods
   - Establish sufficient conditions under which χ^2 critical values can be applied
   - Propose simple bootstrap procedures when these conditions are not met

Related Literature

1. Point Hypothesis Testing under Degeneracy
   - Dufour, Renault, Zinde-Walsh (2025), Dufour and Valéry (2025), etc
     - uses Wald statistic, divergence issues
     - focuses on testing problems where the null itself contains the singularity
   - This paper
     - A minimum distance–based test statistic
     - constructs uniformly valid CI when the null may be near a singularity
2. Econometric Impossibility Results
   - Hirano and Porter (2012): directionally differentiable functions
   - We establish impossibility in new setting not yet considered by literature
3. Testing Non-Linear Restrictions
   - Andrews and Mikusheva (2016)
     - minimum-distance statistics & maximal curvature of the null manifold
   - We show that existing methods can be improved upon by exploiting certain structure on null-manifold.

Contents

Outline

Setup

Assume that the researcher observes data $X^{(n)} = (X_1, \ldots, X_n)$ from a parametric model $P_{n,\theta}$, $\theta \in \Theta$, that is locally asymptotically normal (Le Cam, 1960)

Definition (Local Asymptotic Normality)

There is a sequence $r_n \to \infty$ such that for every $\theta \in \Theta^o$ and every sequence $h_n \to h \in \mathbb{R}^d$

$$\log\left(\frac{dP_{n,\theta+h_n/r_n}}{dP_{n,\theta}}(X^{(n)})\right) = h'\Delta_n - \frac{1}{2}h'\Gamma_\theta h + o_{P_{n,\theta}}(1)$$

where Δ_n converges in distribution to $N(0, \Gamma_\theta)$ under $P_{n,\theta}$ and Γ_θ is invertible.

Impossibility: Local Behavior of Estimators

- Adopt the local parameterization $\theta_{n,h} = \theta_\star + h/r_n$ and let $P_{n,h} = P_{n,\theta_{n,h}}$
- Estimator Ψ_n is an arbitrary function of the data satisfying

$$r_n^2(\Psi_n - g(\theta_{n,h})) \rightsquigarrow \mathcal{L}_h$$

  along the sequence of distributions $P_{n,h}$, for every $h \in \mathbb{R}^d$
- Limit distribution of estimator depends on local parameter h

Limit Experiment

Proposition (Limit Experiment)

There exists a randomized statistic $\Psi(Z, U)$ *where* Z *is drawn from the Gaussian shift experiment*

$$Z \sim N(h, \Gamma_{\theta_\star}^{-1})$$

and $U \sim \mathrm{Unif}(0, 1)$ *independent of* Z*, such that*

$$\Psi(Z, U) - \frac{1}{2} h' \nabla^2 g(\theta_\star) h \overset{h}{\sim} \mathcal{L}_h \quad \text{for all } h \in \mathbb{R}^d$$

- Estimating $g(\theta)$ in local regions of $\theta_\star \Leftrightarrow$ estimating a quadratic form of the shift parameter in a Gaussian shift experiment.

Impossibility: Intuition

Suppose that there is an estimator sequence Ψ_n whose limiting behavior does not depend on h, that is $\mathcal{L}_h = \mathcal{L}$ for all $h \in \mathbb{R}^d$.

- There must be an estimator $\Psi(Z, U)$ such that for all $h \in \mathbb{R}^d$,

$$\Psi(Z, U) - \frac{1}{2} h' \nabla^2 g(\theta_\star) h \overset{h}{\sim} \mathcal{L}$$

- To build intuition suppose that $\mathcal{L}$ has a finite second moment.
- Cramér-Rao lower bound tells us that for any unbiased estimator

$$\text{Var}_h(\Psi(Z, U)) \geq h'(\nabla^2 g(\theta_\star)) \Gamma_{\theta_\star}^{-1} (\nabla^2 g(\theta_\star)) h$$

Contradiction: This bound can be made arbitrarily large by varying h over $\mathbb{R}^d$.

Impossibility: Formal Result

Theorem (No Regular Estimator.)

There is no estimator sequence Ψ_n and law $\mathcal{L}$ on $\mathbb{R}$ such that

$$r_n^2(\Psi_n - g(\theta_{n,h})) \rightsquigarrow \mathcal{L}$$

under the sequence of alternatives $P_{n,h}$ for all $h \in \mathbb{R}^d$.

