Inference under First-Order Degeneracy*

Xinyue Bei[1] Manu Navjeevan[2]

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

February 7, 2026

Abstract

We study inference in models where a transformation of parameters exhibits first-order degeneracy — that is, its gradient is zero or close to zero, making the standard delta method invalid. A leading example is causal mediation analysis, where the indirect effect is a product of coefficients and the gradient degenerates near the origin. In these local regions of degeneracy the limiting behaviors of plug-in estimators depend on nuisance parameters that are not consistently estimable. We show that this failure is intrinsic — around points of degeneracy, both regular and quantile-unbiased estimation are impossible. Despite these restrictions, we develop minimum-distance methods that deliver uniformly valid confidence intervals. We establish sufficient conditions under which standard chi-square critical values remain valid, and propose a simple bootstrap procedure when they are not. We demonstrate favorable power 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

*Emails: xinyue.bei@austin.utexas.edu and mnavjeevan@tamu.edu. We thank Zheng Fang, Matt Masten, Andres Santos, and seminar participants at the UT Austin Econometrics Writing Seminar for their helpful comments. Mengyuan Jiang provided excellent research assistance.

1 Introduction

The delta method is a fundamental tool in econometric analysis for deriving the asymptotic distribution of smooth functions of estimators. In its standard form, the delta method relies on a non-zero gradient of the smooth function with respect to the primitive parameter. When this condition holds, first-order linearization provides an accurate approximation to the sampling distribution. However, in many empirically relevant scenarios, the gradient may be zero at certain points in which case higher-order terms become essential. As a result, the limiting distribution near points of degeneracy typically differs substantially from that obtained under standard regularity. A leading example arises in causal mediation analysis, where the indirect effect is the product of two primitive parameters: the effect of the treatment on the mediator and the effect of the mediator on the outcome. When these effects are zero, the gradient of the indirect effect degenerates, leading to a nonstandard limiting distribution of the Wald statistic (Sobel, 1982).

In practice, the researcher does not know whether the true parameter lies near such points of degeneracy, and thus cannot know which asymptotic approximation is appropriate for inference. This challenge has motivated a number of recent papers on hypothesis testing when the gradient may be degenerate, see, for example, van Garderen and van Giersbergen (2024) for an analysis of the mediation model mentioned above and Dufour et al. (2025) in a more general treatment of Wald-type statistics. These papers acknowledge discontinuities in limiting distributions at the point of degeneracy, and analyze the resulting distortions in Wald-type statistics. Yet, they do not propose a unified asymptotic framework for studying local regions of degeneracy. Moreover, most existing papers are interested in specific point hypotheses, leaving open the broader question of how to construct uniformly valid confidence intervals.

This paper makes two main contributions to the literature. Our first contribution is a formal asymptotic framework for studying the behavior of statistics in local regions of first-order degeneracy. In our framework, the primitive parameter is modeled as local to the point of degeneracy, in the spirit of weak identification asymptotics, where identifying information is local to zero. Under this setup, the behavior of simple plug-in estimators becomes nonstandard and depends on local parameters that cannot be consistently estimated, formally capturing an observation in the existing literature that the standard delta method does not properly approximate the behavior of plug-in estimators near degenerate points (Miller et al., 2025).

Leveraging Le Cam's limit of experiments framework (Le Cam, 1970, 1972), we show that the problem of estimating a smooth function under degeneracy is asymptotically equivalent to estimating a quadratic form of the shift parameter in a Gaussian shift model. Within this model, we show that equivariant-in-law and quantile-unbiased estimators cannot be obtained. Translated back to the original estimation problem, these results imply that regular estimation is impossible — the limiting behavior of *any properly scaled and centered estimator* in local regions of degeneracy depends on nuisance parameters that cannot be consistently estimated. Moreover, there do not exist asymptotically similar confidence intervals for the transformation of interest in local regions of degeneracy.

