Inference under "Near-Failure" of the Delta Method*

Xinyue Bei[1] Manu Navjeevan[2]

[1]The University of Texas at Austin
[2]Texas A&M University

September 21, 2025

Abstract

We study inference in settings where the delta method nearly fails, that is, when the first derivative of a transformation is zero or close to zero so that first-order approximations alone are uninformative and second-order terms must also be taken into account. A leading example is causal mediation analysis, where the indirect effect is a product of coefficients and the gradient becomes small near the origin. In these local regions of degeneracy, plug-in estimators fail to admit regular asymptotic distributions as their limiting behavior depends on nuisance parameters that are not consistently estimable. We formally show that this failure is intrinsic — in local regions of first-order degeneracy, regular estimation is unattainable and α-quantile-unbiased procedures are sharply constrained, echoing related findings for nondifferentiable functionals in Hirano and Porter (2012). Despite these restrictions, we develop minimum-distance based inference procedures that deliver uniformly valid confidence intervals and demonstrate favorable power both in simulations and in an empirical application linking teacher gender attitudes to student outcomes.

Keywords: Delta Method, Higher Order Asymptotics, Impossibility, Minimum Distance Inference

JEL Codes: C12, C13, C18

1 Introduction

The purpose of this document is to consider inference under what we term "Near Failure" of the Delta Method. To be more formal, consider a parameter $\theta \in \Theta \subseteq \mathbb{R}^k$ and a twice continuously differentiable function $g : \Theta \to \mathbb{R}^d$. We are interested in inference, e.g constructing asymptotically valid confidence intervals, in local neighborhoods of a point $\theta_\star$ for which $\nabla g(\theta_\star) = 0$. To capture this phenomena, suppose that we have an estimator of θ, $\hat{\theta}$, that satisfies, uniformly over $\theta \in \Theta$,

$$r_n(\hat{\theta} - \theta) \rightsquigarrow \boldsymbol{W} \sim N(\mathbf{0}, \Sigma)$$

for some diverging sequence r_n and a consistently estimable $\Sigma \in \mathbb{R}^{k \times k}$. We model local neighborhoods of $\theta_\star$ via the drifting parameter sequence $\theta_{n,h} = \theta_\star + h/r_n$. Via the second-order delta method, under these local

*Working Title. Emails: xinyue.bei@austin.utexas.edu and mnavjeevan@tamu.edu. We thank...

sequences the behavior of the plug-in estimator $g(\hat{\theta})$ is described by

$$r_n^2(g(\hat{\theta}) - g(\theta_{n,h})) \overset{h}{\rightsquigarrow} h'\boldsymbol{W} + \frac{1}{2}\boldsymbol{W}'\nabla^2 g(\theta_\star)\boldsymbol{W}$$

Notice that the limiting distribution of the plug-in estimator depends on the local parameter h, which is not consistently estimable. This is problematic when constructing confidence intervals in local regions of $\theta_\star$, e.g for constructing valid tests of the null-alternate pair $H_0 : g(\theta) = g(\theta_{n,h})$ against $H_1 : g(\theta) \neq g(\theta_{n,h})$. Moreover, as we establish below, this problem cannot be avoided by considering some alternative to the plug-in estimator — there is no estimator that behaves "regularly" in local areas of $\theta_\star$. However, it may still be possible to construct tests that are reasonably powerful in these local regions; Section **??** provides new uniformly valid testing procedures based on the minimum distance statistic.

Below, we give some empirically relevant examples of when near failure of the delta method may be a relevant concern.

Example 1 (Mediation Analysis)**.** We can consider the setup of a recent paper by van Garderen and van Giersbergen (2024) where the researcher observes estimates $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)'$ of parameters $\theta = (\theta_1, \theta_2)'$, where θ_1 represents the effect of a treatment variable on a mediator and θ_2 represents the effect of the mediator on the outcome. The total mediation effect of the treatment effect on the outcome is then given by $g(\theta) = \theta_1\theta_2$. When θ is close to zero, $\nabla g(\theta) = (\theta_2, \theta_1)'$ is close to zero and we have a "near failure" problem.

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

Example 2 (Impulse Response Function)**.** Consider a autoregressive AR(1) model of the form $y_t = \theta y_{t-1} + u_t$ where $y_t, y_{t-1} \in \mathbb{R}$, $\theta \in \mathbb{R}$, and $u_t \in \mathbb{R}^d$ is a white noise process. The "impulse response function" is defined as $\varphi_h(\theta) = \theta^h$ and measures the impact at time period h of an initial shock. Due to the importance of this in macroeconomic analysis inference on $\varphi_h(\theta)$ has recieved attention from the econometric literature (see Inoue and Kilian (2002), Gospodinov (2004), and Mikusheva (2012)), mostly related to inference when θ is close to one (the so called "unit-root" problem). However, due to the emergence of near-failure asymptotics, inference on the impulse response function is also complicated when θ is close to zero. This is easiest to see when $h = 2$ and θ is modeled as $\theta = \gamma/\sqrt{n}$, in which case $\varphi_h'(\theta) = 2\gamma/\sqrt{n}$ with $\varphi_h''(\theta) = 2$. However, similar problems also emerge when $h > 2$, though analysis of higher order terms is then required.

Example 3 (Breakdown Point Analysis)**.** Consider a missing data setup where the researcher observes $\{Y_i D_i, D_i, X_i\}_{i=1}^n$, where $D_i \in \{0, 1\}$ represents whether or not an observation's final response is observed and X_i is a set of discrete covariates, i.e $X_i \in \mathcal{X} := \{x_1, \ldots, x_d\}$. In order to achieve identification of parameters of interest, assumptions are typically made about the nature of the selection, for example that the data is "missing conditionally at random", i.e $Y_i \perp D_i \mid X_i$. Ober-Reynolds (2024) proposes assessing the robustness of results to these assumptions through a breakdown point analysis. This assesment involves generating a confidence interval for the squared Hellinger distance between P_0, the distribution of $X_i \mid D_i = 0$, and P_1, the distribution of $X_i \mid D_i = 1$. Since X_i is discrete, this distance can be written

$$\delta_{01} := H^2(P_0, P_1) = \frac{1}{2}\sum_{k=1}^{d}(\sqrt{p_0(x_k)} - \sqrt{p_1(x_k)})^2 = 1 - \sum_{k=1}^{d}\sqrt{p_0(x_k)p_1(x_k)}$$

where $p_0(x_k) = \Pr(X = x_k \mid D = 0)$ and $p_1(x_k) = \Pr(X = x_k \mid D = 1)$. Simple sample analogue estimators of $p_0(x_k)$ and $p_1(x_k)$ are $\sqrt{n}$-consistent and asymptotically normal under mild assumptions. However, the derivative of δ_{01} with respect to the quantities $p_0(x_k)$ and $p_1(x_k)$ are given

$$\frac{\partial\,\delta_{01}}{\partial\,p_0(x_k)} = \frac{1}{2}\frac{\sqrt{p_0(x_k)} - \sqrt{p_1(x_k)}}{\sqrt{p_0(x_k)}}, \qquad \frac{\partial\,\delta_{01}}{\partial p_1(x_k)} = -\frac{1}{2}\frac{\sqrt{p_0(x_k)} - \sqrt{p_1(x_k)}}{\sqrt{p_1(x_k)}}.$$

If the data is missing completely at random, that is $D \perp (Y, X)$, then P_0 is equal to P_1 and these derivatives are uniformly equal to zero. In general, when P_0 is "close" to P_1 the quantities $p_0(x_k)$ will be close to $p_1(x_k)$ for $k = 1, \ldots, d$ and these derivatives will be close to zero. This emergence of near failure asymptotics can complicate the construction of confidence intervals for δ_{01} when the true value of δ_{01} is close to zero.

Example 4 (Explained Variance in Linear Regression)**.** Consider a linear regression model,

$$Y = X'\theta + \epsilon, \quad \mathbb{E}[\epsilon X] = 0 \tag{1}$$

and define $\sigma_Y^2 = \mathrm{Var}(Y)$ and $\Sigma_X = \mathbb{E}[XX']$. A parameter of interest in these models may be the proportion of variance in Y explained by the linear model with X, i.e

$$g(\theta) = \frac{\theta'\Sigma_X\theta}{\sigma_Y^2}. \tag{2}$$

If the true slope parameter θ is close to zero then we have a "near failure" problem. This type of parameter is often of interest in labor economics when explaining variation in wage regressions. Recently, Torres et al. (2018) compare the proportion of variation in wages explained by various fixed-effect specifications in order to compare the empirical relevance of proposed models to explain wage dispersion. If a model under consideration can only explain a weak amount of variation in wage dispersion, we may expect θ to be close to zero.

In a related analysis, Card et al. (2013) compare the baseline Abowd et al. (1999) (AKM) model to various extensions in terms of each models ability to explain increases in wage inequality in West Germany. They find that these extensions provide little explanatory power on top of the baseline AKM model. The additional variance explained by these extensions corresponds to the linear regression model with Y equal to the residual from the AKM model and X equal to the new fixed-effect terms introduced by the extended models. They find that these new fixed-effect terms are close to zero suggesting a near failure problem with the delta method for inference on $g(\theta)$.

Example 5 (Weak IV Bias and Size Distortion)**.** In a standard, homoskedastic, linear IV model,

$$\begin{aligned} y_i &= x_i\beta + \epsilon_i \\ x_i &= z_i'\Pi + v_i \end{aligned} \tag{3}$$

with $y_i, x_i \in \mathbb{R}$, $z_i \in \mathbb{R}^{d_z}$, and $\mathbb{E}[(\epsilon_i, v_i)'|z_i] = 0$, weak identification is modeled through a sequence of first stage coefficients that are local to zero, i.e $\Pi = C/\sqrt{n}$ for some fixed $C \neq 0$. Under this asymptotic framework, the standard IV estimator $\hat{\beta}^{\mathrm{IV}}$ is inconsistent with a non-pivotal limiting distribution – the limiting distribution depends on C which is not consistently estimable.

Stock and Yogo (2005) provide bounds on the bias and size distortion of tests based on $\hat{\beta}^{\mathrm{IV}}$ in terms of the concentration parameter, which is a function of (roughly) the magnitude of C. Ganics et al. (2021) extend this analysis and develop confidence intervals for the bias and size distortion by first constructing confidence intervals for the local parameter C. This is done by examining the F-statistic

$$\hat{\mathcal{F}}_0 = \hat{\Pi}'(Z'Z)^{-1}\hat{\Pi}/\sigma_v^2$$

where $\sigma_v^2 = \text{Var}(v_i)$, $\hat{\Pi}$ is the OLS estimate of Π, and $Z = (z_1', \ldots, z_n')'$ so that $(Z'Z)^{-1}$ is the empirical design matrix. The analysis of Ganics et al. (2021) is complicated by the fact that the shape of the limiting distribution of $\hat{\mathcal{F}}_0$ under weak-IV asymptotics also depends on the local parameter C. This dependence is actually an example of near-failure asymptotics. Since $\hat{\mathcal{F}}_0$ is a quadratic form in $\hat{\Pi}$, when the parameter Π is well separated from zero, the distribution of $\hat{\mathcal{F}}_0 - \mathcal{F}_0$ is pivotal, where

$$\mathcal{F}_0 = \Pi'(Z'Z)^{-1}\Pi/\sigma_v^2$$

is closely related to concentration parameter governing the size of the bias or size distortion. However, when Π is in a $\sqrt{n}$-neighborhood of zero this distribution cannot be consistently estimated.

Remark 1 (Higher Order Failures)**.** Examples 1-5 are examples of near-failure problems associated with quadratic forms in a initially estimable parameter θ. However, near failure problems would also occur when conducting inference on $g(\theta)$ when $g(\theta)$ is a higher-order polynomial as well. When dealing with higher-order polynomials, the rate of convergence of the plug-in estimator $g(\hat{\theta})$ would be faster than r_n^2 and the limiting distribution would contain higher order terms, but the basic problem would remain. A recent paper by Dufour et al. (2025) proposes stochastic bounds on behavior of Wald statistics in this setting, however they are interested in testing for specific point-hypotheses rather than constructing confidence intervals — their bounds depend on the specific type of point-hypotheses that they consider.

2 Impossibility Results

In this section we establish that inference on $g(\theta)$ is fundamentally limited in local regions of first-order degeneracy. We begin in Section 2.1 by introducing a parametric framework to study this problem and defining what it means for an estimator to be "regular." In this setting. Section 2.2 then uses a representation theorem to show that the problem reduces to estimation of quadratic forms in a Gaussian shift experiment, where we prove that well behaved estimators cannot be constructed. Section 2.3 extends the analysis in two directions: first, to hypothesis testing problems where the null value of $g(\theta_\star)$ may hold on a nontrivial subset of the parameter space, and second, to infinite-dimensional models, where we show that the impossibility results remain valid so long as the model contains a suitable parametric submodel. Together, these results demonstrate that standard approaches to inference break down in local regions of delta-method failure.

2.1 Setup and Notation

We begin by assuming that the researcher observes data $\boldsymbol{X}_n = (X_1, \ldots, X_n)$ drawn from a parametric model $P_{n,\theta}$,

$$\boldsymbol{X}_n \sim P_{n,\theta} \tag{4}$$

where $\theta \in \Theta \subset \mathbb{R}^d$ and Θ is a nonempty open set. Let $\mathcal{X}_i$ denote the support of X_i, which could be a general space, and denote $\mathcal{X}^n = \times_{i=1}^n \mathcal{X}_i$. We assume that the sequence of statistical models $\big(P_{n,\theta} : \theta \in \Theta\big)$, indexed by the sample size n, is locally asympotically normal in the sense of Le Cam (1960).

Assumption 1 (Local Asymptotic Normality). *There exists a sequence $r_n \to \infty$ such that for every $\theta \in \Theta$ and every sequence $h_n \to h \in \mathbb{R}^d$*

$$\log\left(\frac{dP_{n,\theta+h_n/r_n}}{dP_{n,\theta}}(\boldsymbol{X})\right) = h'\Delta_n - \frac{1}{2}h'\Gamma_\theta h + Z_n(h) \tag{5}$$

where Δ_n converges in distribution to $N(0,\Gamma_\theta)$ under the sequence of measures $P_{n,\theta}$, $\Delta_n \overset{\theta}{\rightsquigarrow} N(0,\Gamma_\theta)$, and $Z_n(h)$ converges in probability to zero under $P_{n,\theta}$ for every $h \in \mathbb{R}^d$, $Z_n(h) \overset{\theta}{\to}_p 0$.