- There is no regular estimator of $g(\theta)$ in local regions of $\theta_\star$
- Full argument uses characteristic functions

Outline

1. Impossibility
2. Inference
3. Empirical Application
4. Conclusion

Inference

We focus on the setting

- Full rank second order derivative $\frac{\partial^2 g}{\partial\theta\partial\theta'}(\theta_\star) = H$
- Asymptotically normal estimator $\hat{\theta}$

$$r_n(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, \Sigma)$$

The confidence interval is constructed by inverting the hypothesis $H_0 : g(\theta) = \tau$

A minimum distance test statistic

$$\hat{T}_n(\tau) = \inf_{\theta\in\Theta: g(\theta)=\tau} r_n^2(\hat{\theta} - \theta)'\Sigma^{-1}(\hat{\theta} - \theta)$$

Inference: Two-Dimensional θ and Indefinite H

Consider the null hypothesis

$$H_0 : g(\theta_1, \theta_2) := \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}' \begin{bmatrix} \rho - 1 & \\ & \rho + 1 \end{bmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \tau$$

with $|\rho| < 1$

- The restriction $|\rho| < 1$ guarantees that H is indefinite

For simplicity, let $n = r_n = 1$, and assume

$$\hat{\theta} - \theta \sim N(0, \mathcal{I}_2)$$

The quadratic form g can be viewed as second order approximations

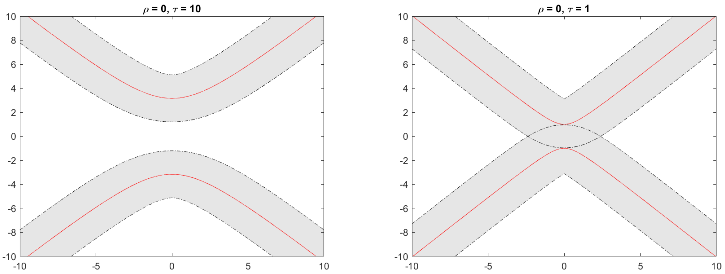
Inference: Two-Dimensional θ and Indefinite H

Let $\mathcal{S}_0$ be the null parameter space

$$\mathcal{S}_0 = \left\{\theta : (1+\rho)\theta_2^2 - (1-\rho)\theta_1^2 = \tau\right\}$$

The acceptance region with critical value c^2 is $\mathcal{S}_0^c$

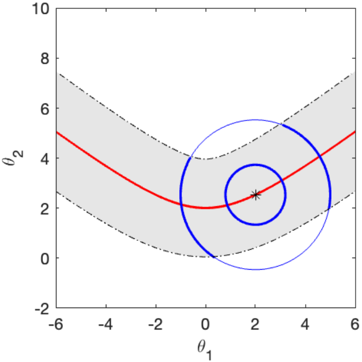
$$\mathcal{S}_0^c = \left\{\theta : \inf_{\vartheta \in \Theta : g(\vartheta) = \tau} (\theta - \vartheta)'(\theta - \vartheta) = d(\theta, \mathcal{S}_0)^2 \leq c^2\right\}$$



Inference: Two-Dimensional θ and Indefinite H

Let $c^2 = \chi^2_{1,1-\alpha}$.

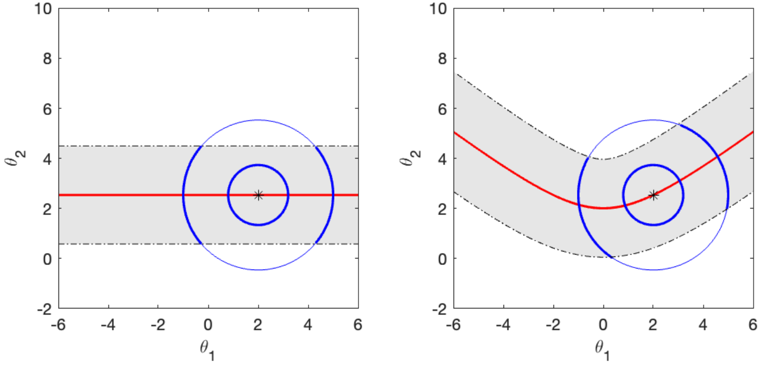
- If $\text{curve}(\mathcal{S}_0) \le \frac{1}{c}$, then $P(\hat{\theta} \in \mathcal{S}_0^c) \ge 1 - \alpha$



Inference: Two-Dimensional θ and Indefinite H

Let $c^2 = \chi^2_{1,1-\alpha}$.