Our second contribution is to construct confidence intervals that are both uniformly valid and exhibit favorable power in local regions of degeneracy compared to the few existing alternatives. The impossibility results mentioned above imply that standard delta-method based confidence intervals may not be uniformly valid in such regions. Indeed, in the context of mediation analysis our simulation study shows that standard Wald-statistic based tests lead to confidence intervals that undercover when the true primitive parameter is near the origin. We thus take a different approach and propose confidence intervals based on test inversion with a minimum-distance test statistic. We first show that when the parameter dimension is two, the standard chi-square critical value is uniformly valid under either of two conditions: (i) the curvature of the null curve is not too large, or (ii) the two branches of the null curve are sufficiently close. These conditions hold in the leading mediation example. In more general settings, we propose a bootstrap critical value based on a quadratic approximation of the test statistic, and show that when the true parameter is well separated from the point of degeneracy, the bootstrap critical value is nearly identical to the efficient one.

We demonstrate the empirical relevance of our results in both simulation study and real world. In simulation study our proposed methods are shown to control size uniformly over the parameter space while standard Wald-based inference can overreject when the true primitive parameter is close to points of degeneracy, a finding also noted by Dufour et al. (2025). We additionally demonstrate favorable power properties of our proposed methods when compared to a method proposed by Andrews and Mikusheva (2016), which turns out to also be applicable in this setting.[1] These improvements in power are also seen in an application to the data of Alan et al. (2018), who consider the effect of teachers' gender attitudes on student outcomes. In revising their mediation analysis, we find that our proposed methods deliver tighter confidence intervals than existing methods for all parameters. Our results support a conclusion by van Garderen and van Giersbergen (2024) that the mediation effect of a one-year exposure to a teacher with traditional negative views is negative, while existing methods cannot rule out an effect of size zero.

The rest of this paper proceeds as follows. This section concludes with a review of the related literature. Section 2 gives examples of when first-order degeneracy may be a concern. Section 3 formally establishes the impossibility results mentioned above and discusses implications for hypothesis testing. Section 4 introduces the minimum distance based inference procedures and discusses how uniformly valid critical values may be constructed. Sections 5 and 6 contain, respectively, the simulation study and empirical application to the data of Alan et al. (2018). Section 7 concludes. Proofs are deferred to Appendices A and B.

1.1 Literature Review

Our paper is related to previous literature on econometrics and statistics studying inference under degeneracy, statistical impossibility results, and testing non-linear restrictions.

There is a growing literature examining hypothesis tests in which the null includes points of singularity. Gaffke et al. (1999) show that the distribution of the Wald statistic at points of degeneracy is nonstandard, and Gaffke et al. (2002) derive its asymptotic distribution under a variety of singular null hypotheses. Drton and Xiao (2016) demonstrate the conservativeness of the Wald test at degeneracy points for quadratic forms and for bivariate monomials of arbitrary degree. Dufour and Valery (2016) propose rank-robust regularized Wald-type tests allowing for singular covariance matrices. Dufour et al. (2025) analyze Wald tests for polynomial restrictions

[1] In their original analysis, Andrews and Mikusheva (2016) were interested in testing non-linear restrictions in the context of weak identification.

with possibly multiple constraints and show that such tests can under-reject, over-reject, or even diverge under the null; see also Dufour et al. (2025) for additional references in this area. Our paper differs from this literature in two key ways. First, prior work focuses on testing problems where the null itself contains the singularity, while our interest lies in constructing uniformly valid confidence intervals when the null may be near a singularity. In simulations, we show that the upper bound on the Wald statistic derived in Dufour et al. (2025) does not, in general, yield valid confidence intervals. Second, instead of using a Wald statistic, we employ a minimum distance–based test statistic, which is bounded in probability by construction. This approach avoids the divergence issues documented in Dufour et al. (2025).

Our paper is also related to the hypothesis testing problem with a curved null. Andrews and Mikusheva (2016) study this problem and show that the distribution of minimum-distance statistics is dominated by a tractable distribution that depends only on the maximal curvature of the null manifold relative to the known variance matrix. Inverting their test leads to uniformly valid confidence intervals. However, when the curvature of the null hypothesis is large, for example, in testing the significance of an indirect effect, their procedure yields critical values that are close to those from projection-based methods. By contrast, our procedure exploits the possibility that the null hypothesis may include multiple manifolds that are close to one another, which in turn reduces the critical value.

