Inference Under First Order Degeneracy

Xinyue Bei [1] Manu Navjeevan [2]

[1]University of Texas, Austin

[2]Texas A&M University

April, 2026

Table of Contents

Introduction

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

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

Object of interest is $g(\theta) : \Theta \to \mathbb{R}$. We are interested in inference on $g(\theta)$ near a point $\theta_\star$ such that $\nabla g(\theta_\star) = 0$.

Examples: Mediation Analysis

Example 1 (Mediation Analysis)

Consider a causal mediation analysis with parameter $\theta = (\theta_1, \theta_2)'$, where θ_1 represents the effect of a treatment on a mediator and θ_2 represents the effect of the mediator on the outcome. The indirect effect is

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

At $\theta_\star = (0, 0)'$, $\nabla g(\theta_\star) = 0$, which complicates inference on $g(\theta)$ in local regions of $\theta_\star$.

Recent works have proposed tests for the specific null-alternate pair $H_0 : g(\theta) = 0$ against $H_1 : g(\theta) \neq 0$ (Garderen and Giersbergen, 2024, Hillier et al., 2024). However, these works do not consider the more general problem of constructing confidence intervals in local regions of the origin.

Examples: Impulse Response Functions

Example 2 (Impulse Response Functions)

Consider an autoregressive AR(1) model

$$y_t = \theta y_{t-1} + u_t$$

where $\theta \in \mathbb{R}$ and u_t is white noise. The impulse response function at horizon h is

$$g(\theta) = \theta^h$$

which measures the impact at time period h of an initial shock.

At $\theta_\star = 0$: $\frac{\partial}{\partial\theta} g(\theta_\star) = h\theta_\star^{h-1} = 0$ for $h \geq 2$.

Existing literature has focused on inference when θ is close to one (the unit-root problem). However, inference on the impulse response is also complicated when θ is close to zero (Benkwitz et al., 2000, Inoue and Kilian, 2002, Mikusheva, 2012).

Examples: Weak IV Bias and Size Distortion

Example 3 (Weak IV Bias and Size Distortion)

Consider a standard homoskedastic linear IV model,

$$y_i = x_i\beta + \varepsilon_i, \qquad x_i = z_i'\theta + v_i$$

where $Z = (z_1', \ldots, z_n')' \in \mathbb{R}^{n \times d_z}$ is treated as fixed. Size distortion of Wald tests and 2SLS bias for β governed by the concentration parameter

$$g(\theta) = \theta'(Z'Z)\theta/\sigma_v^2$$

At $\theta_\star = 0$: $\nabla g(\theta_\star) = 2(Z'Z)\theta_\star/\sigma_v^2 = 0$.

Stock and Yogo (2005) and Ganics et al. (2021) make inferences about the concentration parameter using the F-statistic, a scaled version of $g(\hat{\theta})$. Analyses complicated by the fact that the limiting distribution of $g(\hat{\theta})$ is non-standard when θ is close to zero.

Examples: Breakdown Point Analysis

Example 4 (Breakdown Point Analysis)

Consider a missing data setup: observe $\{Y_i D_i, D_i, X_i\}_{i=1}^n$, where $D_i \in \{0, 1\}$ indicates whether Y_i is observed and $X_i \in \{x_1, \ldots, x_K\}$ is a discrete covariate.

The squared Hellinger distance between the distributions of X_i conditional on D_i can be written as

$$g(\theta) = H^2(P_0, P_1) = \frac{1}{2} \sum_{k=1}^{K} \left(\sqrt{\theta_{0,k}} - \sqrt{\theta_{1,k}} \right)^2$$

where $\theta_{d,k} = \Pr(X = x_k \mid D = d)$.

At $\theta_\star$ such that $\theta_{0,k} = \theta_{1,k}$ for all k (i.e., data is missing completely at random), $\nabla g(\theta_\star) = 0$ and $g(\theta_\star) = 0$ (Ober-Reynolds, 2024).