- If $\text{curve}(\mathcal{S}_0) \leq \frac{1}{c}$, then $P(\hat{\theta} \in \mathcal{S}_0^c) \geq 1 - \alpha$



- Idea: show that

$$P(\hat{\theta} \in \mathcal{S}_0^c) \geq P(\hat{\theta} \in \mathcal{S}_{\text{aux}}) = 1 - \alpha$$

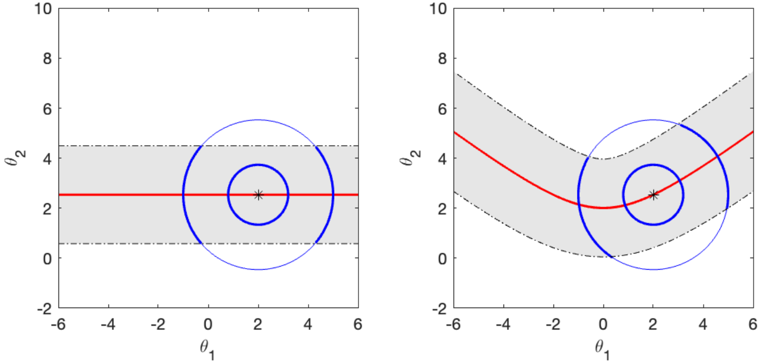
where

$$\mathcal{S}_{\text{aux}} = \left\{ (x_1, x_2) : (x_2 - \theta_2)^2 \leq c^2 \right\}$$

Inference: Two-Dimensional θ and Indefinite H

Let $c^2 = \chi^2_{1,1-\alpha}$.

- If curve$(\mathcal{S}_0) \leq \frac{1}{c}$, then $P(\hat{\theta} \in \mathcal{S}_0^c) \geq 1 - \alpha$



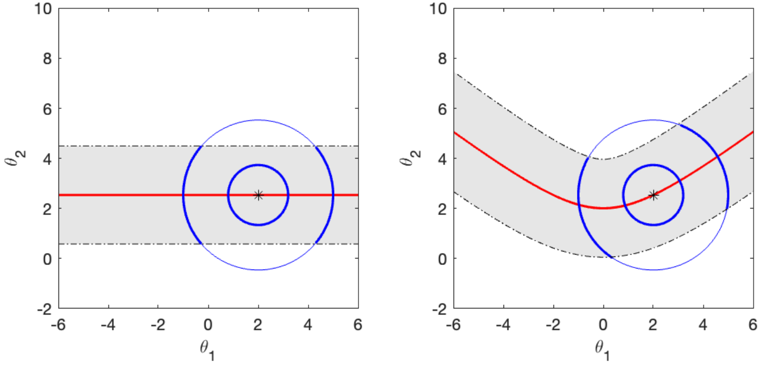
- Switch to polar coordinates $\hat{\theta} = (\theta_1 + r\cos\omega, \theta_2 + r\sin\omega)$

$$P(\hat{\theta} \in \mathcal{S}) = \frac{1}{2\pi} \int_{r=0}^{+\infty} \int_{\omega=0}^{2\pi} 1\left[(\theta_1 + r\cos\omega, \theta_2 + r\sin\omega) \in \mathcal{S}\right] d\omega \; r dr$$

Inference: Two-Dimensional θ and Indefinite H

Let $c^2 = \chi^2_{1,1-\alpha}$.