Finally, our paper contributes to the econometric literature on statistical impossibility results. In particular, it is related to work by Hirano and Porter (2012) who show that regular estimation of directionally, but not fully, differentiable functions is unattainable. Our paper takes a similar approach to that of Hirano and Porter (2012) in that we rule out properties of estimators by analyzing a limiting experiment (Le Cam, 1970, 1972). However, the target functional in our limit experiment is distinct from that of Hirano and Porter (2012). This approach of ruling out behaviors by analyzing limit experiments has also been utilized by Kaji (2021) and Andrews and Mikusheva (2022) in the study of weak identification. Moreover, our work is related to work by Chen and Fang (2019b) who show that all standard bootstrap procedures necessarily fail at points of degeneracy. We view this earlier work as complementary to ours, similarly to how Fang and Santos (2019) establish that the bootstrap necessarily fails as an inference procedure for the functionals considered in Hirano and Porter (2012).

2 Overview and Examples

Consider a parameter $\theta \in \Theta \subseteq \mathbb{R}^d$ and a twice continuously differentiable function $g : \Theta \to \mathbb{R}$. We are interested in inference on $g(\theta)$ in local neighborhoods of a point $\theta_\star$ for which $\nabla g(\theta_\star) = 0$. Below, we give some empirically relevant examples of when such a phenomenon may occur.

Example 2.1 (Mediation Analysis)**.** Consider a causal mediation analysis with parameter $\theta = (\theta_1, \theta_2)'$, 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 indirect effect of the treatment on the outcome is then given by $g(\theta) = \theta_1\theta_2$. At $\theta_\star = (0, 0)'$, we have $\nabla g(\theta_\star) = 0$, which complicates inference on $g(\theta)$ in local regions of $\theta_\star$. As a result, 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 problem of constructing confidence intervals in local regions of the origin. □

Example 2.2 (Impulse Response Function)**.** Consider an 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 $u_t \in \mathbb{R}$ is a white noise process. The "impulse response function" is defined as $g(\theta) = \theta^h$ and measures the impact at time period h of an initial shock. Due to the importance of this in macroeconomic analysis, inference on $g(\theta)$ has received attention from the econometric literature (Inoue and Kilian, 2002; Gospodinov, 2004; Mikusheva, 2012), mostly related to inference when θ is close to one — the so-called "unit-root" problem. However, due to degeneracy, inference on the impulse response function can also be complicated when θ is close to $\theta_\star = 0$ as $\frac{\partial}{\partial\theta} g(\theta_\star) = h\theta_\star^{h-1} = 0$ (Benkwitz et al., 2000; Lütkepohl, 2013). □

Example 2.3 (Breakdown Point Analysis)**.** Consider a missing data setup in which the researcher observes $\{Y_iD_i, D_i, X_i\}_{i=1}^n$, where $D_i \in \{0, 1\}$ represents whether or not an observation's outcome Y_i is observed, and X_i is a set of discrete covariates, i.e., $X_i \in \mathcal{X} := \{x_1, \ldots, x_K\}$. To achieve identification of parameters of interest, assumptions are typically made about the selection mechanism, such as "missing conditionally at random", i.e, $Y_i \perp D_i \mid X_i$. Ober-Reynolds (2026) proposes assessing the robustness of results to these assumptions through a breakdown point analysis. This assessment 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 as

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

where $\theta_{d,k} = \Pr(X = x_k \mid D = d)$. Simple sample analog estimators of $\theta_{0,k}$ and $\theta_{1,k}$ are $\sqrt{n}$-consistent and asymptotically normal under mild assumptions. However, the derivative of $g(\theta)$ with respect to the quantities $\theta_{0,k}$ and $\theta_{1,k}$ are given by

$$\frac{\partial}{\partial\,\theta_{0,k}} g(\theta) = \frac{1}{2}\frac{\sqrt{\theta_{0,k}} - \sqrt{\theta_{1,k}}}{\sqrt{\theta_{0,k}}}, \qquad \frac{\partial}{\partial\theta_{1,k}} g(\theta) = -\frac{1}{2}\frac{\sqrt{\theta_{0,k}} - \sqrt{\theta_{1,k}}}{\sqrt{\theta_{1,k}}}.$$