Problems with Delta-Method

If we know $\theta \neq \theta_\star$, first-order delta method tells us

$$r_n(g(\hat{\theta}) - g(\theta)) \rightsquigarrow \nabla g(\theta)'\mathbf{W}$$

If we know that $\theta = \theta_\star$, second order delta method tells us

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

Problem: Do not know the true value of θ. Asymptotic approximations to distribution of $g(\hat{\theta})$ seem to change abruptly at $\theta_\star$.

- Problem has been noted in literature (e.g Miller et al. (2024), Garderen and Giersbergen (2024), Dufour et al. (2025)) but no formal framework proposed to study this issue.

Problems with Delta-Method

Model the uncertainty by supposing that $\theta = \theta_\star + h/r_n$. Under this asymptotic regime

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

Notice that the limiting distribution now depends on the local parameter h which is not consistently estimable. This formally captures the intuition from before that the researcher cannot know which approximation to use.

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 $\theta_\star$ — the behavior of any estimator must depend on nuisance parameters that cannot be consistently estimated.

2. Given this, are there alternative inference procedures that may still perform well?
   - Answer: Yes. We show that minimum distance methods deliver uniformly valid CIs. Establish sufficient conditions under which standard χ^2 critical values can be applied and propose computationally simple bootstrap procedures when these conditions are not met.

Table of Contents

Impossibility: Setup

Begin by assuming that the researcher observes data $X^{(n)} = (X_1, \ldots, X_n)$ from a parametric model $\mathcal{P}_{n,\theta}$, $\theta \in \Theta$, that is locally asymptotically normal (Le Cam, 1960).

Definition 1 (Local Asymptotic Normality)

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

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

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

Note: This is a slight generalization of the standard concept of a smooth parametric model to accommodate settings where the rate of convergence may be slower than $\sqrt{n}$. Results apply to semi/nonparametric models that contain locally asymptotically normal submodels.

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 can depend on local parameter h.

- The convergence rate is r_n^2 instead of r_n because $g(\theta)$ is "flat" around $\theta_\star$. Tests for $g(\theta)$ based on estimators whose convergence rates are slower than r_n^2 when θ is close to $\theta_\star$ have trivial power against local alternatives of the form $\theta_\star + h/r_n$.

Regularity and Quantile Unbiasedness

Definition 2 (Regularity)

An estimator Ψ_n is **regular** if its limiting distribution does not depend on h: there exists a distribution $\mathcal{L}$ on $\mathbb{R}$ such that $\mathcal{L}_h = \mathcal{L}$ for all $h \in \mathbb{R}^d$.

- Without regular estimators, standard Wald-type inference — comparing a test statistic to a fixed critical value — will not have correct asymptotic size.

Definition 3 (Quantile Unbiasedness)

An estimator Ψ_n is **locally asymptotically α-quantile unbiased** if $\mathcal{L}_h\{(-\infty, 0]\} = \alpha$ for all $h \in \mathbb{R}^d$.

- Any asymptotically similar one-sided CI $(-\infty, \hat{c}]$ can be converted into a quantile-unbiased estimator by setting $\Psi_n = \hat{c}$. Ruling out quantile unbiasedness therefore rules out similar CIs.

We will show that **neither property** can hold in local regions of $\theta_\star$.

Limit Experiment

Proposition 1 (Limit Experiment)

Let Ψ_n be a sequence of estimators as described above. Then, theere 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 $U \sim \text{Unif}(0, 1)$ independent of Z, such that

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

Proposition 1 establishes equivalence between estimating $g(\theta)$ in local regions of $\theta_\star$ and 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 behvavior does not depend on h, that is $\mathcal{L}_h = \mathcal{L}$ for all $h \in \mathbb{R}^d$.

- By Proposition 1, there must be an estimator $\Psi(Z, U)$ in Gaussian shift experiment such that

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

  for all $h \in \mathbb{R}^d$.

- To build intuition suppose that $\mathcal{L}$ has a finite second moment. WLOG, $\Psi(Z, U)$ is an unbiased estimator of $\frac{1}{2} h' \nabla^2 g(\theta_\star) h$.

- However, Cramér-Rao lower bound tells us that for any unbiased estimator

$$\mathrm{Var}_h(\Psi(Z, U)) \geq 4 h' (\nabla^2 g(\theta_\star)) \Gamma_{\theta_\star}^{-1} (\nabla^2 g(\theta_\star)) h$$

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

Impossibility: Formal Result

Full argument uses characteristic functions and considers quantile unbiased estimation.