- If curve$(\mathcal{S}_0) \leq \frac{1}{c}$, then $P(\hat{\theta} \in \mathcal{S}_0^c) \geq 1 - \alpha$

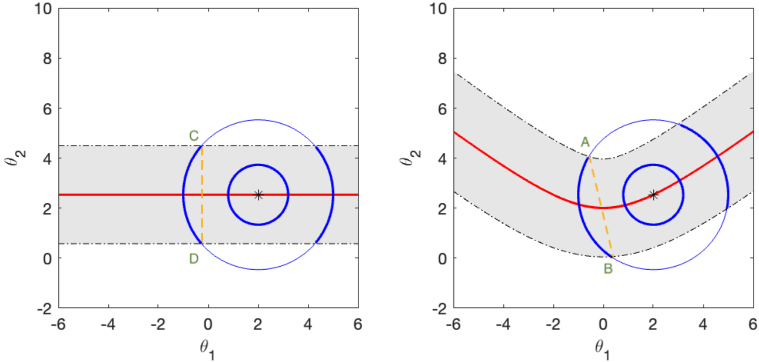


- Switch to polar coordinates $\hat{\theta} = (\theta_1 + r\cos\omega, \theta_2 + r\sin\omega)$
  - $r \leq c$, $\hat{\theta} \in \mathcal{S}$ for all ω

Inference: Two-Dimensional θ and Indefinite H

Let $c^2 = \chi^2_{1,1-\alpha}$.

- If curve$(\mathcal{S}_0) \leq \frac{1}{c}$, then $P(\hat{\theta} \in \mathcal{S}_0^c) \geq 1 - \alpha$

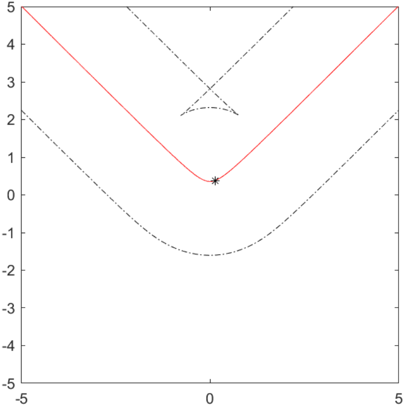


- Switch to polar coordinates $\hat{\theta} = (\theta_1 + r\cos\omega, \theta_2 + r\sin\omega)$
  - $r \leq c$, $\hat{\theta} \in \mathcal{S}$ for all ω
  - $r > c$, $arc(AB) > arc(CD)$

Inference: Two-Dimensional θ and Indefinite H

Let $c^2 = \chi^2_{1,1-\alpha}$.

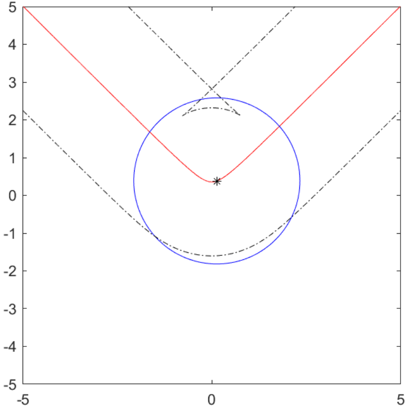
- If $\text{curve}(\mathcal{S}_0) \leq \frac{1}{c}$, then $P(\hat{\theta} \in \mathcal{S}_0^c) \geq 1 - \alpha$
- If $\text{curve}(\mathcal{S}_0) > \frac{1}{c}$, the upper boundary of $\mathcal{S}_0^c$ has a kink



Inference: Two-Dimensional θ and Indefinite H

Let $c^2 = \chi^2_{1,1-\alpha}$.

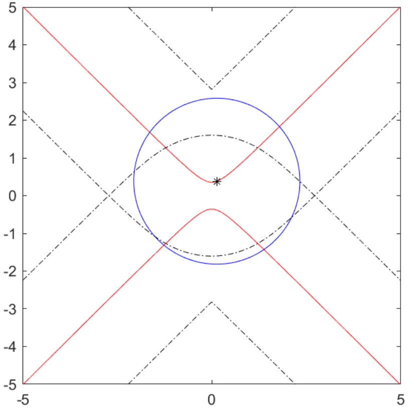
- If $\text{curve}(\mathcal{S}_0) \leq \frac{1}{c}$, then $P(\hat{\theta} \in \mathcal{S}_0^c) \geq 1-\alpha$
- If $\text{curve}(\mathcal{S}_0) > \frac{1}{c}$, the upper boundary of $\mathcal{S}_0^c$ has a kink



Inference: Two-Dimensional θ and Indefinite H

Let $c^2 = \chi^2_{1,1-\alpha}$.