Example (Smooth Parametric Models). A leading example of a model that satisfies (5) is when the researcher observes i.i.d data, $X_i \overset{iid}{\sim} P_\theta$ where $\theta \in \Theta$, an open subset of $\mathbb{R}^d$ and there exists a dominating measure μ such that $P_\theta \ll \mu$ for all $\theta \in \Theta$ and the Radon-Nikodym densities $p_\theta := dP_\theta/d\mu$ are differentiable in quadratic mean, that is there is a function $\dot{\ell}_\theta$ such that for any θ,

$$\int\left[\sqrt{p_{\theta+h}} - \sqrt{p_\theta} - \frac{1}{2}h'\dot{\ell}_\theta\sqrt{p_\theta}\right]^2 d\mu = o(\|h\|^2), \quad h \to 0 \tag{6}$$

and such that the Fisher information, $I_\theta = P_\theta\dot{\ell}_\theta\dot{\ell}_\theta'$, is nonsingular. In for $P_{n,\theta} = \otimes_{i=1}^n P_\theta$ and any $h_n \to h \in \mathbb{R}^d$ the sequence of log likelihood ratios satisfies (see, e.g, van der Vaart (1998), Theorem 7.2):

$$\begin{aligned}
\log\left(\frac{dP_{n,\theta+h/\sqrt{n}}}{dP_\theta}(\boldsymbol{X})\right) &= \log\prod_{i=1}^n \frac{p_{\theta+h/\sqrt{n}}}{p_\theta}(X_i)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n h'\dot{\ell}_\theta(X_i) + h'I_\theta h + R_n(h),
\end{aligned}$$

where $R_n(h) = o_{P_{n,\theta}}(1)$ for every $h \in \mathbb{R}^d$. By the central limit theorem $\frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_\theta(X_i) \overset{P_{n,\theta}}{\rightsquigarrow} N(0, I_\theta)$. Thus, by letting $\Delta_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_\theta(X_i)$ and $\Gamma = I_\theta$ we see that Assumption 1 is satisfied with $r_n = \sqrt{n}$. $\square$

We are interested in a twice continuously differentiable function of θ, $g(\theta)$, and the behavior of estimators of $g(\theta)$ in local regions of a point $\theta_\star \in \Theta$ which is such that the first order derivatives of $g(\cdot)$ at $\theta_\star$ are zero.

Assumption 2 (Differentiability). *The function $g : \Theta \to \mathbb{R}$ is twice continuously differentiable at $\theta_\star$ with $\nabla g(\theta_\star) = 0$ and $\nabla^2 g(\theta_\star) \neq 0$.*

Given the maintained assumption of a locally asymptotically normal model, it is useful to examine regions "close" to $\theta_\star$ by adopting a local paramaterization around $\theta_\star$, defining

$$\theta_{n,h} = \theta_\star + h/r_n$$

and letting $P_{n,h} = P_{n,\theta_\star+h/r_n}$. In our framework an estimator is an arbitrary measurable function of the data $\boldsymbol{X}_n = (X_1,\ldots,X_n)$, $\Psi_n : \mathcal{X}^n \to \mathbb{R}$. We consider sequences of estimators, Ψ_n, that converge in distribution under every sequence of alternative distributions $P_{n,h}$ to some limiting law, $\mathcal{L}_h$. This is denoted

$$r_n^2\left(\Psi_n - g(\theta_{n,h})\right) \overset{h}{\rightsquigarrow} \mathcal{L}_h. \tag{7}$$

where we note that the convergence rate is r_n^2 instead of r_n due to the fact that $g(\theta)$ is "flat" around $\theta_\star$. It is straightforward to show that 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 $g(\theta_\star + h/r_n)$.

Example (Plug-In Estimators)**.** Suppose the researcher has access to an estimator $\hat{\theta}$ of θ that satisfies

$$r_n(\hat{\theta} - \theta_{n,h}) \overset{h}{\rightsquigarrow} \boldsymbol{W}$$

for every $h \in \mathbb{R}^d$. In a smooth parametric model like the one described above, such an estimator could be the MLE or an estimator based on a Bayesian procdures such as the posterior mean. Since $g(\cdot)$ is assumed to be twice continuously differentiable, the second order delta method yields that

$$r_n^2(g(\hat{\theta}) - g(\theta_{n,h})) \overset{h}{\rightsquigarrow} h'\boldsymbol{W} + \frac{1}{2}\boldsymbol{W}'\nabla^2 g(\theta_\star)\boldsymbol{W}$$

Here notice that the behavior of the plug in estimator, $g(\hat{\theta})$, depends on the local parameter h. □

We focus on ruling out "regular" and α-quantile unbiased estimation of $g(\theta_{n,h})$ in local regions of $\theta_\star$.

Definition 1 (Regularity)**.** Let Ψ_n be an estimator satisfying (7), and let $\alpha \in (0, 1)$.

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

(ii) Ψ_n is *α-quantile unbiased* if its limiting α-quantile is zero for every h, i.e. $\mathcal{L}_h\{(-\infty, 0]\} = \alpha$ for all $h \in \mathbb{R}^d$.

2.2 Analysis in the Limiting Experiment

To examine the possible behavior of such estimators, we make use of a representation result, given below in Proposition 1, which is an slight adaptation of Theorem 8.3 in van der Vaart (1998). This earlier result is, in turn, a version of Le Cam's limit of experiments analysis for locally asymptotically normal models (Le Cam, 1970, 1972).

Proposition 1 (Limit Experiment)**.** *Suppose Assumption 1 holds, and let Ψ_n be a sequence of estimators satisfying (7). Then there exists a randomized statistic $\Psi(Z, U)$, where Z is drawn from the Gaussian shift experiment*

$$Z \sim N(h, \Gamma_{\theta_\star}^{-1}), \quad h \in \mathbb{R}^d,$$

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

$$\Psi(Z, U) - \tfrac{1}{2}h^\top \nabla^2 g(\theta_\star)h \ \sim \ L_h \quad \textit{for all } h \in \mathbb{R}^d.$$

Proposition 1 establishes an equivalence between estimating $g(\theta)$ in local regions of delta method failure and estimation of a quadratic form of the mean parameter in a Gaussian shift model where one observes a single draw $Z \sim N(h, \Gamma_{\theta_\star}^{-1})$ where $\Gamma_{\theta_\star}^{-1}$ is known by h is not. In particular, similarly to the approach in Hirano and Porter (2012) , we can rule out sufficiently regular behavior of estimators of $g(\theta)$ in local regions of $\theta_\star$ if the corresponding behavior is not permissible in the Gaussian shift model. Intuitively, sufficiently regular estimation of quadratic forms in the Gaussian shift experiment is not possible since the parameter of interest changes non-linearly as the mean parameter h varies over $\mathbb{R}^d$ while the distribution of Z changes in a linear fashion.

To illustrate, suppose that there was an estimator, $T(Z,U)$, and law, $\mathcal{L}$ with $\int x^2 \, d\mathcal{L}(x) < \infty$, such that $T(Z,U) - h'\nabla^2 g(\theta_\star)h$ is distributed according to $\mathcal{L}$ for all $h \in \mathbb{R}^d$. Since any such estimator can be turned into an unbiased estimator by subtracting off the mean of $\mathcal{L}$, it is without loss of generality to assume that $\mathcal{L}$ is mean zero and thus that $T(Z,U)$ is unbiased. On the other hand the Cramér-Rao lower bound for the variance of any unbiased estimator of $h'\nabla^2 g(\theta_\star)h$ in the Gaussian shift model yields

$$\mathrm{Var}(T(Z,U)) \geqslant 4h'(\nabla^2 g(\theta_\star))'\Gamma_{\theta_\star}^{-1}(\nabla^2 g(\theta_\star))h. \tag{8}$$

By letting h vary over $\mathbb{R}^d$, the right hand side of (8) can be made arbitrarily large while the left hand side is bounded by the second moment of $\mathcal{L}$. Thus, no such estimator can exist. Our full argument relies on analyzing characteristic functions, but the intuition is similar.

Remark 2. It is interesting to compare the argument sketched above to the argument of Hirano and Porter (2012), who rule out regular estimation of $g(\theta)$ when g is directionally, but not fully differentiable at a point $\theta_\star$. The Hirano and Porter (2012) argument relies on analyzing the behavior of a potential regular estimator as the local parameter h approaches zero while our argument relies on analyzing the "global" behavior of a potential regular estimator, i.e the behavior as h varies over $\mathbb{R}^d$. Intuitively, the discrepancy comes because in our limiting experiment our parameter of interest is a quadratic form, which looks approximately linear in local regions of zero but non-linear globally. In contrast, in the limiting experiment of Hirano and Porter (2012) the parameter of interest is a function $\kappa(h)$ which is exactly linear around values of $h \neq 0$, but is not linear around zero. The argument of Hirano and Porter (2012) is able to additionally rule out locally unbiased estimation whereas in our setting locally unbiased estimation is possible. □

Proposition 2. *Let $Z \sim N(h, \Gamma_{\theta_\star}^{-1})$ and $U \sim \mathrm{Unif}(0,1)$ independently of Z. Let J be a $k \times k$ non-zero, symmetric matrix.*

(a) There is no randomized statistic $\Psi(Z,U)$ and law $\mathcal{L}$ on $\mathbb{R}$ with $\Psi(Z,U) - h'Jh \overset{h}{\sim} \mathcal{L}$ for all $h \in \mathbb{R}^d$.

(b) Let $\{\mathcal{L}_h\}_{h\in\mathbb{R}^d}$ be a system of probability measures on $\mathbb{R}$ such that (i) $\mathcal{L}_h\{(-\infty, 0]\} = \alpha$ for some $\alpha \in (0,1)$ and (ii) the CDFs associated with $\mathcal{L}_h$, $F_h(\cdot)$, are differentiable at zero with derivative bounded below by some $\epsilon > 0$. Then, there does not exist a randomized statistic $\Psi(Z,U)$ such that $\Psi(Z,U) - h'Jh \sim \mathcal{L}_h$ for all $h \in \mathbb{R}^d$.

Together, Propositions 1 and 2 can be combined for the main result of this section, which rules out sufficiently regular estimation in local areas of first-order degeneracy.

Theorem 1 (Impossibility of Regular Estimation)**.** *Suppose Assumptions 1 and 2 hold.*

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

$$r_n^2\big(\Psi_n - g(\theta_{n,h})\big) \overset{h}{\rightsquigarrow} \mathcal{L} \quad \textit{for all } h \in \mathbb{R}^d.$$

(b) Let $\{\mathcal{L}_h\}_{h\in\mathbb{R}^d}$ be a family of distributions such that (i) $\mathcal{L}_h\{(-\infty, 0]\} = \alpha$ for some fixed $\alpha \in (0,1)$ and all h, and (ii) the CDFs, $F_h(\cdot)$, of L_h are differentiable at zero with derivatives bounded below by $\epsilon > 0$.

Then there is no estimator sequence Ψ_n *such that*

$$r_n^2\big(\Psi_n - g(\theta_{n,h})\big) \overset{h}{\rightsquigarrow} \mathcal{L}_h \quad \textit{for all } h \in \mathbb{R}^d.$$

Theorem 1 rules out sufficiently well behaved estimation of $g(\theta)$ when the true parameter is "close" to $\theta_\star$. In particular Theorem 1(a) rules out the possibility of regular estimation — the behavior of any estimator Ψ_n of $g(\theta)$ must depend, in local regions of $\theta_\star$, on the local parameter h, which cannot be consistently estimated. Apart from suggesting that standard approaches to inference, such as Wald-type inference procedures, are not available in this setting, the lack of regular estimators implies that our standard theory of efficient estimation does not apply in local regions of delta-method failure. In particular, one can show that estimators of $g(\theta)$ that are efficient under a standard asymptotic regime are dominated in local asymptotic mean squared error under near-failure asymptotics.

Similarly, Theorem 1(b) rules out the possibility of a sufficiently well behaved α-quantile unbiased estimator. Since any asymptotically similar confidence interval of the form $(-\infty, \hat{c}]$ can be converted into a locally asymptotically α-quantile unbiased estimator by taking $\Psi_n = \hat{c}$, Theorem 1(b) essentially rules out the possibility of similar one sided confidence intervals for $g(\theta)$ in local regions of $\theta_\star$. As with the impossibility of locally asymptotically α-quantile unbiased estimation for directionally but not fully differentiable parameters established in Hirano and Porter (2012), the result in Theorem 1 requires some regularity conditions on the system of limiting laws $\{\mathcal{L}_h\}_{h\in\mathbb{R}^d}$.[1] The regularity condition in Theorem 1(b) implies that, if a locally asymptotically α-quantile estimator were to exist, its associated limiting laws $\mathcal{L}_h$ must become "flat". In particular, if each limiting law $\mathcal{L}_h$ has a density with respect to Lebesgue measure, these densities evaluated at zero, which is by definition the α-quantile of each $\mathcal{L}_h$, must be able to be made arbitrarily small.[2]

2.3 Extensions: Hypothesis Testing and Infinite Dimensional Models

The above analysis rules out standard approaches to inference on $g(\theta)$ in local regions of $\theta_\star$. These results are informative when one is interested in constructing confidence intervals for $g(\theta)$ around points of first-order degeneracy when the data is drawn from a parametric model satisfying Assumption 1. In this subsection, we consider two extensions of our results. In the first, we consider the somewhat simpler problem of testing the null hypothesis $H_0 : g(\theta) = g(\theta_\star)$. We show that, if the null hypothesis contains a sufficiently rich set of values, and a similar test exists, this similar test must have low power in local regions of $\theta_\star$. In the second extension we generalize Theorem 1 to infinite dimensional, i.e semiparametric or nonparametric, models.

2.3.1 Hypothesis Testing

In this subsection we consider the problem of testing the null hypothesis $H_0 : g(\theta) = g(\theta_\star)$ where the alternative can be one sided, i.e $H_1 : g(\theta) > g(\theta_\star)$ or two-sided, $H_1 : g(\theta) \neq g(\theta_\star)$. To setup, define $\mathcal{H}_\star$ to be the set of local parameters, $h \in \mathbb{R}^d$, such that $g(\theta_{n,h})$ is asymptotically indistinguishable from $g(\theta_\star)$;

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

[1] The impossibility result in Hirano and Porter (2012) requires that the CDF associated with $\mathcal{L}_0$ is differentiable at zero with positive derivative.

[2] For example, suppose that $\{\mathcal{L}_h\}_{h\in\mathbb{R}^d}$ is a family of Gaussian distributions on $\mathbb{R}$ associated with a locally α-quantile unbiased estimator. Then, the variance of these Gaussian distributions must be able to be made arbitrarily large as h ranges over $\mathbb{R}^d$.

To build a similar test for the null-alternate pairs described above we only need an estimator which is well behaved on $\mathcal{H}_\star$ rather than on the entirety of $\mathbb{R}^d$. We describe these estimators as locally $\mathcal{H}_\star$-α-quantile unbiased.