Theorem 1 (Impossibility of Estimation)

1. *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* $h \in \mathbb{R}^d$.

2. *Let* $\{\mathcal{L}_h\}$ *be a family of distributions on* $\mathbb{R}$ *such that (i)* $\mathcal{L}_h\{(-\infty, 0]\} = \alpha$ *for some* $\alpha \in (0, 1)$, *and (ii) each* $\mathcal{L}_h$ *has a CDF differentiable at zero with derivatives at zero bounded below by* $\epsilon > 0$. *There is no estimator sequence* Ψ_n *such that*

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

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

Impossibility: Discussion

1. The first part of Theorem 1 establishes that there is no regular estimator of $g(\theta)$ in local regions of $\theta_\star$ — any estimator must have a limiting distribution that depends on nuisance parameters that cannot be consistently estimated.
   - Implication for Inference: Standard Wald-type approaches to inference will not work.
2. The second part of Theorem 1 additionally rules out the existence of asymptotically similar one-sided confidence intervals — there is no alternate approach to inference can that can deliver similar confidence intervals.
   - Any asymptotically similar one-sided CI of the form $(-\infty, \hat{c}]$ can be transformed into a quantile-unbiased estimator by setting $\Psi_n = \hat{c}$.
   - As in Hirano and Porter (2012), this result requires some conditions on the family of distributions $\{\mathcal{L}_h\}$. We discuss these conditions in the paper.

Impossibility for Hypothesis Testing

Consider testing $H_0 : g(\theta) = g(\theta_\star)$. Define the set of local parameters under which the null holds asymptotically:

$$\mathcal{H}_\star = \{h \in \mathbb{R}^d : h'\nabla^2 g(\theta_\star)h = 0\}$$

Asymptotically similar test: A test $\Xi_n : \mathcal{X}^n \to \{0, 1\}$ satisfying

$$\limsup_{n\to\infty} P_{\theta_{n,h}}(\Xi_n = 1) = \alpha \quad \text{for all } h \in \mathcal{H}_\star$$

Local asymptotic power curve:

$$\mathcal{P}(h) = \limsup_{n\to\infty} P_{\theta_{n,h}}(\Xi_n = 1)$$

Unlike quantile unbiased estimation (which is ruled out by Theorem 1), asymptotically similar tests can exist — one can always construct a similar test by flipping a weighted coin. The question is whether such tests can have non-trivial power.

Flat Power at Degeneracy

Proposition 2 (Flat Power at Degeneracy)

Let Ξ_n be an asymptotically similar test such that $\mathcal{P}(h)$ is differentiable at $h = 0$. Then, the directional derivative of the local asymptotic power curve, in directions $h \in \mathcal{H}_\star$, is equal to zero:

$$D_h \mathcal{P}(0) = 0 \quad \textit{for all } h \in \mathcal{H}_\star$$

In particular, if $\nabla^2 g(\theta_\star)$ is indefinite then $\mathcal{H}_\star$ spans $\mathbb{R}^d$ and $\nabla \mathcal{P}(0) = 0$.

Interpretation: When $\nabla^2 g(\theta_\star)$ is indefinite, the power of any asymptotically similar test cannot increase at a linear rate in any direction away from $\theta_\star$. This is a strong restriction on the local power of hypothesis tests near points of degeneracy.

Hypothesis Testing: Examples

1. **Mediation (cont.):** $g(\theta) = \theta_1\theta_2$, $\nabla^2 g(\theta_\star) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is indefinite.
   - $\mathcal{H}_\star = \{h \in \mathbb{R}^2 : h_1 h_2 = 0\}$ is the union of the two coordinate axes, which spans $\mathbb{R}^2$.
   - Therefore, $\nabla\mathcal{P}(0) = 0$ for any asymptotically similar test.
2. **Squared mean (contrast):** $g(\theta) = \theta_1^2 + \theta_2^2$, $\nabla^2 g(\theta_\star) = 2I$ is positive definite.
   - $\mathcal{H}_\star = \{0\}$, so Proposition 2 does not bind.
   - Powerful similar tests can be constructed in this case (Chen and Fang, 2019).
3. **Standard inference (contrast):** Univariate θ, $g'(\theta_0) \neq 0$. Here $\mathcal{H}_\star = \{0\}$ and the power curve has strictly positive derivative at zero — standard inference works.