- If $\text{curve}(\mathcal{S}_0) \leq \frac{1}{c}$, then $P(\hat{\theta} \in \mathcal{S}_0^c) \geq 1-\alpha$
- If $\text{curve}(\mathcal{S}_0) > \frac{1}{c}$, if $\rho \geq 0$, then $P(\hat{\theta} \in \mathcal{S}_0^c) \geq 1-\alpha$



Inference: Two-Dimensional θ and Indefinite H

Proposition (Regularity Conditions)

1. *For all* $\theta \in \Theta \backslash \{\theta_\star\}$, $\frac{\partial g(\theta)}{\partial \theta} \neq 0$.
2. *Let* BL_1 *denote the set of Lipchitz functions. There exists* $r_n \to \infty$ *such that*

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} \sup_{f\in BL_1} \left| E_P \left[f \left(\sqrt{r_n} \left(\hat{\theta}_n - \theta_P \right) \right) \right] - E_P \left[f \left(\xi_P \right) \right] \right| = 0,$$

   where $\xi_P \sim N(0, \Sigma_P)$.
3. *Let* $\mathcal{S}$ *denote the set of matrices with eigenvalues bounded below by* $\underline{e} > 0$ *and above by* $\bar{e} \geq \underline{e}$. *For all* $P \in \mathcal{P}$, $\Sigma_P \in \mathcal{S}$.
4. *For all* $\varepsilon > 0$, $\lim_{n\to\infty} \sup_{P\in\mathcal{P}} P\left(\left\| \hat{\Sigma}_n - \Sigma_P \right\| > \varepsilon \right) = 0$.

Inference: Two-Dimensional θ and Indefinite H

Theorem

Suppose $d = 2$. *Let* $(\lambda_{P,1}, \lambda_{P,2})$ *be the eigenvalues of* $\text{sign}\left(g(\theta_P)\right) \Sigma_P^{1/2} H \Sigma_P^{1/2}$. *Let* $\rho_P = \frac{\lambda_{P,1}+\lambda_{P,2}}{|\lambda_{P,1}-\lambda_{P,2}|}$. *Assume that Assumption RC hold. If either*

$$\mathcal{P}_n \subseteq \{P \in \mathcal{P} : \rho_P \in [0, 1-\eta]\} \text{ for some } \eta > 0$$

or

$$\mathcal{P}_n \subseteq \left\{P \in \mathcal{P} : \frac{1}{2} \frac{(1-\rho_P)^2}{1+\rho_P} \frac{|\lambda_{P,1} - \lambda_{P,2}|}{r_n^2 \, |g(\theta_P) - g(\theta_\star)|} \le \frac{1}{c^2}\right\},$$

then

$$\liminf_n \inf_{P \in \mathcal{P}_n} P\left(g(\theta_P) \in CI\right) \ge 1 - \alpha.$$

Inference: General Case

The test statistic is

$$\hat{T}_n = \inf_{\vartheta : g(\vartheta) = g(\theta_n)} r_n^2 \left(\hat{\theta}_n - \vartheta\right)' (\hat{\Sigma}_n)^{-1} \left(\hat{\theta}_n - \vartheta\right)$$

- If $r_n(\theta_n - \theta_\star) = h_n \to h \in \mathbb{R}^d$, a second order approximation gives

$$\hat{T}_n \sim \hat{T}_n^*(h_n) := \inf_{t : t' H t = h_n' H h_n} \left\| \mathbb{Z} - (\hat{\Sigma}_n)^{-1/2} (t - h_n) \right\|^2$$

  where $\mathbb{Z} | \{X_i\} \sim \mathcal{N}(0, \hat{\Sigma}_n)$

Inference: General Case

The test statistic is

$$\hat{T}_n = \inf_{\vartheta: g(\vartheta)=g(\theta_n)} r_n^2 \left(\hat{\theta}_n - \vartheta\right)' (\hat{\Sigma}_n)^{-1} \left(\hat{\theta}_n - \vartheta\right)$$

- If $r_n(\theta_n - \theta_\star) = h_n \to h \in \mathbb{R}^d$, a second order approximation gives

$$\hat{T}_n \sim \hat{T}_n^*(h_n) := \inf_{t: t'Ht = h_n' H h_n} \left\| \mathbb{Z} - (\hat{\Sigma}_n)^{-1/2}(t - h_n) \right\|^2$$

  where $\mathbb{Z} | \{X_i\} \sim \mathcal{N}(0, \hat{\Sigma}_n)$

- If $||h_n|| \to \infty$, the restriction $t'Ht = h_n' H h_n$ is approximately linear

$$\frac{h_n' H}{||h_n' H||}(t - h_n) = -\frac{(t - h_n)' H (t - h_n)}{||h_n' H||} = o(1),$$