Definition 2 ($\mathcal{H}_\star$-α-quantile unbiased)**.** Let Ψ_n be an estimator satisfying (7), and let $\alpha \in (0,1)$. We say that Ψ_n is *locally $\mathcal{H}_\star$-α-quantile unbiased* if for every $h \in \mathcal{H}_\star$ the α-quantile of its limiting distribution is zero, i.e. $\mathcal{L}_h\{(-\infty, 0]\} = \alpha$ for all $h \in \mathcal{H}_\star$.

Notice that this definitions are less restrictive than the standard definitions locally-α-quantile unbiased estimators since it only require the estimator to be well behaved on the set of local parameters $\mathcal{H}_\star$ such that $g(\theta_{n,h}) = g(\theta_\star)$. Our main result in this subsection notes that, under a milder regularity, if such an estimator sequence exists then the local asymptotic power curve of a level α-test based on such an estimator sequence must be flat at $\theta_\star$ in the sense that the derivative of the local asymptotic power curve with respect to the local parameter h exists and is equal to zero. Define the local asymptotic power curve as

$$\mathcal{P}(h) = \limsup_{n\to\infty} \Pr_{\theta_{n,h}}(\Psi_n - g(\theta_\star) \leqslant 0)) \tag{9}$$

Proposition 3. *Let Ψ_n be a locally asymptotically $\mathcal{H}_\star$-α-quantile unbiased estimator sequence with limiting with limiting CDF under the $h = 0$, $F_0(a) = \mathcal{L}_0\{(-\infty, 0]\}$, differentiable at zero. Then the directional derivative of the local asymptotic power curve, $\mathcal{P}(h)$, in directions $h \in \mathcal{H}_\star$ exists and is equal is zero, i.e*

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

In particular, if $\mathcal{H}_\star$ spans $\mathbb{R}^d$ then $\nabla\mathcal{P}(0)$ exists and is equal to zero.

Example (Mediation Model)**.** Consider a mediation model where the original model is given ($P_\theta : \theta \in \Theta \subseteq \mathbb{R}^2$). Suppose that the researcher is interested in testing the null hypothesis $H_0 : \theta_1\theta_2 = 0$, that is $g(\theta) = \theta_1\theta_2$ and $\theta_\star = \mathbf{0}$. We can see that $\mathcal{H}_\star$ is given $\mathcal{H}_\star = \{h \in \mathbb{R}^2 : h_1h_2 = 0\}$. Since $\operatorname{span}(\mathcal{H}_\star) = \mathbb{R}^2$ we have that $\nabla\mathcal{P}(0) = 0$ for the test based on any locally $\mathcal{H}_\star$-α-quantile unbiased estimator. □

Example (Squared Mean)**.** On the other hand consider the case where $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$ and the researcher is interested in testing the null hypothesis $H_0 : \theta_1^2 + \theta_2^2 = 0$. In this case $g(\theta) = \theta_1^2 + \theta_2^2$, $\theta^\star = 0$, and $\mathcal{H}_\star = \{(0,0)\}$. Since $\operatorname{span}(\mathcal{H}_\star) = \{0\}$ Proposition 3 does not apply and powerful similar tests for the null hypothesis $H_0 : \theta_1^2 + \theta_2^2 = 0$ can be constructed, see e.g Chen and Fang (2019). □

The upshot of Proposition 3 is that even if there exists an asymptotically similar level α test, the aforementioned test will have a power curve that is flat in local regions of $\theta_\star$. As such, such a test may have low power in regions close to $\theta_\star$. This is not the case in traditional settings[3] and emerges because the derivative of a quadartic form with respect to its arguments is zero at zero.

It also may be notable that the regularity condition in Proposition 3 is milder than that of Theorem 1(b). Using the proof of Proposition 3 one can show that if there does exist a α-quantile-unbiased estimator such that with limiting CDF under $h = 0$ differentiable at zero then the local power curve $\mathcal{P}(h)$, defined in (9), will satisfy $\nabla\mathcal{P}(0) = 0$.

[3] In a standard one sided test with an estimator Ψ_n that satisfies $r_n(\Psi_n - \theta_{n,h}) \overset{h}{\rightsquigarrow} N(0, \sigma^2)$ the local asymptotic power curve is given $\mathcal{P}(h) = 1 - \Phi(c_{1-\alpha} - h/\sigma)$, where $\Phi(\cdot)$ is the standard normal CDF and $c_{1-\alpha}$ is it's $1 - \alpha$ quantile. Here $\frac{\partial}{\partial h}\mathcal{P}(h)\Big|_{h=0} \neq 0$

Remark. Recent papers by van Garderen and van Giersbergen (2024) and Dufour et al. (2025) also consider the problem of testing the null hypothesis $H_0 : g(\theta) = g(\theta_\star)$ in various contexts. van Garderen and van Giersbergen (2024) consider the case of the mediation model, that is where $\theta = (\theta_1, \theta_2)'$ and $g(\theta) = \theta_1\theta_2$. They assume that the researcher has access to an asymptotically normal estimate of θ, $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)'$ and show that there is no reasonable similar test of the form reject if $\max\{|\hat{\theta}_1|, |\hat{\theta}_2|\} > g(\min\{|\hat{\theta}_1|, |\hat{\theta}_2|\})$, where $g(\cdot)$ may be an arbitrary function. Similarly, Dufour et al. (2025) consider the behavior of Wald type tests based on the test statistic $W_n = \frac{g(\hat{\theta}) - g(\theta_\star)}{\nabla g(\hat{\theta})' J \nabla g(\hat{\theta})}$, where J represents the asymptotic variance of $\hat{\theta}$. The authors show that, when θ is close to $\theta_\star$, the behaviour of the Wald statistic can be irregular and propose alternate critical values for testing the null hypothesis $H_0 : g(\theta) = g(\theta_\star)$ using W_n.

We view our results as complementary to these existing results. While the results in Proposition 3 are more limited in scope, we show only that similar tests may not be powerful around $\theta_\star$, their breadth is significantly wider — we cover a broad class of problems and the testing procedures considered Proposition 3 can depend arbitrarily on the data, $\boldsymbol{X}_n$, they do not have to be based on an initial estimator of $\hat{\theta}$ nor take the form of the tests considered by van Garderen and van Giersbergen (2024). □

2.3.2 Infinite Dimensional Models

Of course, in many setting the researcher may not be willing to assume that the data comes from a finite dimensional parametric model as described in the previous section. Following a tradition on studying semi-parametric efficiency Bickel et al. (1993), we show that this does not affect our impossibility results so long as the larger model contains a parametric submodel satisfying Assumptions 1 and 2.

To be formal, let the model $\mathcal{P}$ be a collection of sequences of probability measures on the sample space $\mathcal{X}^n$ from the previous section. That is, each element of $\mathcal{P}$ looks like a sequence of probability measures, $\{P_n\}$ where each probability measure P_n is defined on the sample space $\mathcal{X}^n$. A finite dimensional submodel, $\mathcal{P}_f$, is some smaller collection of sequences of probability measures that can be parameterized as $\mathcal{P}_f = (\{P_{n,\theta}\}_{n\in\mathbb{N}} : \theta \in \Theta)$ for an open set $\Theta \in \mathbb{R}^{d_f}$. Fix a "centering" sequence of probability measures $\{P_{0,n}\}_{n\in\mathbb{N}} \in \mathcal{P}$. We say that the submodel passes through $\{P_{0,n}\}$ if $\{P_{0,n}\} \in \mathcal{P}_f$, that is $\{P_{0,n}\} = \{P_{n,\theta}\}$ for some $\theta \in \Theta$. We will call such a parametric model "regular" if Assumption 1 holds and the model passes through $\{P_{0,n}\}$.

We suppose that the object of interest is a quantity that depends on the sequence of underlying probability measures, that is we can think of the estimand $g[\{P_n\}]$ as a functional defined on $\mathcal{P}$. For any regular parametric model, $\mathcal{P}_f$, this implicitly defines a function on θ via the relation $g_f(\theta) = g[\{P_{n,\theta}\}]$.

With this notation defined, we show that the results of Theorem 1 can be extended in a straightforward fashion 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 Assumption 2 with $\{P_{n,\theta_\star}\} = \{P_{0,n}\}$.*

(a) There is no estimator sequence Ψ_n and law $\mathcal{L}$ on $\mathbb{R}$ such that, along every regular parametric submodel $\mathcal{P}_f$,

$$r_n^2\big(\Psi_n - g_f(\theta_\star + h/r_n)\big) \overset{h}{\rightsquigarrow} \mathcal{L} \quad \text{for all } h \in \mathbb{R}^{d_f}.$$

(b) Let $\{\mathcal{L}_h\}_{h\in\mathbb{R}^{d_f}}$ be a family of distributions such that (i) $\mathcal{L}_h\{(-\infty, 0]\} = \alpha$ for some $\alpha \in (0,1)$ and all h, and (ii) each $\mathcal{L}_h$ has a CDF, F_h, differentiable at zero with derivative bounded below by $\epsilon_f > 0$. Then

there is no estimator sequence Ψ_n *such that, along every regular parametric submodel* $\mathcal{P}_f$*,*

$$r_n^2\big(\Psi_n - g_f(\theta_\star + h/r_n)\big) \overset{h}{\rightsquigarrow} \mathcal{L}_h \quad \textit{for all } h \in \mathbb{R}^{d_f}.$$

3 Minimum Distance Based Inference

In this section, we construct a uniformly valid confidence interval for $g(\theta)$ using estimator $\hat{\theta}$. The confidence interval is constructed by interting the hypothesis $H_0 : g(\theta) = \tau$, and we use a minimum distance test statistic

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

We focus on the setting where the standard first order approximation of $g(\theta)$ fails at θ_*, but the second order derivative is non-degenerate, i.e. $\frac{\partial g}{\partial \theta}(\theta_*) = 0$ and $\frac{\partial^2 g}{\partial \theta \partial \theta'}(\theta_*) = H$ with full rank H.

In Section 3.1, we discuss a simple case where $\theta \in \mathbb{R}^2$ and H is indefinite, and we provide sufficient conditions under which the standard critical value $q_{\chi_1^2, 1-\alpha}$ is uniformly valid. In Section 3.2, we propose a computationally simple method to construct a critical value for g with $d_\theta > 2$. The inference procedure can be easily generalized to cases with higher order singularity, see Remark 4.

3.1 Two-Dimensional θ and Indefinite H

To simplify notation, consider the null hypothesis

$$H_0 : g(\theta_1, \theta_2) := (1 + \rho)\theta_2^2 - (1 - \rho)\theta_1^2 = \tau, \tag{10}$$

with $|\rho| < 1$ and $\tau \geqslant 0$. The restriction $|\rho| < 1$ guarantees that H is indefinite, while $\tau \geqslant 0$ is a normalization. For simplicity, let $n = 1$, and assume $\hat{\theta} - \theta \sim \mathcal{N}(0, \mathcal{I}_2)$. The quadratic form g and the normality of $\hat{\theta}$ can be viewed as second order approximations, with general asymptotic results provided in Theorem 2.

Let $X_2(\theta_1) = \sqrt{\frac{\tau + (1-\rho)\theta_1^2}{1+\rho}}$ be the solution of θ_2 such that (10) holds. Let $\mathcal{S}_0(\tau)$ be the null parameter space, which contains two separate curves $\mathcal{S}_0^+(\tau)$ and $\mathcal{S}_0^-(\tau)$,

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

where

$$\mathcal{S}_0^+(\tau) = \Big\{\big(x_1, X_2(x_1)\big) : x_1 \in \mathbb{R}\Big\}, \quad \mathcal{S}_0^-(\tau) = \Big\{\big(x_1, -X_2(x_1)\big) : x_1 \in \mathbb{R}\Big\}.$$

Let $\mathcal{S}(\tau, c)$ be the acceptance region with critical value c^2, equivalently the c-enlargement of $\mathcal{S}_0(\tau)$,

$$\mathcal{S}(\tau, c) = \Big\{(x_1, x_2) : (x_1 - \theta_1)^2 + (x_2 - \theta_2)^2 \leqslant c^2, (\theta_1, \theta_2) \in \mathcal{S}_0(\tau)\Big\}.$$

Proposition 4. *Suppose either* $\tau \geqslant \frac{c^2(1-\rho)^2}{1+\rho}$ *or* $\rho \geqslant 0$*. For all* $\theta \in \mathcal{S}_0$*,* $\hat{\theta} - \theta \sim \mathcal{N}(0, \mathcal{I}_2)$*,*

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

Proposition 4 shows that the standard MD test with critical value $c = \sqrt{q_{\chi_1^2,1-\alpha}}$ remains valid under a curved null. The proof builds on Lemma 1, which connects the coverage rate of the MD test under a curved null to that under a standard linear null, and provides a sufficient condition for validity.

Let $B\left((x_1, x_2), r\right)$ denote the ball centered at (x_1, x_2) with radius r, and $\partial B\left((x_1, x_2), r\right)$ its boundary (i.e. the circle). Let $\widehat{AB}$ be the arc from A to B, and $\overline{AB}$ the line segment.

Lemma 1. *Fix $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$, $c = \sqrt{q_{\chi_1^2,1-\alpha}}$. If the set $\bar{\mathcal{S}}$ satisfies*

1. *$B(\theta, c) \subset \bar{\mathcal{S}}$.*
2. *For all $r > c$, $\left|\text{length}\left(\partial B(\theta, r) \cap \bar{\mathcal{S}}\right)\right| \geqslant 4r \arcsin \frac{c}{r}$.*

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

$$P\left(\hat{\theta} \in \bar{\mathcal{S}}\right) \geqslant 1 - \alpha.$$

Proof. The key idea of Lemma 1 is to compare the coverage probability of $\bar{\mathcal{S}}$ that of an auxiliary acceptance set

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

It is trivial that $P\left(\hat{\theta} \in \mathcal{S}_{\text{aux}}\right) = 1 - \alpha$. We will show that the coverage probability of $\bar{\mathcal{S}}$ is bounded below by that of $\mathcal{S}_{\text{aux}}$.