Extension to Infinite-Dimensional Models

Corollary 1 (Impossibility in Infinite-Dimensional Models)

Suppose the data are generated from a sequence of distributions $\{P_{0,n}\} \in \mathcal{P}$. Let $\mathcal{P}_f \subset \mathcal{P}$ be a regular parametric submodel passing through $\{P_{0,n}\}$, and suppose that g_f satisfies the differentiability assumption with $\{P_{n,\theta_\star}\} = \{P_{0,n}\}$. Then:

1. *There is no estimator sequence Ψ_n and law $\mathcal{L}$ such that $r_n^2(\Psi_n - g_f(\theta_\star + h/r_n)) \rightsquigarrow \mathcal{L}$ for all $h \in \mathbb{R}^{d_f}$, along every regular parametric submodel $\mathcal{P}_f$.*
2. *The quantile-unbiasedness impossibility likewise holds along every regular submodel.*

The impossibility results are not an artifact of the parametric assumption — they apply to any semiparametric or nonparametric model that contains a parametric submodel satisfying local asymptotic normality.

Table of Contents

Minimum Distance Inference: Overview

Given the impossibility results, standard Wald-type inference fails in local regions of $\theta_\star$. We take a different approach.

Strategy: Construct confidence intervals by inverting the hypothesis $H_0 : g(\theta) = \tau$ using a minimum distance (MD) test statistic:

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

i.e., the squared distance from $\hat{\theta}$ to the nearest point on the null manifold $\{\theta : g(\theta) = \tau\}$.

Key question: What critical value should we compare $\hat{T}_n(\tau)$ to?

- If the null manifold were a **hyperplane**, the answer is $Q(\chi_1^2, 1 - \alpha)$.
- Near $\theta_\star$, the null manifold is **curved** — standard results do not directly apply.

Our main result: The χ_1^2 critical value remains valid under two geometric conditions. When these fail, we propose a simple bootstrap procedure.

The Null Manifold Near Degeneracy

When the Hessian $H = \nabla^2 g(\theta_\star)$ is indefinite and $d = 2$, the null manifold $\{\theta : g(\theta) = \tau\}$ near $\theta_\star$ is approximately a **hyperbola** with two branches:

$$\mathcal{S}_0(\tau) = \mathcal{S}_0^+(\tau) \cup \mathcal{S}_0^-(\tau)$$

The MD test accepts $H_0 : g(\theta) = \tau$ when $\hat{\theta}$ falls within distance c of some point on $\mathcal{S}_0(\tau)$.

Two features of this geometry matter for coverage:

1. **Curvature** of each branch — how much it bends away from a hyperplane.
2. **Proximity** of the two branches — how close they are to each other.

When the curvature is small or the branches are close, the χ^2_1 critical value $c = \sqrt{Q(\chi^2_1, 1-\alpha)}$ provides valid coverage.

When Does χ_1^2 Work? Two Sufficient Conditions

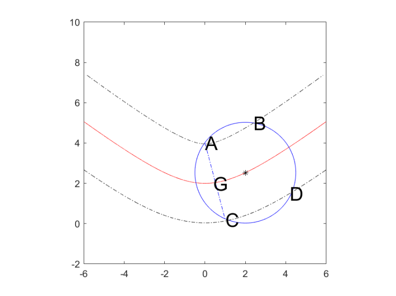
The standard χ_1^2 critical value delivers uniformly valid coverage if **either**:

1. **Low curvature:** The null manifold is close to linear near the point of interest — the acceptance region behaves like a band around a hyperplane.
2. **Close branches:** The two branches of the null manifold are sufficiently close to each other — even when one branch is highly curved, proximity to the other branch ensures coverage.

The formal conditions are given in the appendix. The key geometric intuition is illustrated on the next two slides.