Let $\theta_\star$ be a point such that $\theta_{0,k} = \theta_{1,k}$ for all k. This occurs if the data is missing completely at random, that is $D \perp (Y, X)$. At $\theta_\star$, the derivatives above are uniformly equal to zero and the squared Hellinger distance $g(\theta)$ between P_0 and P_1 is zero. Thus, when θ is close to $\theta_\star$, that is when P_0 is close to P_1, standard approaches to inference on $g(\theta)$ will fail. □

Example 2.4 (Weak IV Bias and Size Distortion). Consider a standard homoskedastic linear IV model,

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

where $y_i, x_i \in \mathbb{R}$, $\mathbb{E}[(\epsilon_i, v_i)'] = 0$, and $Z = (z_i', \ldots, z_n')' \in \mathbb{R}^{n \times d_z}$ is treated as fixed. When θ is close to zero, identification is referred to as "weak" and it is well known that standard inference procedures for β fail to control size (Staiger and Stock, 1994). Stock and Yogo (2005) provide bounds on the size distortion of Wald tests for β in terms of the concentration parameter,

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

Ganics et al. (2021) extend this analysis and develop confidence intervals for the bias and size distortion. In both papers, the researcher makes inferences about the concentration parameter by examining the distribution of the F-statistic, a scaled version of $g(\hat{\theta})$, where $\hat{\theta}$ is the OLS estimate of θ. The analyses of both Stock and

Yogo (2005) and Ganics et al. (2021) are complicated by the fact that the limiting distribution of $g(\hat{\theta})$ is non-standard when θ is close to zero, that is, when identification is weak. This can also be seen as inference in local regions of degeneracy — at $\theta_\star = 0$ we have that $\nabla g(\theta_\star) = 2(Z'Z)\theta/\sigma_v^2 = 0$. □

Example 2.5 (Explained Variance in Linear Regression)**.** Consider a linear regression model,

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

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

$$g(\theta) = \theta'\Sigma_X\theta/\sigma_Y^2,$$

that is, the population R^2. Although empirical work typically reports only a point estimate of R^2, reporting a confidence set for R^2 is informative for comparing the explanatory power or predictive performance of competing models (Hawinkel et al., 2024). When R^2 is bounded away from zero and one, standard errors and confidence intervals can be obtained using conventional asymptotic approximations (Cohen et al., 2013). However, at $\theta_\star = 0$ we have that $\nabla g(\theta) = 2\Sigma_X\theta_\star/\sigma_Y^2 = 0$. Consequently, inference on the explained variance is non-standard when θ is close to zero, or equivalently, when R^2 is close to zero.

This type of parameter is also of interest in labor economics when explaining variation in wage regressions. 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. Card et al. (2013) compare the baseline Abowd, Kramarz, and Margolis (1999) (AKM) model with 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 that degeneracy may be a concern when conducting inference on $g(\theta)$. □

3 Impossibility Results

In this section we establish that standard approaches to inference on $g(\theta)$ necessarily fail in local regions of first-order degeneracy. We begin in Section 3.1 by introducing a parametric framework to study this problem and defining what it means for an estimator to be "regular" in this setting. Section 3.2 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 3.3 extends the analysis in two directions: first, to hypothesis testing problems where the null hypothesis that $g(\theta) = g(\theta_\star)$ 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 necessarily break down in local regions of degeneracy.

3.1 Preliminaries

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

$$X^{(n)} \sim P_{n,\theta} \tag{3.1}$$

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

Assumption 3.1 (Local Asymptotic Normality)**.** *There exists a sequence $r_n \to \infty$ such that for every $\theta \in \Theta^\circ$ and every sequence $h_n \to h \in \mathbb{R}^d$*

$$\log\left(\frac{dP_{n,\theta+h_n/r_n}}{dP_{n,\theta}}(X^{(n)})\right) = h'\Delta_n - \frac{1}{2}h'\Gamma_\theta h + Z_n(h) \tag{3.2}$$

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 $h \in \mathbb{R}^d$, $Z_n(h) \xrightarrow{p} 0$.