To simplify the comparison, we switch to polar coordinates. Let $\hat{\theta} = (\theta_1 + r\cos\omega, \theta_2 + r\sin\omega)$, then

$$\begin{aligned} P\left(\hat{\theta} \in \mathcal{S}_{\text{aux}}\right) &= \frac{1}{2\pi}\int_{r=0}^{+\infty}\int_{\omega=-\frac{\pi}{2}}^{\frac{3}{2}\pi} \mathbf{1}\left[(r\sin\omega)^2 \leqslant c^2\right] \mathrm{d}\omega \exp(-\frac{r^2}{2}) r\mathrm{d}r \\ &= \int_{r=0}^{c} \exp(-\frac{r^2}{2}) r\mathrm{d}r + \int_{r=c}^{+\infty} \frac{4\arcsin\frac{c}{r}}{2\pi} \exp(-\frac{r^2}{2}) r\mathrm{d}r. \end{aligned} \tag{11}$$

To see (11), note that if $r \leqslant c$, then $(r\sin\omega)^2 \leqslant c^2$ holds for all $\omega \in [-\frac{1}{2}\pi, \frac{3}{2}\pi]$; if $r > c$, then

$$\begin{aligned} &(r\sin\omega)^2 \leqslant c^2, \omega \in [-\frac{\pi}{2}, \frac{3}{2}\pi] \\ \Leftrightarrow &\omega \in \left[-\arcsin(\frac{c}{r}), \arcsin(\frac{c}{r})\right] \cup \left[\pi - \arcsin(\frac{c}{r}), \pi + \arcsin(\frac{c}{r})\right]. \end{aligned}$$

Now consider $P\left(\hat{\theta} \in \bar{\mathcal{S}}\right)$. By Condition 1 and 2,

$$\begin{aligned} P\left(\hat{\theta} \in \bar{\mathcal{S}}\right) &= \frac{1}{2\pi}\int_{r=0}^{+\infty} \frac{1}{r}\left|\text{length}\left(\partial B(\theta, r) \cap \bar{\mathcal{S}}\right)\right| \exp(-\frac{r^2}{2}) r\mathrm{d}r \\ &\geqslant \int_{r=0}^{c} \exp(-\frac{r^2}{2}) r\mathrm{d}r + \int_{r=c}^{+\infty} \frac{4\arcsin\frac{c}{r}}{2\pi} \exp(-\frac{r^2}{2}) r\mathrm{d}r. \end{aligned} \tag{12}$$

This lower bound matches the expression in (11), which completes the proof. An illustration is provided in Figure 1. □

Note that Condition 1 of Lemma 1 holds for all $\theta \in \mathcal{S}_0(\tau)$ and $\bar{\mathcal{S}} = \mathcal{S}(\tau, c)$, since $\mathcal{S}(\tau, c)$ is the union of $B(\theta, c)$ over $\theta \in \mathcal{S}_0(\tau)$. Hence, it suffices to verify Condition 2. The proof of Proposition 4 considers two

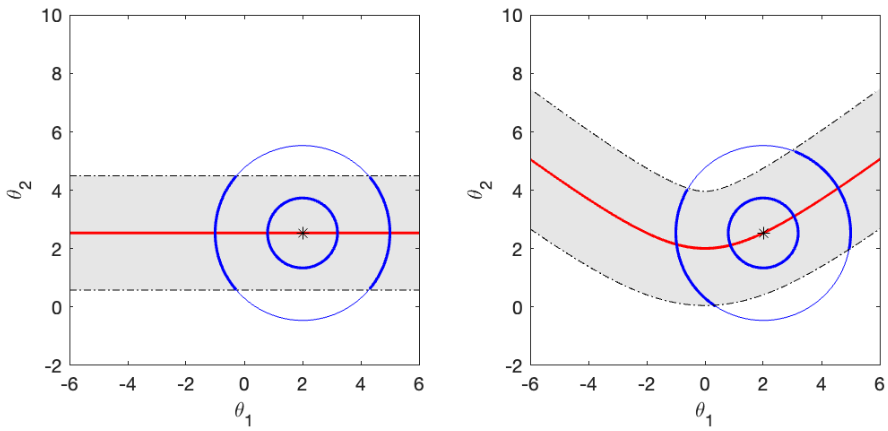


Figure 1: Lemma 1: Acceptance Region of Linear and Curved Null.

The red curve shows the null parameter space $\mathcal{S}_0(\tau)$. Shaded areas denote the acceptance regions $\mathcal{S}_{\text{aux}}$ (left) and $\bar{\mathcal{S}}$ (right). "*" represents the true value θ, and the blue circles represent $B(\theta, r)$ with bold segments indicating the portions inside the acceptance regions. If, all r, the bold segment in the right panel is longer than that in the left, then the acceptance rate of $\bar{\mathcal{S}}$ is at least $1 - \alpha$.

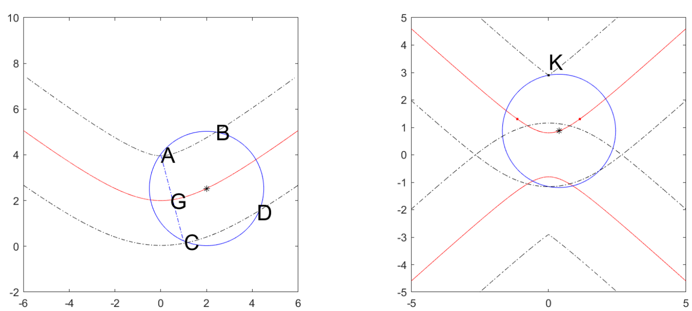


Figure 2: Proposition 4: Acceptance Regions under Low (Left) and High (Right) Curvature.

Red curves represent $\mathcal{S}_0^+(\tau)$ (left) and $\mathcal{S}_0(\tau)$ (right). Black dash curves represents the boundaries of $\mathcal{S}^+(\tau, c)$ (left) and $\mathcal{S}^+(\tau, c)$ and $\mathcal{S}^-(\tau, c)$ (right). "*" represents θ, and the blue curves represents $\partial B(\theta, r)$ for some $r > c$.

cases: (i) $|\tau| \geqslant \frac{c^2(1-\rho)^2}{1+\rho}$, where the curvature of $\mathcal{S}_0(\tau)$ is not too large; (ii) $|\tau| < \frac{c^2(1-\rho)^2}{1+\rho}$ with $\rho \geqslant 0$, where the curvature is large but $\mathcal{S}_0^+(\tau)$ and $\mathcal{S}_0^-(\tau)$ are sufficiently close.

We begin with case (i). Without loss of generality, let $\theta \in \mathcal{S}_0^+(\tau)$. In this setting, Lemma 1 holds with $\bar{\mathcal{S}} = \mathcal{S}^+(\tau)$, where $\mathcal{S}^+(\tau, c)$ is the c-enlargement of $\mathcal{S}_0^+(\tau)$. Let $\mathcal{C}_u(\tau)$ and $\mathcal{C}_\ell(\tau)$ denote the upper and lower boundaries of $\mathcal{S}^+(\tau, c)$, see Figure 2 left panel. For $r > c$, $\partial B(\theta, r)$ intersects $\mathcal{C}_u(\tau)$ at points A and B (with A to the left of B) and $\mathcal{C}_\ell(\tau)$ at points C and D (with C to the left of D). To show that length $\left(\widehat{AC}\right)$ + length $\left(\widehat{BD}\right) \geqslant 4r \arcsin \frac{c}{r}$, it suffices to show that

$$\text{length}\left(\overline{AC}\right) \geqslant 2c \text{ and length}\left(\overline{BD}\right) \geqslant 2c.$$

Suppose, for contradiction, that length $\left(\overline{AC}\right) < 2c$. Let AC intersects $\mathcal{S}_0^+(\tau)$ at point G. Then $B(G, c) \nsubseteq \mathcal{S}^+(\tau)$, which contradicts the definition of $\mathcal{S}^+(\tau, c)$. This verifies Condition 2 of Lemma 1, and hence MD is valid.

Next we consider case (ii). Due to the high curvature, $\mathcal{C}_u(\tau)$ has a kink. Thus, there exists $\theta \in \mathcal{S}_0^+(\tau)$ and $r > c$ such that $\partial B(\theta, r)$ does not intersect $\mathcal{C}_u(\tau)$, see Figure 2 right panel, and the argument from case (i) no longer applies. We therefore analyze the full region $\mathcal{S}(\tau, c)$ rather than just $\mathcal{S}^+(\tau)$. With $\rho \geqslant 0$, such $B(\theta, r)$ is sufficiently close to $\mathcal{S}_0^-(\tau)$, and we can show that

$$\left|\text{length}\left(\partial B_r\left((\theta_1,\theta_2), r\right) \cap \mathcal{S}(c)\right)\right| = 2\pi r \geqslant 4r \arcsin\frac{c}{r}. \tag{13}$$

The remainder of the proof is the same as in case (i).

Next, we present the asymptotic results for general data generating processes.

Assumption 3. *(Regularity Conditions)Suppose that Combine this with previous assumptions if needed*

1. *$\Theta = \{\theta_P : P \in \mathcal{P}\}$ is a compact set.*
2. *$g : \Theta \to \mathbb{R}$ is twice continuously differentiable on Θ^ϵ for some $\epsilon > 0$, where $\Theta^\epsilon = \left\{\theta : ||\theta - \tilde{\theta}|| \leqslant \epsilon, \tilde{\theta} \in \Theta\right\}$.*
3. *For all $\theta \in \Theta \backslash \{\theta_\star\}$, $\frac{\partial g(\theta)}{\partial \theta} \neq 0$.*
4. *Let BL_1 denote the set of Lipchitz functions which are bounded by 1 in absolute value and have Lipschitz constant bounded by 1. Assume 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 $\xi_P \sim \mathcal{N}(0, \Sigma_P)$.

5. *Let $\mathcal{S}$ denote the set of matrices with eigenvalues bounded below by $\underline{e} > 0$ and above by $\bar{e} \geqslant \underline{e}$. For all $P \in \mathcal{P}$, $\Sigma_P \in \mathcal{S}$.*
6. *For all $\varepsilon > 0$,*

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} P\left(\left\|\hat{\Sigma}_n - \Sigma_P\right\| > \varepsilon\right) = 0.$$

Assumption 4. $\frac{\partial g(\theta_*)}{\partial \theta} = 0$. $\frac{\partial^2 g(\theta_*)}{\partial\theta\partial\theta'} = H$ *has full rank. Combine this with previous assumptions if needed*

Theorem 2. *Suppose d =2, Assumption 3 and 4 hold. 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 $\rho_P = \frac{\lambda_{P,1}+\lambda_{P,2}}{|\lambda_{P,1}-\lambda_{P,2}|}$. If either*

$$\mathcal{P}_n \subseteq \left\{P \in \mathcal{P} : \rho_P \in [0, 1)\right\} \tag{14}$$

or

$$\mathcal{P}_n \subseteq \left\{P \in \mathcal{P} : 2\left|\lambda_{P,1} - \lambda_{P,2}\right|^{-1}\left|g(\theta_P)\right| \geqslant \frac{c^2(1-\rho_P)^2}{1+\rho_P}\right\}, \tag{15}$$

then

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

Remark 3. To illustrate Theorem 2, consider $g(\theta) = \theta_1\theta_2$ as motivated by Example X. In this case, $H = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. With $\Sigma_P = \begin{bmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$, we have $\lambda_{P,1} = \text{sign}\left(g(\theta_P)\right)(r-1)\sigma_1\sigma_2$ and $\lambda_{P,2} =$

sign $\left(g(\theta_P)\right)(r+1)\,\sigma_1\sigma_2,\ \rho_P = \text{sign}\left(g(\theta_P)\right) r$. Therefore, (14) holds when $r=0$, and the MD test with the simple critical value yields a uniformly valid confidence interval for the mediation effect. If sign $\left(g(\theta_P)\right) r < 0$, then Condition (15) is satisfied when

$$\frac{\left|g(\theta_P)\right|}{\sigma_1\sigma_2} \geqslant \frac{c^2(1+|r|)^2}{1-|r|}.$$

3.2 General Case

In this section, we present the inference procedure for a general function g. The procedure is based on a local approximation of the test statistic. First, consider the case where the true parameter value θ_n satisfies $\theta_n = \theta_* + h_n/r_n$. The test statistic is given by

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

where $\tilde{\theta}_n$ denotes the minimizer. Under H_0, we have $\tilde{\theta}_n = \theta_n + O_p(\frac{1}{r_n})$. A second order Taylor expansion of the restriction $g(\tilde{\theta}_n) = g(\theta_n)$ gives

$$r_n^2(\tilde{\theta}_n - \theta_*)' H (\tilde{\theta}_n - \theta_*) = h_n' H h_n + o_p(1).$$

In addition, let $\mathbb{Z}_n = r_n(\hat{\theta} - \theta_n)$, we can write

$$\begin{aligned} r_n(\hat{\theta}_n - \tilde{\theta}_n) &= r_n \left((\hat{\theta}_n - \theta_n) + (\theta_n - \theta_*) - (\tilde{\theta}_n - \theta_*)\right) \\ &= \mathbb{Z}_n + h_n - \sqrt{n}(\tilde{\theta}_n - \theta_*) \end{aligned}$$

In sum, given h_n, we can approximate $\hat{T}$ by

$$\hat{T}^*(h_n) = \inf_{t: t'Ht = h_n' H h_n} \left(\mathbb{Z}_n^* + h_n - t\right)' \hat{\Sigma}_n^{-1} \left(\mathbb{Z}_n^* + h_n - t\right) \tag{16}$$

where $\mathbb{Z}_n^* | (\hat{\theta}_n, \hat{\Sigma}_n) \sim \mathcal{N}(0, \hat{\Sigma}_n)$. In Lemmas **??** and **??**, we show that $\hat{T}$ and $\hat{T}^*(h_n)$ have the same asymptotic distribution, regardless of whether h_n converges to $h \in \mathbb{R}$ or diverges to infinity. Intuitively, if $h_n \to \infty$, the restriction for the optimizer $\tilde{t}$ in (16) $\tilde{t}' H \tilde{t} = h_n' H h_n$ implies

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

i.e., an approximately linear restriction. The equality follows from $\tilde{t} - h_n = O_p(1)$. In this case, both $\hat{T}$ and $\hat{T}^*(h_n)$ are approximated χ_1^2.

Given h_n, we can easily get the quantile of $\hat{T}^*$ by simulation. However, h_n is a nuisance parameter that cannot be consistently estimated. Next, we propose a two step feasible critical value. In the first step, we construct a $(1-\eta)$ confidence set for h_n. Let $c_h = Q\left(\max_j |\frac{\mathbb{Z}_{n,j}^*}{\hat{\sigma}_{n,j}}|; 1-\eta\right)$, this set be defined as

$$\mathcal{H} = \prod_{j=1}^{d} \left[\sqrt{n}(\hat{\theta}_n - \theta_*)_j - \hat{\sigma}_{n,j} c_h,\ \sqrt{n}(\hat{\theta}_n - \theta_*)_j + \hat{\sigma}_{n,j} c_h\right]. \tag{17}$$

In the second step, we construct the critical value based on the $\frac{1-\alpha}{1-\eta}$ quantile of $\hat{T}^*$ conditional on the first step. That is, let

$$\hat{c} = \sup_{h \in \mathcal{H}} Q\left(\hat{T}^*(h) \Big| \max_j |\frac{\mathbb{Z}^*_{n,j}}{\hat{\sigma}_{n,j}}| \leqslant c_h; \frac{1-\alpha}{1-\eta}\right), \tag{18}$$

and reject H_0 if $\hat{T} > \hat{c}$. In (18), the construction of $\hat{c}$ takes into account the first step selection, thus it is less conservative than simple Bonferroni correction.