Why this matters: Existing approaches (Andrews and Mikusheva, 2016) rely only on curvature, and become overly conservative when curvature is large. Our approach additionally exploits branch proximity.

Geometric Intuition: Low Curvature



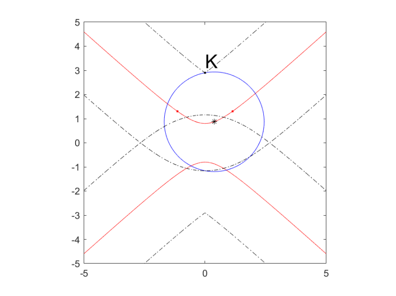
Red curves: null manifold $\mathcal{S}_0(\tau)$. Black dashes: boundary of acceptance region. Blue circle: $\partial B(\theta, r)$ for $\theta \in \mathcal{S}_0^+(\tau)$.

When curvature is small:

- The acceptance region looks like a band around a nearly-linear curve.
- Each circle centered at θ has at least as much arc inside the acceptance region as it would if the null were exactly linear.

Coverage is at least $1 - \alpha$.

Geometric Intuition: High Curvature



When curvature is large, the null manifold develops a kink (K) and the low-curvature argument fails.

However, when the two branches are close:

- Points on $\mathcal{S}_0^+(\tau)$ are near $\mathcal{S}_0^-(\tau)$.
- Circles centered on one branch are partially covered by the other branch's acceptance region.

Key insight: Our method exploits the proximity of multiple branches, unlike curvature-only approaches.

Application: Mediation Analysis

For the mediation effect $g(\theta) = \theta_1 \theta_2$:

- The "close branches" condition holds whenever the correlation between $\hat{\theta}_1$ and $\hat{\theta}_2$ is non-negative.
- In this case, the χ^2_1 critical value is uniformly valid.

The rejection region for testing $\theta_1 \theta_2 = 0$ simplifies to

$$\min\{|\hat{\theta}_1|, |\hat{\theta}_2|\} > \sqrt{Q(\chi^2_1, 1-\alpha)}$$

Contrast with Andrews and Mikusheva (2016): Their critical value is based on the maximal curvature of the null manifold. As $\tau \to 0$, curvature grows without bound and their critical value approaches $Q(\chi^2_2, 1-\alpha)$. Our critical value remains $Q(\chi^2_1, 1-\alpha)$ because we exploit branch proximity — producing 5–18% tighter intervals.

Bootstrap Procedure for General Cases

When $d > 2$ or the sufficient conditions for χ^2_1 are not met, we propose a computationally simple **two-step bootstrap**:

1. **First step:** Construct a confidence set for the local parameter $h_n = r_n(\theta - \theta_\star)$, which governs the shape of the null manifold near $\theta_\star$.
2. **Second step:** For each h in this confidence set, simulate the distribution of the MD test statistic. Take the most conservative critical value.

Properties:

- Uniformly valid: coverage is at least $1 - \alpha$ over all DGPs.
- Less conservative than Bonferroni — the second step conditions on the first.
- When θ is far from $\theta_\star$, the bootstrap critical value converges to the standard χ^2_1 critical value automatically.
- Tuning parameter: $\eta = \alpha/10$ in practice.

Formal details in appendix.

Table of Contents

Simulation Design

Setting: Confidence intervals for the mediation effect $g(\theta) = \theta_1\theta_2$.

DGP: Estimators are simulated from

$$\hat{\theta} - \theta \sim \mathcal{N}\left(0, \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}\right)$$

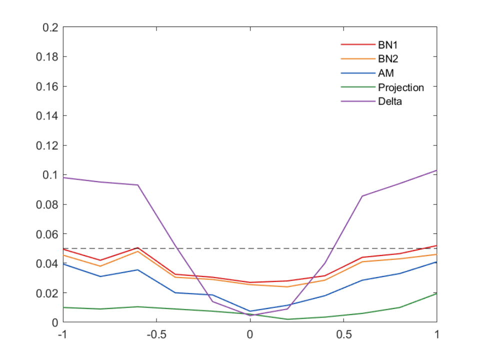
with $r \in \{0, 0.5\}$, $\theta_2 \in \{2, 6\}$, $\theta_1/\theta_2 \in [-1, 1]$.