Example 3.1 (Smooth Parametric Models)**.** A leading example of a model that satisfies (3.2) is when the researcher observes i.i.d data, $X_i \overset{iid}{\sim} P_\theta$ where $\theta \in \Theta$. Assume that there exists a dominating measure μ such that $P_\theta \ll \mu$ for all $\theta \in \Theta^\circ$ 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 $\theta \in \Theta^\circ$,

$$\int \left[\sqrt{p_{\theta+h}} - \sqrt{p_\theta} - \frac{1}{2} h' \dot{\ell}_\theta \sqrt{p_\theta} \right]^2 d\mu = o(\|h\|^2), \quad h \to 0 \tag{3.3}$$

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

$$\begin{aligned} \log \left(\frac{dP_{n,\theta+h/\sqrt{n}}}{dP_\theta} (X^{(n)}) \right) &= \log \prod_{i=1}^n \frac{p_{\theta+h/\sqrt{n}}}{p_\theta}(X_i) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n h' \dot{\ell}_\theta(X_i) - \frac{1}{2} h' \Gamma_\theta h + R_n(h), \end{aligned}$$

where $R_n(h) = o_{P_{n,\theta}}(1)$ for all $h \in \mathbb{R}^d$. By the central limit theorem, $\frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\ell}_\theta(X_i) \overset{P_{n,\theta}}{\rightsquigarrow} N(0, \Gamma_\theta)$. Thus, by letting $\Delta_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\ell}_\theta(X_i)$ we see that Assumption 3.1 is satisfied with $r_n = \sqrt{n}$. $\square$

We study a twice continuously differentiable scalar functional $g : \Theta \to \mathbb{R}$, and the behavior of estimators of $g(\theta)$ in local regions of a point $\theta_\star \in \Theta^\circ$ which is such that the first-order derivatives of $g(\cdot)$ at $\theta_\star$ are zero. We will refer to $\theta_\star$ as the "point of degeneracy" and local neighborhoods of $\theta_\star$ as "local regions of degeneracy."

Assumption 3.2 (Differentiability)**.** *The function $g : \Theta \to \mathbb{R}$ is twice continuously differentiable on Θ, a compact subset of $\mathbb{R}^d$, with $\nabla g(\theta_\star) = 0$ and $\nabla^2 g(\theta_\star) \neq 0$ for some $\theta_\star \in \Theta^\circ$.*

Given the maintained assumption of a locally asymptotically normal model, it is useful to examine regions close to $\theta_\star$ by adopting a local parameterization 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, $\Psi_n : \mathcal{X}^n \to \mathbb{R}$. 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

$$r_n^2 \left(\Psi_n - g(\theta_{n,h})\right) \overset{h}{\rightsquigarrow} \mathcal{L}_h. \tag{3.4}$$

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

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

for every $h \in \mathbb{R}^d$. In the smooth parametric models described in Examples 3.1, such an estimator could be the maximum likelihood estimator or a Bayes estimator such as the posterior mean. Since $g(\cdot)$ is assumed to be twice continuously differentiable, the limiting behavior of the plug-in estimator $g(\hat{\theta})$ can be found via the second order delta method

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

The behavior of the plug in estimator, $g(\hat{\theta})$, depends on the local parameter h. $\square$

We focus on ruling out regular and locally asymptotically α-quantile unbiased estimation of $g(\theta_{n,h})$ in local regions of $\theta_\star$. Formally, these notions are defined as follows.

Definition 3.1 (Regularity)**.** Let Ψ_n be an estimator satisfying (3.4), 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 *locally asymptotically α-quantile unbiased* if its limiting α-quantile is zero for every h, i.e. $\mathcal{L}_h\{(-\infty, 0]\} = \alpha$ for all $h \in \mathbb{R}^d$.