Theorem 3. *Under Assumption 3 and 4, it holds that*

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

In addition, if $\left\|\sqrt{n}(\theta_{P_n} - \theta_*)\right\| \to \infty$,

$$\lim_n P_n\left(\hat{T} \leqslant \hat{c}\right) \in [1-\alpha, 1-\alpha+\eta). \tag{19}$$

The slight conservativeness arises from the two-step procedure. Alternatively, we can introduce a pretest to check whether h_n is far away from zero, e.g. $||h_n|| > \ln r_n$. If so, we can use the standard critical value $q_{\chi^2_1, 1-\alpha}$. The cost is that we need to introduce an extra tuning parameter.

Remark 4. In general, if H is singular and g is higher order identified, we can construct the critical value using a similar two step procedure. In the first step, we construct a $1-\eta$ confidence set $\hat{\Theta}$ for θ. In the second step, we define the critical value as

$$\sup_{\theta \in \hat{\Theta}, g(\theta) = \tau} Q\left(\inf_{\vartheta : g(\vartheta) = \tau} r_n^2 \left(r_n^{-1}\mathbb{Z}^*_n + \theta - \vartheta\right)' \hat{\Sigma}_n^{-1} \left(r_n^{-1}\mathbb{Z}^*_n + \theta - \vartheta\right); 1 - \alpha + \eta\right).$$

Remark 5. Dufour, Renault, and ZindeWalsh (2025) and Dufour and Valery (2025) study hypothesis testing under first-order degeneracy. They propose testing procedures based on Wald-type statistics. Dufour, Renault, and ZindeWalsh (2025) show that when g is a vector-valued function and the degree of singularity differs across elements of g, the Wald-type test statistic may diverge, complicating inference. In contrast, the MD test considered in this paper yields a test statistic that is first-order stochastically dominated by χ^2_d, regardless of the level of singularity in g. Moreover, Dufour, Renault, and ZindeWalsh (2025) and Dufour and Valery (2025) focus solely on hypothesis tests at a fixed point, i.e., testing $g(\theta) = g(\theta_*)$, whereas this paper aims to construct uniformly valid confidence intervals.

Remark 6. Andrews and Mikusheva (2016) construct a uniformly valid MD test based on a geometric approach that incorporates the curvature of the null restriction $g(\theta) = \tau$. When the curvature is large, their procedure may yield overly conservative critical values. For example, consider testing $\theta_1\theta_2 = 0$. Around the point $(0,0)$, the curvature is infinite, and their corresponding critical value approaches $q_{\chi^2_2, 1-\alpha}$. However, as shown in Section 3.1 of this paper, a uniformly valid critical value in this setting is $q_{\chi^2_1, 1-\alpha}$.

4 Simulation

In this section, we examine the size and power properties of the proposed procedures and compare them with several alternatives. We focus on the context of Example X, namely the construction of confidence intervals for the mediation effect. In addition to the two MD-based methods proposed in Section ??, one using the critical value (BN1; see Section ??.3.1) and one using a bootstrapped critical value (BN2; Section ??.3.2), we consider two uniformly valid MD-based alternatives: (i) the procedure of Andrews and Mikusheva (AM). [1](ii) the MD-based method with projection critical value $Q(\chi^2_2, 1-\alpha)$. For comparison, we also include the naive Delta method, i.e. a Wald-type test with critical value $Q(\chi^2_2, 1-\alpha)$.

We study confidence intervals for $g(\theta) = \theta_1\theta_2$, where the estimators are simulated from

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

Without loss of generality, we normalize the variance of $\hat{\theta}$ to one. We consider $r = 0$ (as in the mediation example) and $r = 0.5$, with $\theta_2 \in \{2, 6\}$ and $\theta_1 = [-1 : 0.25 : 1] \times \theta_2$.

In Figure 3a, we plot the probability that the confidence intervals exclude the true value $g(\theta)$. The naive method has correct size when $\theta_1\theta_2 = 0$, but its rejection probability is very low near the origin, consistent with earlier findings (e.g. van~X; Dufour, Renault, and Zinde-Walsh, 2025). Away from the origin (see, e.g. $\theta_2 = 2$, $\theta_1 = 1.5$), the naive Wald test overrejects. According to Remark ??, BN1 is valid for $r = 0$. For $r = 0.5$, Theorem~1 only guarantees validity when $\theta_1 \geqslant 0.5$ under $\theta_2 = 6$, and gives no result for $\theta_2 = 2$. Nevertheless, BN1 maintains correct size across all designs, even when conditions fail, suggesting that the condition is sufficient but not necessary. As expected, all MD-based methods control size, with rejection rates below the nominal level.

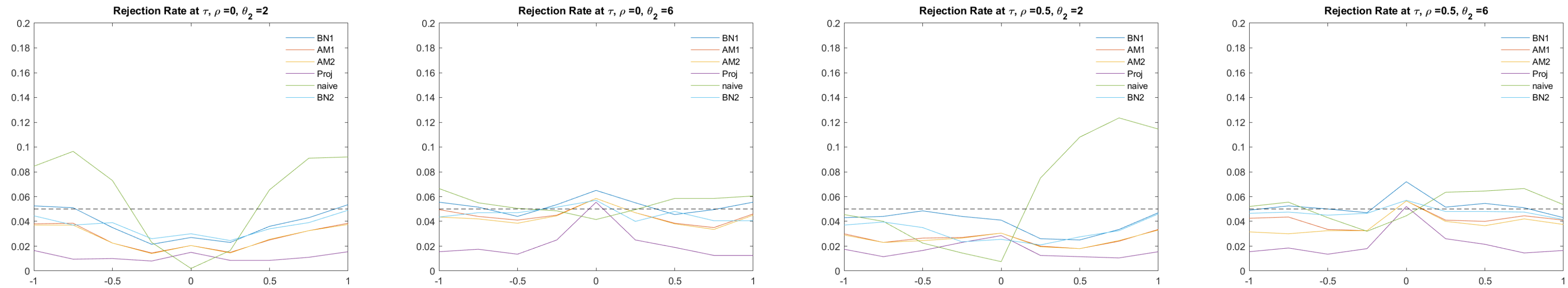
reports the probability that the confidence interval does not include zero, i.e. the probability that the CI gives a significant result. When θ is close to the origin, i.e. $\theta_2 = 2$, our methods have significantly higher power than AM, and the power for AM is close to the simple projection based method. When θ is far away from the origin, i.e. $\theta_2 = 6$, the power curves for different methods are indistinguishable to each other.

Figure 3b shows the probability that the confidence intervals exclude zero, i.e. the probability of a significant result. When θ is close to the origin ($\theta = 2$), our methods have substantially higher power than AM, whose performance is close to the simple projection method. When θ is further from the origin ($\theta_2 = 6$), power curves across methods are nearly identical. AM1, AM2 and BN2 have some asymmetry for $\rho = 0$, rej rate at 0, which is surprising to me. I will check the code.

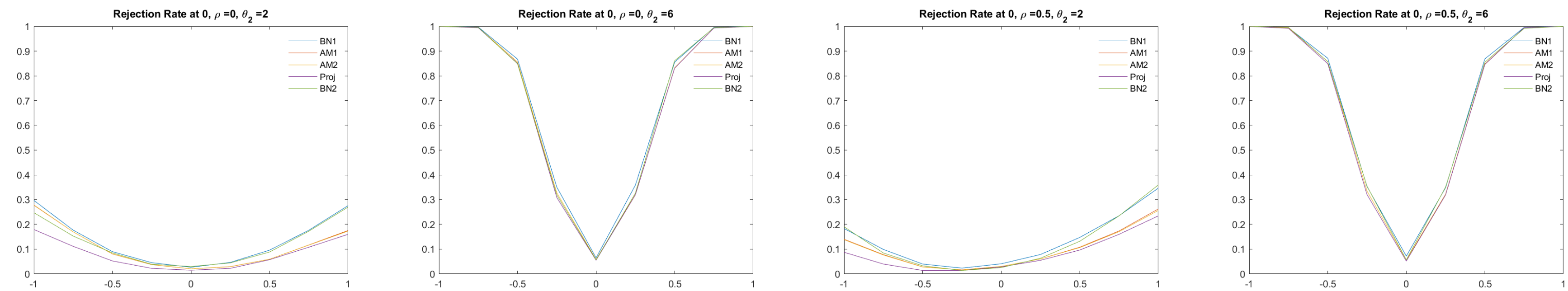
reports the median of the length of the confidence interval, where the median is taken over S samples. BN1 is the shortest in all designs, and BN2 is very close. Projection base method is the most conservative one, and it can be 19%~30% wider than BN1. AM is bwteen BN1 and the projection method, and the median CI can be 5% to 18% wider than BN1. The length difference is larger when θ is closer to the origin.

Finally, Figure 3c reports the median length of the confidence intervals, computed across S replications. BN1 consistently yields the shortest intervals, with BN2 close behind. The projection method is the most conservative, producing intervals 19–30% longer than BN1. AM lies between BN1 and the projection method, with median lengths 5–18% longer than BN1. The differences are most pronounced when θ is near the origin.

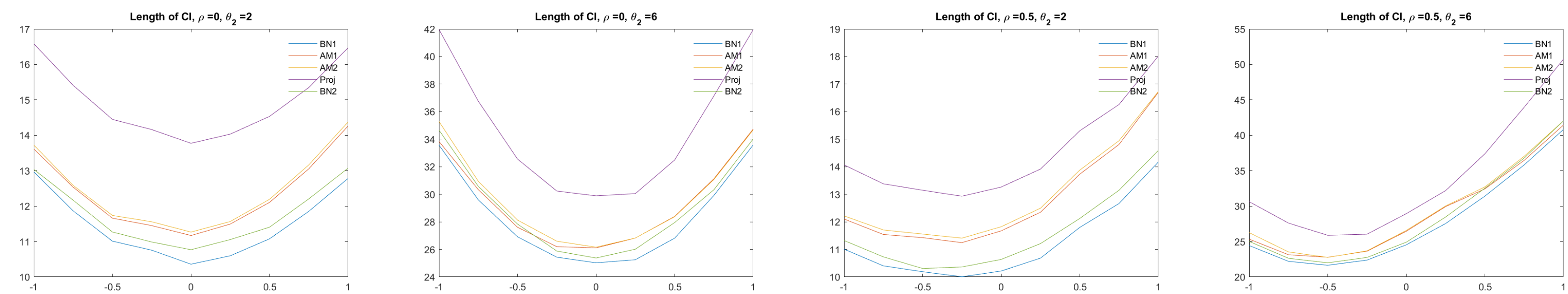
[1]We report results using their Section 4.1 implementation, which computes curvature over a restricted set with tuning parameter $\eta = \alpha/10$. We also implemented their worst-case curvature procedure from Section~2; the two produce nearly identical power, with the latter slightly worse.



(a) Rejection Rate at True Value. Left to Right: $(\theta_2, \rho) = (2, 0), (6, 0), (2, 0.5), (6, 0.5)$.



(b) Rejection Rate at Zero. Left to Right: $(\theta_2, \rho) = (2, 0), (6, 0), (2, 0.5), (6, 0.5)$.



(c) Median Length of the Confidence Interval. Left to Right: $(\theta_2, \rho) = (2, 0), (6, 0), (2, 0.5), (6, 0.5)$.

Figure 3: Simulation Results

The horizontal axis represents θ_1/θ_2. The first and second panels report the probability than the confidence intervals do not contain the true value $\theta_1\theta_2$ and zero, respectively. The third panel reports the median length of the CIs, where the median is taken over $S = 2,000$ samples.

5 Empirical Application

We illustrate the empirical relevance of our results using the setting analyzed by Alan et al. (2018). Their study takes advantage of a distinctive feature of the Turkish education system, in which elementary school teachers are randomly allocated across schools. This institutional detail generates plausibly exogenous variation in teacher characteristics that can be used to study how teachers' gender role attitudes influence student outcomes. The data include roughly 4,000 third- and fourth-grade students taught by 145 teachers, and 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. The treatment variable is whether a teacher is identified as holding traditional rather than progressive gender beliefs, while the mediator of interest is the student's own gender role beliefs. Following a similar analysis of this data in van Garderen and van Giersbergen (2024) we focus on verbal test scores as the outcome. Alan et al. (2018) argue that, after controlling for an extensive set of student, family, teacher, and school characteristics, the identifying assumptions for causal mediation analysis are satisfied in this context.

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 1: Estimates of Mediation Effects by Teacher Exposure

Table 1 reports estimates from Alan et al. (2018) analysis linking teachers' gender role attitudes to students' verbal test performance. The first coefficient, $\hat{\theta}_1$, comes from a regression of students' gender role beliefs on the gender role attitudes of their teachers, with the standard set of student, family, teacher, and school controls included. This coefficient summarizes the extent to which progressive teachers transmit their views to students. The second coefficient, $\hat{\theta}_2$, is estimated from a regression of test scores on both student gender beliefs and teacher attitudes, again with the full set of controls. It reflects how student beliefs are associated with verbal performance once teacher attitudes are held constant. Multiplying these two coefficients gives the mediated, or indirect, effect: the part of the teacher's influence on scores that operates through the channel of student beliefs. The estimates show that this indirect pathway is negative and relatively small, although it varies across exposure groups, being largest in the one-year sample and smallest for students exposed for four years.

Because the true mediation, or indirect, effect appears to be close to zero, Section 2 suggest that standard approaches to inference will fail. In particular, we cannot construct valid confidence intervals via the typical approach of inverting a t-test. Instead, we construct confidence intervals using the newly proposed methods of Section 3.

We compare our confidence intervals to two other inference procedures that might be applied in this setting, both of which are based on the minimum distance statistic. The first alternate procedure is that of Andrews and Mikusheva (2016), who propose a simulated critical value based on measuring the maximal curvature of the null manifold.[1] This testing procedure technically does not cover the case where we are testing the

[1] In implementing the test, we follow the empirical application in Andrews and Mikusheva (2012) and only calculate the maximum curvature over a set "close" to the point estimate, adjusting the critical value accordingly.

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

Table 2: Mediation Effect Point Estimates, 95% Confidence Intervals, and Confidence Interval Lengths in the data of Alan et al. (2018). Confidence Intervals are generated by inverting the corresponding tests. Values are rounded to three significant figures.

null that the mediation effect is equal to zero since the null manifold is not smooth in this case. However, the Andrews and Mikusheva (2016) critical value approaches $Q(\chi^2_2, 1-\alpha)$ from below as the null hypothesis value approaches zero and $Q(\chi^2_2, 1-\alpha)$ is a valid critical value for testing the null that the mediation effect is equal to zero so we simply modify the procedure slightly to directly use a $Q(\chi^2_2, 1-\alpha)$ when the null value is equal to zero. The second method, "Projection", simply uses the $Q(\chi^2_2, 1-\alpha)$ at all points, which is justified since, under the null hypothesis, the distance to the null manifold is always less than the distance to the point $(\theta_1, \theta_2)'$.