Methods compared:

- **BN1:** MD test with χ^2_1 critical value
- **BN2:** MD test with bootstrapped critical value ($\eta = \alpha/10$)
- **AM:** Andrews and Mikusheva (2016) geometric approach
- **Projection:** MD test with χ^2_2 critical value
- **Naive Wald** and **Naive Bootstrap**: Standard delta method approaches

Nominal level $\alpha = 0.05$, $S = 2{,}000$ replications.

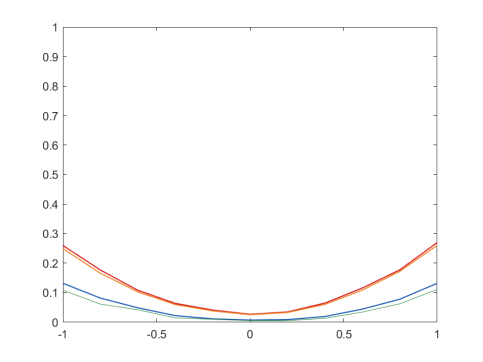
Simulation: Size Control ($r = 0$)



Rejection rate at true value $g(\theta)$, $\theta_2 = 2$, $r = 0$.

- Naive Wald (Delta): overrejects ($\sim 10\%$) away from origin.
- BN1 (red): near nominal level everywhere.
- AM, Projection: conservative.
- BN1 controls size even when sufficient conditions are not met.

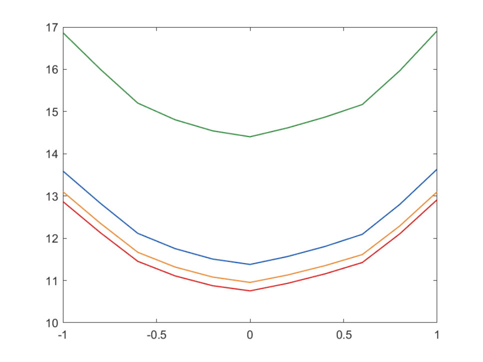
Simulation: Power ($r = 0$)



Probability CI excludes zero, $\theta_2 = 2$, $r = 0$.

- BN1/BN2 have substantially higher power than AM near the origin.
- AM performance close to that of the projection method.
- Power differences most pronounced when θ is near degeneracy.

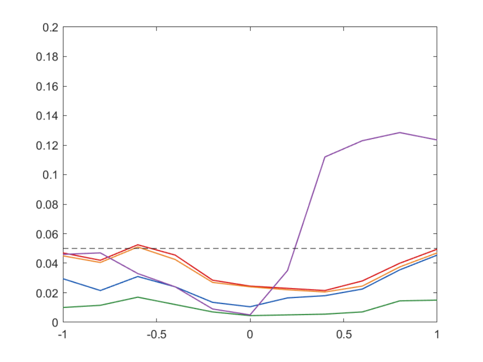
Simulation: Confidence Interval Length ($r = 0$)



Median CI length, $\theta_2 = 2$, $r = 0$.

- BN1 (red) consistently produces the shortest intervals.
- BN2 (orange) close behind.
- AM (blue): 5–18% longer than BN1.
- Projection (green): 19–30% longer than BN1.
- Differences most pronounced near the origin.

Simulation: Correlated Estimates ($r = 0.5$)



Rejection rate at true value, $\theta_2 = 2$, $r = 0.5$.

- When $r = 0.5$, the "close branches" condition fails for some parameter values ($\text{sign}(g(\theta_P)) \cdot r < 0$).
- Nevertheless, BN1 maintains correct size across all designs.
- The sufficient conditions for χ_1^2 are **not necessary**.
- Same qualitative patterns: BN1/BN2 dominate AM and Projection.

Table of Contents

Empirical Application: Background

We illustrate the empirical relevance of our results using the setting analyzed by Alan et al. (2018).