  Both $\hat{T}_n$ and $\hat{T}_n^*(h_n)$ are approximated χ_1^2

Given h_n, we can easily get the quantile of $\hat{T}_n^*$ by simulation. However, h_n is a nuisance parameter that cannot be consistently estimated.

Inference: General Case

Two step feasible critical value.

1. Construct a $(1-\eta)$ confidence set for h_n,

$$\mathcal{H} = \hat{\theta}_n + r_n^{-1}(\hat{\Sigma}_n)^{1/2}\mathcal{H}_z$$

   where set $\mathcal{H}_z$ satisfies $P\left(N(0, I_d) \in \mathcal{H}_z\right) = 1-\eta$.

2. Construct the critical value based on dist. of $\hat{T}_n^*$ conditional on the first step,

$$\hat{c} = \sup_{h \in \mathcal{H}} Q\left(\hat{T}_n^*(h) \middle| \mathbb{Z} \in \mathcal{H}_z; \frac{1-\alpha}{1-\eta} \right)$$

   $\hat{c}$ is less conservative than simple Bonferroni correction.

Inference: General Case

Theorem

Under Assumption RC, it holds that

$$\liminf_n \inf_{P \in \mathcal{P}} P\left(\hat{T}_n \leq \hat{c}\right) \geq 1 - \alpha.$$

In addition, if $\|r_n(\theta_{P_n} - \theta_\star)\| \to \infty$,

$$\lim_n P_n\left(\hat{T}_n \leq \hat{c}\right) \in \left[1 - \alpha, 1 - \alpha + \eta\right).$$

Outline

Empirical Application

Alan et al. (2018) study how teachers' gender role attitudes influence student outcomes

- Treatment: whether a teacher is identified as holding traditional rather than progressive gender beliefs
- Mediator: the student's own gender role beliefs
- Outcome: the student's verbal test scores

Empirical Application

Exposure	$\hat{\theta}_1$	$t(\hat{\theta}_1)$	$\hat{\theta}_2$	$t(\hat{\theta}_2)$	$\hat{\theta}_1 \cdot \hat{\theta}_2$	n
Full sample	0.199	3.140	-0.119	-5.343	-0.024	1885
1 year	0.256	2.052	-0.097	-1.941	-0.025	499
2–3 years	0.109	1.065	-0.125	-4.163	-0.014	906
4 years	0.064	0.513	-0.113	-1.931	-0.007	468

Table: Estimates of Mediation Effects by Teacher Exposure

- Students can be grouped according to the length of their exposure to a given teacher: at most one year, two to three years, or up to four years.

Empirical Application

Exposure	Full	1-Year	2-3 Year	4 Year
Point Estimate	-0.024	-0.025	-0.014	-0.007
←Interval Length→	←0.032→	←0.070→	←0.053→	←0.070→
95% BN1 CI	$[-0.042, -0.010]$	$[-0.071, -0.001]$	$[-0.042, 0.010]$	$[-0.045, 0.025]$
	←0.034→	←0.076→	←0.058→	←0.076→
95% BN2 CI	$[-0.044, -0.010]$	$[-0.075, 0.001]$	$[-0.046, 0.012]$	$[-0.049, 0.027]$
	←0.038→	←0.086→	←0.068→	←0.094→
95% AM CI	$[-0.046, -0.008]$	$[-0.083, 0.003]$	$[-0.052, 0.016]$	$[-0.059, 0.035]$
	←0.042→	←0.092→	←0.070→	←0.096→
95% Projection CI	$[-0.048, -0.006]$	$[-0.085, 0.007]$	$[-0.052, 0.018]$	$[-0.059, 0.037]$

Table 1: Mediation Effect Point Estimates, 95% Confidence Intervals

Outline

Conclusion

- We examine inference in local regions of first-order degeneracy
- We show that regular estimation is infeasible
- We develop inference procedures based on minimum-distance statistics