Consistent with the discussion in Section 3, the confidence intervals based on either the χ^2_1 critical value (BN1) or the two-step procedure (BN2) are uniformly tighter than those obtained from the Andrews and Mikusheva (2016) simulated critical value (AM); in all specifications, our intervals are strict subsets of theirs. The difference is not only theoretical but also empirically relevant. Using the χ^2_1 critical value, for instance, the confidence interval supports the conclusion of van Garderen and van Giersbergen (2024) that the mediation effect of a one-year exposure to a teacher with traditional views is negative, whereas the alternative methods cannot reject a null of zero at the five-percent level. As expected, the Andrews and Mikusheva (2016) intervals lie strictly inside those generated by the projection method, which uses a χ^2_2 critical value at all points. However, because the true mediation effect appears small in this setting, their simulated critical value converges toward $Q(\chi^2_1, 1-\alpha)$, which accounts for the close similarity between the two sets of intervals.

A Proofs and Supporting Results for Section 2

Proof of Proposition 1. Define $S_n = n(T_n - g(\theta_\star))$. Via a second order Taylor expansion we can see that

$$\begin{aligned} S_n &= n(T_n - g(\theta_\star)) \\ &= n(T_n - g(\theta_\star + h/\sqrt{n})) + n(g(\theta_\star + h/\sqrt{n} - g(\theta_\star)) \\ &\overset{h}{\rightsquigarrow} \mathcal{L}_h + \frac{1}{2}h'\nabla^2 g(\theta_\star)h \end{aligned}$$

where in the last line we use the fact that equation (7) holds for any $h \in \mathbb{R}^d$ by hypothesis. Since the experiment $\{P_\theta : \theta \in \Theta\}$ satisfies Assumption 1 with non-singular Fisher information $\Gamma_{\theta_\star}$, by Theorem 7.10 in van der Vaart (1998) there is a randomized statistic $\Psi(X, U)$ in the Gaussian shift experiment $\{N(h, \Gamma_{\theta_\star}^{-1}) : h \in \mathbb{R}^d\}$ such that $\Psi(X, U)$ has distribution $\mathcal{L}_h + \frac{1}{2}h'\nabla^2 g(\theta_\star)h$ when $X \sim N(h, \Gamma_{\theta_\star}^{-1})$. Equivalently, $\Psi(X, U) - \frac{1}{2}h'\nabla^2 g(\theta_\star)h \overset{h}{\sim} \mathcal{L}_h$. □

Proof of Proposition 2. (a) We proceed by contradiction, assuming there is an equivariant in law estimator. The characteristic function of the recentered estimator is given by

$$\psi(s) = \mathbb{E}_h[\exp(is(\Psi(Z, U) - h'Jh))] \tag{20}$$

where, by assumption, $\psi(s)$ does not depend on h. Let $\Phi_h(s) = \mathbb{E}_h[\exp(is\Psi(Z, U))]$ and notice that (20) implies that we can decompose $\psi(s)\exp(isf(h)) = \Phi_h(s)$ where we let $f(h) = h'Jh$ to save notation. We start by showing that $\Phi_h(s)$ is twice continuously differentiable in h and deriving expressions for the derivatives.

For the first derivative, consider a point $h_0 \in \mathbb{R}^d$ and a deviation in the direction h of size r. We save notation by letting $\Gamma = \Gamma_{\theta_\star}$ and justify bringing the limit inside the integral by the uniform integrability condition of Hirano and Porter (2012), Lemma 1(b).

$$\begin{aligned} \lim_{r\downarrow 0}\frac{1}{r}\left[\Phi_{h_0+rh}(s) - \Phi_{h_0}(s)\right] &= \lim_{r\downarrow 0}\frac{1}{r}\left[\int_{[0,1]}\int \exp(is\Psi(z,u))\{\phi(z|h_0 + rh, \Gamma^{-1}) - \phi(z|h_0, \Gamma^{-1}\}\,dz\,du\right] \\ &= \int_{[0,1]}\int \exp(is\Psi(z,u))\lim_{r\downarrow 0}\{\phi(z|h_0 + rh, \Gamma^{-1}) - \phi(z|h_0, \Gamma^{-1})\}\,dz\,du \\ &= \int_{[0,1]}\int \exp(is\Psi(z,u))(z - h_0)'\Gamma h\phi(z|h_0, \Gamma^{-1})\,dz\,du \\ &= \mathbb{E}_{h_0}[\exp(is\Psi(Z,U))(Z - h_0)']\Gamma h \end{aligned}$$

Since h is arbitrary here, we can rewrite the above as

$$\nabla\Phi_{h_0}(s) = \mathbb{E}_{h_0}[\exp(is\Psi(Z,U))(Z - h_0)']\Gamma$$

where the gradient is understood to be with respect to the argument h_0, i.e s is kept fixed. For the second derivative, we repeat the argument, again letting h be an arbitrary direction in $\mathbb{R}^d$ and justifying bringing the limit into the integral via Hirano and Porter (2012), Lemma 1(b) along with the fact that

$\mathbb{E}_{h_0}[\|\exp(is\Psi)(Z-h_0)\|]$ is uniformly bounded over h_0:

$$\begin{aligned}
&\lim_{r\downarrow 0}\frac{1}{r}\big[\nabla\Phi_{h_0+rh}(s)-\nabla\Phi_{h_0}(s)\big]\\
&\quad=\lim_{r\downarrow 0}\frac{1}{r}\Big\{\int_{[0,1]}\int \exp(is\Psi(z,u))(z-h_0)'\Gamma\{\phi(z|h_0+rh,\Gamma^{-1})-\phi(z|h_0,\Gamma^{-1})\}\,dz\,du\\
&\qquad\quad-\int_{[0,1]}\int \exp(is\Psi(z,u))rh'\Gamma\phi(z|h_0+rh,\Gamma^{-1})\,dz\,du\Big\}\\
&\quad=\int_{[0,1]}\int \exp(is\Psi(z,u))(z-h_0)'\lim_{r\downarrow 0}\frac{1}{r}\Gamma\{\phi(z|h_0+rh,\Gamma^{-1})-\phi(z|h_0,\Gamma^{-1})\}\,dz\,du\\
&\qquad\quad-\int_{[0,1]}\int \exp(is\Psi(z,u))h'\Gamma\lim_{r\downarrow 0}\phi(z|h_0+rh,\Gamma^{-1})\,dz\,du\Big\}\\
&\quad=h'\Gamma\int_{[0,1]}\int \exp(is\Psi(z,u))(z-h_0)(z-h_0)'\,\phi(z|h_0,\Gamma^{-1})\,dz\,du\Gamma\\
&\qquad-h'\Gamma\int_{[0,1]}\int \exp(is\Psi(z,u))\phi(z|h_0,\Gamma^{-1})\,dz\,du\\
&\quad=h'\Gamma\mathbb{E}_{h_0}[\exp(is\Psi(Z,U))(Z-h_0)(Z-h_0)']\Gamma-h'\Gamma\mathbb{E}_{h_0}[\exp(is\Psi(Z,U))]
\end{aligned}$$

Again, since h is arbitrary we can write this

$$\nabla^2\Phi_{h_0}(s)=\Gamma\mathbb{E}_{h_0}[\exp(is\Psi(Z,U))(Z-h_0)(Z-h_0)']\Gamma-\Phi_{h_0}(s)\Gamma \tag{21}$$

The first and second derivatives of $\exp(isf(h_0))$ with respect to h_0 can be expressed

$$\begin{aligned}
\nabla\exp(isf(h_0))&=2is\exp\big(isf(h_0)\big)\,Jh_0\\
\nabla^2\exp\big(isf(h_0)\big)&=\exp\big(isf(h_0)\big)\Big(2isJ-4s^2(Jh_0)(Jh_0)'\Big)
\end{aligned} \tag{22}$$

Recall that, by assumption, $\Phi_{h_0}(s)=\psi(s)\exp(isf(h_0))$ for all h_0. Pick an $s\neq 0$ such that $\psi(s)\neq 0$. This is possible since $\psi(0)=1$ and $\psi(\cdot)$ is continuous. Combining (21) and (22) yields, for any h_0, that

$$\psi(s)\exp\big(isf(h_0)\big)\Big(2isJ-4s^2(Jh_0)(Jh_0)'\Big)=\Gamma\mathbb{E}_{h_0}[\exp(is\Psi(Z,U))(Z-h_0)(Z-h_0)']\Gamma-\Phi_{h_0}(s)\Gamma \tag{23}$$

Notice that since $|\exp(is\Psi(z,u))|=1$ and $|\Phi_{h_0}(s)|\leqslant 1$ for all h_0, the operator norm of the RHS of (23) is bounded uniformly over $h_0\in\mathbb{R}^d$. On the other hand, looking at the LHS of (23) we can see, using $\|A+B\|\geqslant\|B\|-\|A\|$, that

$$\|\text{LHS}\|\geqslant|\psi(s)|\Big(4s^2\|Jh_0\|^2-2|s|\|J\|\Big)$$

Let v be such that $\|Jv\|\neq 0$ and let $h_0=cv$ for some $c>0$ so that $\|Jh_0\|^2=c^2\|Jv\|^2$. By sending $c\to\infty$ we can thus make $\|\text{LHS}\|$ arbitrarily large, leading to a contradiction since $\|\text{RHS}\|$ is uniformly bounded over $h_0\in\mathbb{R}^d$.

(b) Let h be such that $h'Jh\neq 0$. Since J is assumed symmetric and non-zero, it is guaranteed that such an h exists. For any $r\geqslant 0$ we have that

$$\alpha=\mathbb{P}_{(1+r)h}(\Psi(Z,U)\leqslant((1+r)h)'\boldsymbol{J}((1+r)h))$$

In particular

$$0 = \alpha - \alpha = \mathbb{P}_{(1+r)h}(T \leqslant ((1+r)h)'\boldsymbol{J}((1+r)h)) - \mathbb{P}_h(T \leqslant h'\boldsymbol{J}h)$$

and thus

$$\begin{aligned} 0 = \lim_{r \downarrow 0} \Bigg\{ \frac{1}{r} \Big[&\mathbb{P}_{(1+r)h}(T \leqslant (rh)'\boldsymbol{J}(rh)) - \mathbb{P}_h\left(T \leqslant ((1+r)h)'\boldsymbol{J}((1+r)h)\right) \Big] \\ &+ \frac{1}{r}\Big[\mathbb{P}_h\left(T \leqslant ((1+r)h)'\boldsymbol{J}((1+r)h)\right) - \mathbb{P}_h\left(T \leqslant h'\boldsymbol{J}h\right)\Big] \Bigg\} \end{aligned} \tag{24}$$

Applying the uniform integrability in Lemma 1(a) of Hirano and Porter (2012) to justify exchanging limits and integrals as in the proof of Lemma 2, we obtain for any h

$$h'\Gamma_{\theta_\star}\mathbb{E}_h[\mathbf{1}\{\Psi(X,U) \leqslant h'\boldsymbol{J}h\}(X-h)] = 2(h'\boldsymbol{J}h)F_h'(h'\boldsymbol{J}h)$$

From here, take a constant $c > 0$ and consider the behavior of the LHS and RHS as $c \to \infty$. Notice that for any $c > 0$ $\|ch'\Gamma_{\theta_\star}\| \lesssim c$ while $\|\mathbb{E}_h[\mathbf{1}\{\Psi(X,U) \leqslant h'\boldsymbol{J}h\}(X-h)]\| \lesssim 1$ by Cauchy-Schwarz. Meanwhile, $2((ch)'\boldsymbol{J}(ch)) \propto c^2 F_{ch}'((ch)'\boldsymbol{J}(ch))$. Since $c^2 F_{ch}'((ch)'\boldsymbol{J}(ch)) \to \infty$ as $c \to \infty$ we arrive at a contradiction. $\square$

Proof of Theorem 1. Theorem 1 follows directly from Proposition 1 along with Proposition 2. $\square$

Proof of Proposition 3. From the proof of Proposition 1 we see that

$$n(T_n - g(\theta_\star)) \overset{h}{\rightsquigarrow} \mathcal{L}_h + \frac{1}{2}h'\nabla^2 g(\theta_\star)h$$

Thus, under the sequence of alternatives $P_{\theta_\star + h/\sqrt{n}}$ the asymptotic power of the test that rejects if $n(T_n - g(\theta_\star)) \leqslant 0$ is $\mathcal{L}_h(\{-\infty, -\frac{1}{2}h'\nabla^2 g(\theta_\star)h\})$. Let $\mathcal{P}(h) = \mathcal{L}_h(\{-\infty, -\frac{1}{2}h'\nabla^2 g(\theta_\star)h\}$. Via Proposition 1, there is a randomized statistic T in the Gaussian shift experiment $\{N(h, \Gamma_{\theta_\star}^{-1}) : h \in \mathbb{R}^d\}$ such that $T - \frac{1}{2}h'\nabla^2 g(\theta_\star)h \overset{h}{\sim} \mathcal{L}_h$. Notice that, in the limiting Gaussian shift experiment, we have that for any h

$$\begin{aligned} \mathbb{P}_h\left(\Psi(Z,U) \leqslant 0\right) &= \mathbb{P}_h\left(\Psi(Z,U) - \frac{1}{2}h'\nabla^2 g(\theta_\star)h \leqslant -\frac{1}{2}h'\nabla^2 g(\theta_\star)h\right) \\ &= \mathcal{L}_h(\{-\infty, -\frac{1}{2}h'\nabla^2 g(\theta_\star)h\}) \\ &= \mathcal{P}(h) \end{aligned} \tag{25}$$

Thus, in the limiting experiment we wish to show that $\nabla_h \Pr_h(\Psi(Z,U) \leqslant 0)|_{h=0} = 0$. This is obtained in Lemma 3 under the maintained assumption that Ψ is α-quantile unbiased. $\square$

Proof of Corollary 1. Follows directly from Theorem 1. $\square$

Lemma 2. *Suppose that $\Psi(Z,U)$ is a statistic in the Gaussian shift experiment $\{N(h, \Gamma_{\theta_\star}^{-1}) : h \in \mathbb{R}^d\}$ and let $\mathcal{H} \subseteq \mathbb{R}^d$ be a cone such that, for some $\alpha \in (0,1)$,*

$$\alpha = \mathbb{P}_h\left(\Psi \leqslant h'\nabla^2 g(\theta_\star)h\right), \qquad \text{for all } h \in \mathcal{H}$$

Let $F_\Psi(\cdot)$ denote the CDF of $\Psi(Z,U)$ under $h = 0$. Assume that the derivative of F_Ψ exists at zero. Then $h'\Gamma_{\theta_\star}\mathbb{E}_0[\mathbf{1}\{\Psi(Z,U) \leqslant 0\}Z] = 0$ for all $h \in \mathcal{H}$.