- **Setting:** Turkish education system — elementary school teachers are randomly allocated across schools.
- **Treatment:** Whether a teacher holds traditional (vs. progressive) gender role beliefs.
- **Mediator:** Student's own gender role beliefs.
- **Outcome:** Verbal test scores.
- **Sample:** ~4,000 third- and fourth-grade students taught by 145 teachers.
- Students grouped by exposure to a given teacher: full sample, ≤ 1 year, 2–3 years, 4 years.

Random assignment of teachers generates plausibly exogenous variation. Following Garderen and Giersbergen (2024), the identifying assumptions for causal mediation analysis are satisfied.

Estimates of Mediation Effects

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

- $\hat{\theta}_1$: effect of teacher attitudes on student beliefs.
- $\hat{\theta}_2$: effect of student beliefs on test scores.
- $\hat{\theta}_1 \cdot \hat{\theta}_2$: indirect (mediated) effect — small and negative across all exposure groups.
- The indirect effect appears close to zero, suggesting that degeneracy may be a concern for inference.

Confidence Intervals for Mediation Effect

	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]

- BN1 produces the tightest intervals in all subsamples.
- Garderen and Giersbergen (2024) show the correlation between $\hat{\theta}_1$ and $\hat{\theta}_2$ is zero, so both BN methods are uniformly valid.

Empirical Application: Discussion

Key finding: For the 1-year exposure subsample, the BN1 confidence interval $[-0.071, -0.001]$ excludes zero, while all other methods do not.

- BN1 supports the conclusion of Garderen and Giersbergen (2024) that the mediation effect of a one-year exposure to a teacher with traditional views is negative.
- AM and Projection cannot reject a null of zero effect at the five-percent level.
- The difference is not only theoretical but **empirically relevant**: different methods lead to different inferential conclusions.

The AM intervals lie strictly inside the Projection intervals — as the true mediation effect approaches zero, the AM critical value converges toward $Q(\chi^2_2, 1-\alpha)$, which accounts for the close similarity between the two sets of intervals.

Conclusion

1. Developed a formal asymptotic framework for studying inference in local regions of **first-order degeneracy**, where $\nabla g(\theta_\star) = 0$.
2. **Impossibility results:** No regular estimator of $g(\theta)$ exists. No asymptotically similar one-sided CIs. Similar hypothesis tests must have flat power at the point of degeneracy.
3. **Minimum distance inference:** Standard χ_1^2 critical value is uniformly valid when null curve curvature is small or branches are close. Bootstrap procedure for general cases.
4. **Simulations:** BN1/BN2 control size uniformly. Substantially higher power and shorter intervals than AM and Projection methods near degeneracy.
5. **Empirical application:** Tighter confidence intervals for the mediation effect of teacher gender attitudes on student outcomes. Different inferential conclusions from existing methods.

References

1 S. Alan, S. Ertac, and I. Mumcu, "Gender stereotypes in the classroom and effects on achievement", The Review of Economics and Statistics **100**, 876–890 (2018).

2 I. Andrews and A. Mikusheva, "A geometric approach to nonlinear econometric models", Econometrica **84**, 1249–1264 (2016).

3 A. Benkwitz, M. H. Neumann, and H. Lütekpohl, "Problems related to confidence intervals for impulse responses of autoregressive processes", Econometric Reviews **19**, 69–103 (2000).

4 Q. Chen and Z. Fang, "Improved inference on the rank of a matrix", Quantitative Economics **10**, 1787–1824 (2019).

5 J.-M. Dufour, E. Renault, and V. Zinde-Walsh, "Wald tests when restrictions are locally singular", The Annals of Statistics **53**, 457–476 (2025).

6 G. Ganics, A. Inoue, and B. R. and, "Confidence intervals for bias and size distortion in iv and local projections-iv models", Journal of Business & Economic Statistics **39**, 307–324 (2021).

7 K. J. van Garderen and N. P. van Giersbergen, "A nearly similar powerful test for mediation", Review of Economics and Statistics, 1–26 (2024).

References

8 G. Hillier, K. J. van Garderen, and N. van Giersbergen, "Improved tests for mediation", arXiv preprint arXiv:2403.02144 (2024).

9 K. Hirano and J. R. Porter, "Impossibility results for nondifferentiable functionals", Econometrica **80**, 1769–1790 (2012).