Proof of Lemma 2. The proof of the following lemma closely follows that of Proposition 1(c) in Hirano and Porter (2012). To simplify notation, let $\boldsymbol{J} = \frac{1}{2}\nabla^2 g(\theta_\star)$. For any $r \geqslant 0$ we have that

$$\alpha = \mathbb{P}_{rh}\left(T \leqslant (rh)'\boldsymbol{J}(rh)\right)$$

Evaluating the above expression at $r > 0$ and $r = 0$ yields

$$0 = \alpha - \alpha = \mathbb{P}_{rh}\left(T \leqslant (rh)'\boldsymbol{J}(rh)\right) - \mathbb{P}_0(T \leqslant 0)$$

and thus

$$\begin{aligned} 0 = \lim_{r\downarrow 0}\bigg\{ &\frac{1}{r}\Big[\mathbb{P}_{rh}(T \leqslant (rh)'\boldsymbol{J}(rh)) - \mathbb{P}_0\left(T \leqslant (rh)'\boldsymbol{J}(rh)\right)\Big] \\ &+ \frac{1}{r}\Big[\mathbb{P}_0\left(T \leqslant (rh)'\boldsymbol{J}(rh)\right) - \mathbb{P}_0\left(T \leqslant 0\right)\Big]\bigg\} \end{aligned} \tag{26}$$

Each of the terms on the RHS of (26) exist, so we can write the limit of the sum as the sum of the limits. Let $\phi(\cdot|\mu,\Sigma)$ denote the pdf of a normal distribution with mean μ and variance Σ. Consider the first term. Applying the uniform integrability condition in Lemma 1(a) of Hirano and Porter (2012) to justify interchanging limits and integrals below, we obtain

$$\begin{aligned} &\lim_{r\downarrow 0}\frac{1}{r}\big[\mathbb{P}_{rh}(T \leqslant (rh)'\boldsymbol{J}(rh)) - \mathbb{P}_0(T \leqslant (rh)'\boldsymbol{J}(rh))\big] \\ &\quad = \lim_{r\downarrow 0}\int_{[0,1]}\int \mathbf{1}\left\{\Psi(z,u) \leqslant (rh)'\boldsymbol{J}(rh)\right\} \times \frac{1}{r}[\phi(z|rh,\Gamma_{\theta_\star}^{-1}) - \phi(z|0,\Gamma_{\theta_\star}^{-1})]\,dz\,du \\ &\quad = \int_{[0,1]}\int \lim_{r\downarrow 0}\mathbf{1}\left\{\Psi(z,u) \leqslant (rh)'\boldsymbol{J}(rh)\right\} \times \frac{1}{r}[\phi(z|rh,\Gamma_{\theta_\star}^{-1}) - \phi(z|0,\Gamma_{\theta_\star}^{-1})]\,dz\,du \\ &\quad = \int_{[0,1]}\int \mathbf{1}\{\Psi(z,u) \leqslant 0\}\left(\frac{\partial}{\partial\tilde{h}}\phi(z|\tilde{h},\Gamma_{\theta_\star}^{-1}\right)_{\tilde{h}=0} h\,dz\,du \\ &\quad = h'\Gamma_{\theta_\star}\left\{\int_{[0,1]}\int \mathbf{1}\{\Psi(z,u) \leqslant 0\}z\phi(z|0,\Gamma_{\theta_\star}^{-1})\,dz\,du\right\} \end{aligned}$$

Since the since the derivative of $F_\Psi(\cdot)$ at zero exists and $\frac{\partial}{\partial r}(rh)'\boldsymbol{J}(rh)\big|_{r=0} = 0$, the second term on the RHS of (26) evaluates to zero. Thus, we obtain for any $h \neq 0$ that

$$0 = h'\Gamma_{\theta_\star}\left\{\int_{[0,1]}\int \mathbf{1}\{\Psi(z,u) \leqslant 0\}(\Gamma_{\theta_\star}^{-1})z\phi(z|0,\Gamma_{\theta_\star}^{-1})\,dz\,du\right\}$$

which gives the result □

Lemma 3. *Let $\Psi(Z,U)$ be a statistic in the Gaussian shift experiment $\{N(h,\Gamma_{\theta_\star}^{-1}), h \in \mathbb{R}^d\}$ such that for (i) for some $\alpha \in (0,1)$ and cone $\mathcal{H} \subset \mathbb{R}^d$,*

$$\alpha = \mathbb{P}_h\left(\Psi \leqslant \frac{1}{2}h'\nabla^2 g(\theta_\star)h\right) \quad \textit{for all } h \in \mathcal{H},$$

and (ii) the CDF of $\Psi(Z,U)$ under $h = 0$, $F_\Psi(\cdot)$ is differentiable at zero. Consider the level α test based on T, that is the test that rejects if $\{\Psi(Z,U) \leqslant 0\}$. Define $\mathcal{P}(h) = \Pr_h(\Psi(Z,U) \leqslant 0)$ the power curve for this *test. This power curve is flat around zero in the direction h in the sense that $D_h\mathcal{P}(0)$ exists and is equal to zero.*

Proof of Lemma 3. Consider a deviation in the direction h. Define $\mathcal{P}_h(r) = \Pr_{rh}(\Psi(Z,U) \leqslant 0)$. We want to show that

$$\frac{\partial}{\partial r}\mathcal{P}_h(r)\big|_{r=0} = \lim_{r\downarrow 0} \frac{\mathbb{P}_{rh}(\Psi(Z,U) \leqslant 0) - \mathbb{P}_0(\Psi(Z,U) \leqslant 0)}{r} = 0$$

Let us expand the above limit and, as in the proof of Lemma 2, invoke Lemma 1(a) in Hirano and Porter (2012) to justify exchanging a limit and an integral below.

$$\begin{aligned}
\frac{\partial}{\partial r}\mathcal{P}_h(r)\big|_{r=0} &= \lim_{r\downarrow 0}\int_{[0,1]}\int \mathbf{1}\{\Psi(z,u) \leqslant 0\} \times \frac{1}{r}[\phi(z|rh,\Gamma_{\theta_\star}^{-1}) - \phi(z|0,\Gamma_{\theta_\star}^{-1})]\, dz\, du \\
&= \int_{[0,1]}\int \mathbf{1}\{\Psi(z,u) \leqslant 0\} \times \lim_{r\downarrow 0}\frac{1}{r}[\phi(z|rh,\Gamma_{\theta_\star}^{-1}) - \phi(z|0,\Gamma_{\theta_\star}^{-1})]\, dz\, du \\
&= h'\Gamma_{\theta_\star}\int_{[0,1]}\int \mathbf{1}\{\Psi(z,u) \leqslant 0\} z\phi(z|0,\Gamma_{\theta_\star}^{-1})\, dz\, du \\
&= h'\Gamma_{\theta_\star}\mathbb{E}_0[\mathbf{1}\{\Psi(Z,U) \leqslant 0\}Z] \\
&= 0
\end{aligned}$$

where the final equality comes from Lemma 2. □

Remark 7. The proof of Lemma 3 could be obtained almost directly from the proof of Lemma 2. However, the statement of Lemma 2 additionally implies that $\mathrm{Cov}_0(\mathbf{1}\{T \leqslant 0\}, Z) = 0$, which is also an interesting restriction on any α-quantile unbiased estimate. □

B Supporting Results for ??

Lemma 4. *The boundary $\partial\mathcal{S}^+(\tau)$ can be characterized by two curves*

$$\mathcal{C}_1(\tau) = \begin{cases} \left\{\left(X_1(x), X_1(\frac{\tau}{x})\right) : x \in \mathbb{R}_{++}\right\} & \textit{if } \tau > \frac{c^2}{2} \\ \left\{\left(X_1(x), X_1(\frac{\tau}{x})\right) : x \in (0, x_1^*] \cup [x_2^*, +\infty)\right\} & \textit{if } 0 < \tau \leqslant \frac{c^2}{2} \end{cases}$$

$$\mathcal{C}_2(\tau) = \left\{\left(X_2(x), X_2(\frac{\tau}{x})\right) : x \in \mathbb{R}_{++}\right\},$$

where

$$X_1(x) = x + \frac{c\tau}{\sqrt{\tau^2 + x^4}},\ X_2(x) = x - \frac{c\tau}{\sqrt{\tau^2 + x^4}}.$$

$$x_1^* = \sqrt{\frac{c^2}{2} - \sqrt{\frac{c^4}{4} - \tau^2}},\ x_2^* = \sqrt{\frac{c^2}{2} + \sqrt{\frac{c^4}{4} - \tau^2}}. \tag{27}$$

Similarly let $\mathcal{C}_3$ and $\mathcal{C}_4$ be the boundary of $\mathcal{S}^-(\tau)$ where

$$\begin{aligned}
\mathcal{S}^-(\tau) &= \left\{(X,Y) : \inf_{xy=\tau, x<0} (X-x)^2 + (Y-y)^2 \leqslant c^2\right\} \\
\mathcal{C}_3(\tau) &= \left\{\left(-X_1(x), -X_1(\frac{\tau}{x})\right) : x \in (0, x_1^*] \cup [x_2^*, +\infty)\right\} \\
\mathcal{C}_4(\tau) &= \left\{\left(-X_2(x), -X_2(\frac{\tau}{x})\right) : x \in \mathbb{R}_{++}\right\}.
\end{aligned}$$

Proof. $\mathcal{C}_1(\tau)$ and $\mathcal{C}_2(\tau)$ are obtained by shifting $\mathcal{S}_0^+(\tau)$ a distance c along its normal direction. $\mathcal{C}_1$ has two expressions. Note that with $\tau > \frac{c^2}{2}$

$$\frac{\mathrm{d}X_1(x)}{\mathrm{d}x} = 1 - \frac{2c}{\tau^2}\left(\frac{x^2}{\tau^2} + \frac{1}{x^2}\right)^{-3/2} \geqslant 1 - \frac{2c}{\tau^2}\left(\frac{2}{\tau}\right)^{-3/2} \geqslant 0.$$

With $\tau \in (0, \frac{c^2}{2}]$, for all $x \in (x_1^*, x_2^*)$, $\left(X_1(x), X_1(\frac{\tau}{x})\right)$ is an interior point of $\mathcal{S}^+(\tau)$, thus not included in $\mathcal{C}_1(\tau)$. Note that $x_1^* \leqslant \sqrt{\tau} \leqslant x_2^*$, thus for $x \in (0, x_1^*) \cup (x_2^*, +\infty)$

$$\begin{aligned}
\frac{\mathrm{d}X_1(x)}{\mathrm{d}x} &= 1 - \frac{2c}{\tau^2}\left(\frac{x^2}{\tau^2} + \frac{1}{x^2}\right)^{-3/2} \\
&\geqslant 1 - \frac{2c}{\tau^2}\left(\min\left\{\frac{x_1^{*2}}{\tau^2} + \frac{1}{x_1^{*2}}, \frac{x_2^{*2}}{\tau^2} + \frac{1}{x_2^{*2}}\right\}\right)^{-3/2} = 1 - \frac{2c}{\tau^2}\left(\frac{c^2}{\tau^2}\right)^{-3/2} \geqslant 0.
\end{aligned}$$

In addition $\frac{\mathrm{d}X_2(x)}{\mathrm{d}x} \geqslant 0$ for all $\tau \geqslant 0$. □

Proof of Theorem 4.

Proof. This follows from Proposition 5, 6 and 7. □

Proposition 5. *For all $(x_1, x_2) \in \mathcal{S}_0(0)$,*

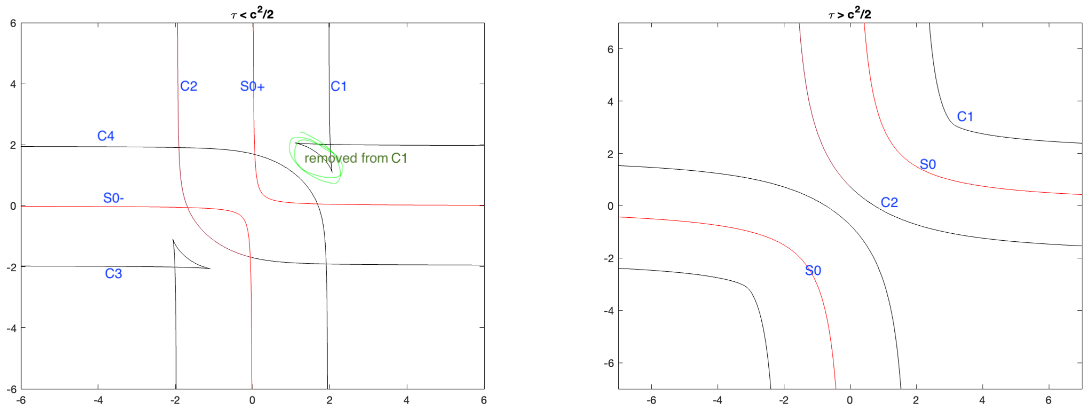
$$P\left((Z_1 + x_1, Z_2 + x_2) \in \mathcal{S}(0)\right) \geqslant 1 - \alpha.$$

Proof. WLOG, assume $x = 0$. Then

$$\begin{aligned}
P\left((Z_x + x, Z_y + y) \in \mathcal{S}(0)\right) &= P\left(\inf_{\tilde{x}\tilde{y}=0} (Z_x + x - \tilde{x})^2 + (Z_y + y - \tilde{y})^2 \leqslant c^2\right) \\
&\geqslant P\left((Z_x + x - \tilde{x})^2 + (Z_y + y - \tilde{y})^2 \leqslant c^2 \text{ where } \tilde{x} = 0, \tilde{y} = Z_y + y\right) \\
&= P\left(Z_x^2 \leqslant c^2\right) = 1 - \alpha.
\end{aligned}$$

□

Figure 4: $\mathcal{S}(\tau)$ and $\mathcal{C}(\tau)$



Proposition 6. *Let* $|\tau| > \frac{c^2}{2}$. *For all* $(\vartheta_1, \vartheta_2) \in \mathcal{S}_0^+(\tau) = \{(x_1, x_2) : x_1x_2 = \tau, x_1 > 0\}$,

$$P\left((Z_1 + \vartheta_1, Z_2 + \vartheta_2) \in \mathcal{S}^+(\tau)\right) \geqslant 1 - \alpha.$$

Proof. Fix $(x_1, x_2) \in \mathcal{S}_0^+(\tau)$. Condition 1 of Lemma 1 holds trivially. We now verify Condition 2 of Lemma 1. Let $r > c$. By Lemma 5.1, $\partial B\left((x_1, \frac{\tau}{x_1}), r\right)$ intersects $\mathcal{C}_1(\tau)$ at (and only at) A and B, with A to the left of B. By Lemma 5.4, $\partial B\left((x_1, \frac{\tau}{x_1}), r\right)$ intersects $\mathcal{C}_2(\tau)$ at (and only at) C and D, with C to the left of D. By Lemma 5.1 with $c = 0$, $\partial B\left((x_1, \frac{\tau}{x_1}), r\right)$ intersects $\mathcal{S}_0^+(\tau)$ at (and only at) E and F, with E on the left of F. Since $\mathcal{S}_0^+(\tau)$ separates $\mathcal{C}_1$ and $\mathcal{C}_2$, the clockwise order of these six points on $\partial B\left((x, \frac{\tau}{x}), r\right)$ is $BFDCEA$. Since $E, F \in \mathcal{S}^+(\tau)$, we have $\widehat{AEC} \subset \mathcal{S}^+(\tau)$ and $\widehat{BFD} \subset \mathcal{S}^+(\tau)$. See Figure ??.

To show that length $\left(\widehat{AEC}\right)$ + length $\left(\widehat{BFD}\right) \geqslant 4r \arcsin \frac{c}{r}$, it suffices to show that

$$\text{length}\left(\overline{AC}\right) \geqslant 2c \text{ and length}\left(\overline{BD}\right) \geqslant 2c.$$

By contradiction, assume that length $\left(\overline{AC}\right) < 2c$. Let AC intersects $\mathcal{S}_0^+(\tau)$ at point G, then $B(G, c) \nsubseteq \mathcal{S}^+(\tau)$, which contradicts the definition of $\mathcal{S}^+(\tau)$. Therefore, Condition 1 and 2 of Lemma 1 hold, and

$$P\left((Z_1 + x_1, Z_2 + x_2) \in \mathcal{S}^+(\tau)\right) \geqslant 1 - \alpha. \tag{28}$$

This completes the proof. □

Proposition 7. *Let* $0 < |\tau| \leqslant \frac{c^2}{2}$. *The standard GMM is valid, i.e. for all* $(x_1, x_2) \in \mathcal{S}_0(\tau)$,

$$P\left((Z_1 + x_1, Z_2 + x_2) \in \mathcal{S}(\tau)\right) \geqslant 1 - \alpha.$$

Proof. By Lemma 5.2, for all $x \leqslant x_1^*$ or $x \geqslant x_2^*$, the coverage is at least $1 - \alpha$ using the same argument as in Proposition 6.

For $x \in (x_1^*, x_2^*)$, we check Condition 2 of Lemma 1. Let the kink of $\mathcal{C}_1$, i.e. $\mathcal{C}_1 \cap \{(x,x) : x \in \mathbb{R}\}$, be

$$H = \left(X_1(x_1^*), X_1(\frac{\tau}{x_1^*})\right) = \left(X_1(x_2^*), X_1(\frac{\tau}{x_2^*})\right). \tag{29}$$

Let $r(x_1)$ denote the distance between $O = (x_1, \tau/x_1)$ and H. If $r > r(x)$, $\partial B_r(O, r)$ intersects with $\mathcal{C}_1$ at exactly two points, therefore

$$\left|\text{length}\left(\partial B_r(O,r) \cap \bar{\mathcal{S}}\right)\right| \geqslant 4r \arcsin\frac{c}{r}$$

following from the same argument in Proposition 6.

Then we show that for $r \in (c, r(x)]$,

$$\left|\text{length}\left(\partial B_r\left((x_1, \tau/x_1), r\right) \cap \bar{\mathcal{S}}\right)\right| = 2\pi r \geqslant 4r \arcsin\frac{c}{r}. \tag{30}$$

Let I and J denote $\mathcal{C}_2 \cap \mathcal{C}_4$, i.e.

$$I = \left(X_2(x_1^*), X_2(\frac{\tau}{x_1^*})\right),\ J = \left(X_2(x_2^*), X_2(\frac{\tau}{x_2^*})\right).$$

By Lemma 6, the perpendicular bisector of IH is tangent of $\mathcal{S}_0^+(\tau)$ at $(x_1^*, \frac{\tau}{x_1^*})$. Since $\mathcal{S}_0$ is concave, $O = (x, \frac{\tau}{x})$ is on the right hand side of the perpendicular bisector of $\overline{IH}$, which implies that $\overline{IO} \geqslant \overline{HO} = r(x) \geqslant r$. Similarly $JO \geqslant r$. Let $O' = \left(X_2(x), X_2(\frac{\tau}{x})\right)$, and $\overline{O'O} = c < r$. By the continuity of distance, $\partial B(O,r)$ intersects $\mathcal{C}_2$ at point I' on curve $O'I$ and at point J' on curve $\overline{O'J}$. By Lemma 5.4, $\partial B(O,r) \cap \mathcal{C}_2 = \{I', J'\}$. Lastly, we show that $\overset{\frown}{I'J'} \subseteq \mathcal{S}^-(\tau)$. It suffices to show that $\partial B(O,r) \cap \mathcal{C}_3 = \varnothing$. By contradiction, if $K' = \left(-X_1(x'), -X_1(\frac{\tau}{x'})\right) \in \partial B(O,r) \cap \mathcal{C}_3$ for some $x' \in \mathbb{R}_{++}$, then we have $|\overline{OK}| < |\overline{OK'}| = r$ where $K = \left(X_1(x'), X_1(\frac{\tau}{x'})\right)$. However, by Lemma 5.3, we have $r \leqslant r(x) = \overline{OH} \leqslant \overline{OK}$, which is a contradiction. ☐

Lemma 5. *Let $r > c$.*

1. *For all $\tau \geqslant \frac{c^2}{2}$, for all $x_0 > 0$, $\mathcal{C}_1$ intersects $\partial B\left((x_0, \frac{\tau}{x_0}), r\right)$ at two and only at two points.*
2. *For all $0 < \tau \leqslant \frac{c^2}{2}$, for all $x_0 \in (0, x_1^*] \cup [x_2^*, +\infty)$, $\mathcal{C}_1$ intersects $\partial B\left((x_0, \frac{\tau}{x_0}), r\right)$ at two and only at two points.*
3. *For all $0 < \tau \leqslant \frac{c^2}{2}$, for all $x_0 \in (x_1^*, x_2^*)$, $d\left((x_0, \frac{\tau}{x_0}), \mathcal{C}_1\right) = d\left((x_0, \frac{\tau}{x_0}), H\right)$, where H is defined in (29).*
4. *For all $\tau > 0$, for all $x_0 > 0$, $\mathcal{C}_2$ intersects $\partial B\left((x_0, \frac{\tau}{x_0}), r\right)$ at two and only at two points.*

Proof. Part I $\mathcal{C}_1$. Let $h_1(x)$ being the distance between $(x_0, \frac{\tau}{x_0})$ and $(X_1(x), X_1(\frac{\tau}{x})) \in \mathcal{C}_1$, i.e.

$$h_1(x) = \left(x + c\tau(\tau^2 + x^4)^{-1/2} - x_0\right)^2 + \left(\frac{\tau}{x} + cx^2(\tau^2 + x^4)^{-1/2} - \frac{\tau}{x_0}\right)^2.$$

It's easy to see that

$$h_1(0_+) = \infty,\ h_1(+\infty) = \infty,\ h_1(x_0) = c^2 < r^2. \tag{31}$$

The derivative has form

$$h_1'(x) = \frac{2\left(\tau^2 + x_0 x^3\right)}{x_0\left(\tau^2 + x^4\right)^{3/2}}(x - x_0)\left(\left(\frac{\tau^2}{x^2} + x^2\right)^{\frac{3}{2}} - 2c\tau\right). \tag{32}$$

To show 1, by the intermediate value theorem and (31), we have at least $x_1 \in (0, x_0)$, $x_2 \in (x_0, \infty)$ such that $h_1(x_1) = h_1(x_2) = r^2$. In addition, with $\tau \geqslant \frac{c^2}{2}$,

$$\left(\frac{\tau^2}{x^2} + x^2\right)^{\frac{3}{2}} - 2c\tau \geqslant (2\tau)^{\frac{3}{2}} - 2c\tau = 2\tau\left(\sqrt{2\tau} - c\right) \geqslant 0. \tag{33}$$

Therefore, $h_1'(x) > 0$ for $x > x_0$ and $h_1'(x) < 0$ for $x < x_0$. By the monotonicity, the solution $h_1(x) = r^2$ is unique for $x \in (0, x_0)$ and $x \in (x_0, +\infty)$.

To show 2. WLOG, let $x_0 \in (0, x_1^*]$. Note that

$$\left(\frac{\tau^2}{x_1^{*2}} + x_1^{*2}\right)^{\frac{3}{2}} - 2c\tau = \left(\frac{\tau^2}{x_2^{*2}} + x_2^{*2}\right)^{\frac{3}{2}} - 2c\tau = c(c^2 - 2\tau) \geqslant 0, \tag{34}$$

thus for all $x \in (0, x_0)$, $h_1'(x) \leqslant 0$; $x \in (x_0, x_1^*)$, $h_1'(x) > 0$; $x \in (x_2^*, +\infty)$, $h_1'(x) \geqslant 0$. Together with $h_1(x_1^*) = h_1(x_2^*)$, we have that

$$h_1(x) = r^2,\ x \in (0, x_1^*] \cup [x_2^*, +\infty)$$

has exactly two solution, one in $(0, x_0)$ and one in $(x_0, x_1^*) \cup (x_2^*, +\infty)$.

To show 3. By (32) and (34), $h_1'(x) < 0$ for $x \in (0, x_1^*) \subseteq (0, x_0)$, and $h_1'(x) > 0$ for $x \in (x_2^*, \infty) \subseteq (x_0, +\infty)$. Therefore,

$$d\left((x_0, \frac{\tau}{x_0}), \mathcal{C}_1\right) = d\left((x_0, \frac{\tau}{x_0}), H\right).$$

Part II $\mathcal{C}_2$. To show 4, let $h_2(x)$ being the distance between $(x_0, \frac{\tau}{x_0})$ and $(X_2(x), X_2(\frac{\tau}{x})) \in \mathcal{C}_1$, i.e.

$$h_2(x) = \left(x - c\tau(\tau^2 + x^4)^{-1/2} - x_0\right)^2 + \left(\frac{\tau}{x} - cx^2(\tau^2 + x^4)^{-1/2} - \frac{\tau}{x_0}\right)^2.$$

It's easy to see that

$$h_2(0_+) = \infty,\ h_2(+\infty) = \infty, \quad h_2(x_0) = c^2 < r^2.$$

Therefore, by the intermediate value theorem, we have at least $x_1 \in (0, x_0)$, $x_2 \in (x_0, \infty)$ such that $h_2(x_1) = h_2(x_2) = r^2$. To prove the uniqueness, note that

$$h_2'(x) = \frac{2\left(2c\tau x^3 + \left(\tau^2 + x^4\right)^{3/2}\right)\left(\tau^2 + x_0 x^3\right)}{x_0 x^3\left(\tau^2 + x^4\right)^{3/2}}(x - x_0)$$

Therefore, $h_1'(x) > 0$ for $x > x_0$ and $h_1'(x) < 0$ for $x < x_0$. By the monotonicity, the solution $h_1(x) = r^2$ is unique for $x \in (0, x_0)$ and $x \in (x_0, +\infty)$. □

Lemma 6. *The perpendicular bisector of* $(X_1(x), X_1(\frac{\tau}{x}))$ *and* $(X_2(x), X_2(\frac{\tau}{x}))$ *is tangent to* $\mathcal{S}_0^+(\tau)$ *at* $(x, \frac{\tau}{x})$.

Proof. It is easy to verify that

$$\frac{1}{2}\big(X_1(x)+X_2(x)\big)=x,$$
$$\frac{1}{2}\left(X_1(\frac{\tau}{x})+X_2(\frac{\tau}{x})\right)=\frac{\tau}{x}.$$

In addition,

$$\frac{X_2(\frac{\tau}{x})-X_1(\frac{\tau}{x})}{X_2(x)-X_1(x)}=\frac{x^2}{\tau},\ \frac{\mathrm{d}(\frac{\tau}{x})}{\mathrm{d}x}=-\frac{\tau}{x^2}\ \Rightarrow \frac{X_2(\frac{\tau}{x})-X_1(\frac{\tau}{x})}{X_2(x)-X_1(x)}\frac{\mathrm{d}(\frac{\tau}{x})}{\mathrm{d}x}=-1.$$

This completes the proof. □

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–333.

Alan, S., S. Ertac, and I. Mumcu (2018, 12). Gender stereotypes in the classroom and effects on achievement. *The Review of Economics and Statistics 100*(5), 876–890.

Andrews, I. and A. Mikusheva (2012). A geometric approach to weakly identified econometric models.

Andrews, I. and A. Mikusheva (2016). A geometric approach to nonlinear econometric models. *Econometrica 84*(3), 1249–1264.

Bickel, P. J., C. A. Klaassen, P. J. Bickel, Y. Ritov, J. Klaassen, J. A. Wellner, and Y. Ritov (1993). *Efficient and adaptive estimation for semiparametric models*, Volume 4. Springer.

Card, D., J. Heining, and P. Kline (2013, 05). Workplace heterogeneity and the rise of west german wage inequality*. *The Quarterly Journal of Economics 128*(3), 967–1015.

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

Dufour, J.-M., E. Renault, and V. Zinde-Walsh (2025). Wald tests when restrictions are locally singular. *The Annals of Statistics 53*(2), 457–476.

Ganics, G., A. Inoue, and B. R. and (2021). Confidence intervals for bias and size distortion in iv and local projections-iv models. *Journal of Business & Economic Statistics 39*(1), 307–324.

Gospodinov, N. (2004). Asymptotic confidence intervals for impulse responses of near-integrated processes. *The Econometrics Journal 7*(2), 505–527.

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

Hirano, K. and J. R. Porter (2012). Impossibility results for nondifferentiable functionals. *Econometrica 80*(4), 1769–1790.

Inoue, A. and L. Kilian (2002). Bootstrapping smooth functions of slope parameters and innovation variances in var(∞) models. *International Economic Review 43*(2), 309–331.

Le Cam, L. (1960). Locally asymptotically normal families of distributions; certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. Technical Report 2, University of California, Berkeley, Department of Statistics.

Le Cam, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *The Annals of Mathematical Statistics 41*(3), 802–828.

Le Cam, L. (1972). Limits of experiments. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, pp. 245–261. University of California Press.

Mikusheva, A. (2012). One-dimensional inference in autoregressive models with the potential presence of a unit root. *Econometrica 80*(1), 173–212.

Ober-Reynolds, D. (2024). Robustness to missing data: Breakdown point analysis.

Stock, J. and M. Yogo (2005). *Testing for Weak Instruments in Linear IV Regression.* New York: Cambridge University Press.

Torres, S., P. Portugal, J. T. Addison, and P. Guimarães (2018). The sources of wage variation and the direction of assortative matching: Evidence from a three-way high-dimensional fixed effects regression model. *Labour Economics 54*, 47–60.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge: Cambridge University Press.

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