10 A. Inoue and L. Kilian, "Bootstrapping smooth functions of slope parameters and innovation variances in var(∞) models", International Economic Review **43**, 309–331 (2002).

11 L. Le Cam, *Locally asymptotically normal families of distributions; certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses*, tech. rep. 2 (University of California, Berkeley, Department of Statistics, 1960), pp. 37–98.

12 A. Mikusheva, "One-dimensional inference in autoregressive models with the potential presence of a unit root", Econometrica **80**, 173–212 (2012).

13 D. L. Miller, F. Molinari, and J. Stoye, "Testing sign congruence between two parameters", arXiv preprint arXiv:2405.11759 (2024).

14 D. Ober-Reynolds, *Robustness to missing data: breakdown point analysis*, 2024.

References

[15] J. Stock and M. Yogo, *Testing for weak instruments in linear iv regression*, (Cambridge University Press, New York, 2005), pp. 80–108.

Appendix: Technical Details

Formal Result: χ_1^2 Critical Value ($d = 2$)

Proposition 3 (χ_1^2 Critical Value)

Let $c = \sqrt{Q(\chi_1^2, 1-\alpha)}$ *and* $\hat{\theta} - \theta \sim N(0, I_2)$*. Suppose either*

1. $\dfrac{1-\rho}{\sqrt{\tau(1+\rho)}} \le \dfrac{1}{c}$ *(curvature of* $\mathcal{S}_0^+(\tau)$ *is sufficiently small), or*
2. $\rho \ge 0$ *(the two branches* $\mathcal{S}_0^+(\tau)$ *and* $\mathcal{S}_0^-(\tau)$ *are sufficiently close).*

Then for all $\theta \in \mathcal{S}_0(\tau)$*:*

$$P\left(\hat{\theta} \in \mathcal{S}(\tau, c)\right) \ge 1 - \alpha$$

Formal Result: Asymptotic Validity ($d = 2$)

Theorem 2 (Two-Dimensional Case)

Suppose $d = 2$ and let $c = \sqrt{Q(\chi_1^2, 1-\alpha)}$. Let $(\lambda_{P,1}, \lambda_{P,2})$ be the eigenvalues of $\text{sign}(g(\theta_P) - g(\theta_\star))\, \Sigma_P^{1/2} H \Sigma_P^{1/2}$, and define

$$\rho_P = \frac{\lambda_{P,1} + \lambda_{P,2}}{|\lambda_{P,1} - \lambda_{P,2}|}$$

Under regularity conditions on $\hat{\theta}$, if for some $\eta > 0$ either

1. $\dfrac{(1-\rho_P)\sqrt{|\lambda_{P,1} - \lambda_{P,2}|}}{2r_n\sqrt{|g(\theta_P) - g(\theta_\star)|(1+\rho_P)}} \leq \dfrac{1}{c}, \quad \rho_P \in [\eta - 1, 1 - \eta], \quad$ *or*
2. $\rho_P \in [0, 1-\eta]$,

then $\liminf_n \inf_{P \in \mathcal{P}_n} P\left(\hat{T}_n(g(\theta_P)) \leq c^2\right) \geq 1 - \alpha$.

Formal Result: Bootstrap Procedure

Local approximation: Given $\theta_n = \theta_\star + h_n/r_n$, approximate $\hat{T}_n$ by

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

Two-step critical value:

$$\mathcal{H} = r_n(\hat{\theta} - \theta_\star) - \hat{\Sigma}^{1/2}\mathcal{H}_z, \qquad \hat{c} = \sup_{h \in \mathcal{H}} Q\left(\hat{T}_n^*(h) \,\middle|\, \mathbb{Z} \in \mathcal{H}_z;\ \frac{1-\alpha}{1-\eta} \right)$$

Theorem 3 (Uniform Validity of Bootstrap)

Under regularity conditions on g and $\hat{\theta}$, it holds that

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

In addition, if $\|r_n(\theta_{P_n} - \theta_\star)\| \to \infty$: $\lim_n P_n\left(\hat{T}_n(g(\theta_{P_n})) \leq \hat{c} \right) \in [1-\alpha,\ 1-\alpha+\eta)$.