The existence of regular estimators is closely tied to the validity of Wald-type inference procedures — without regular estimators, standard Wald-type inference procedures that compare test statistics to fixed critical values will not have correct asymptotic size. Similarly, the existence of locally asymptotically α-quantile unbiased estimators is closely tied to the existence of asymptotically similar confidence intervals for $g(\theta)$ in local regions of $\theta_\star$. 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 $\Psi_n = \hat{c}$, by ruling out such estimators we also rule out the possibility of similar one-sided confidence intervals.[1]

3.2 Analysis in the Limiting Experiment

To examine the possible behavior of such estimators, we make use of a representation result, given below in Proposition 3.1, which is a 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 3.1 (Limit Experiment). *Suppose Assumption 3.1 holds, and let Ψ_n be a sequence of estimators satisfying (3.4). 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

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

Proposition 3.1 establishes an equivalence between estimating $g(\theta)$ in local regions of first-order degeneracy and estimating of a quadratic form of the mean parameter in a Gaussian shift model in which one observes a single draw $Z \sim N(h, \Gamma_{\theta_\star}^{-1})$, where $\Gamma_{\theta_\star}^{-1}$ is known but h is not. In particular, in a spirit similar 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

[1] The focus on one-sided confidence intervals is largely for simplicity of exposition. If $(-\infty, \hat{c}_1]$ and $[\hat{c}_2, \infty)$ are two asymptotically similar confidence intervals for $g(\theta)$ each with coverage rate $1 - \alpha/2$ and $\hat{c}_1 \leqslant \hat{c}_2$ with probability approaching one, then $[\hat{c}_2, \hat{c}_1]$ is asymptotically similar with coverage rate $1 - \alpha$.

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, $\Psi(Z,U)$, and law, $\mathcal{L}$ with $\int x^2\, d\mathcal{L}(x) < \infty$, such that $\Psi(Z,U) - \frac{1}{2}h'\nabla^2 g(\theta_\star)h$ 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 $\Psi(Z,U)$ is unbiased. On the other hand, the Cramér-Rao lower bound for the variance of any unbiased estimator of $\frac{1}{2}h'\nabla^2 g(\theta_\star)h$ in the Gaussian shift model yields

$$\operatorname{Var}(\Psi(Z,U)) \geqslant h'(\nabla^2 g(\theta_\star))'\Gamma_{\theta_\star}^{-1}(\nabla^2 g(\theta_\star))h. \tag{3.5}$$

By letting h vary over $\mathbb{R}^d$, the right hand side of (3.5) 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 3.1. It is instructive 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. Our arguments, on the other hand, rule out regular estimation by analyzing the "global" behavior of a potential regular estimator, that is, the behavior as h varies over $\mathbb{R}^d$. The approach taken by Hirano and Porter (2012) does not apply in the present setting as the parameter of interest in the limit experiment is non-linear but continuously differentiable at zero. In contrast, in the limit 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 continuously differentiable at 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 3.2. *Let $Z \sim N(h, \Gamma_{\theta_\star}^{-1})$ and $U \sim \operatorname{Unif}(0,1)$ independently of Z. Let J be a $d \times d$ non-zero, symmetric matrix.*

1. *There is no randomized statistic $\Psi(Z,U)$ and law $\mathcal{L}$ on $\mathbb{R}$ with $\Psi(Z,U) - h'Jh \overset{h}{\rightsquigarrow}$*

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

2. *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 3.1 and 3.2 can be combined for the main result of this section, which rules out sufficiently regular estimation in local areas of first-order degeneracy.[2]

Theorem 3.1 (Impossibility of Regular Estimation)**.** *Suppose Assumptions 3.1 and 3.2 hold.*

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

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

2. *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 $\mathcal{L}_h$ are differentiable at zero with derivatives bounded below by $\epsilon > 0$. Then there is no estimator sequence Ψ_n such that*

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

Theorem 3.1 rules out sufficiently well-behaved estimation of $g(\theta)$ when the true parameter is "close" to $\theta_\star$. In particular, Theorem 3.1(a) rules out the possibility of regular estimation — the properly scaled and centered 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. Similarly, Theorem 3.1(b) rules out the possibility of α-quantile unbiased estimation. As mentioned below Definition 3.1, this result has profound implications for inference on $g(\theta)$ in local regions of $\theta_\star$. In particular, both asymptotically exact Wald-type inference procedures and asymptotically similar confidence intervals for $g(\theta)$ are unavailable in local regions of degeneracy $\theta_\star$.

[2] In the application of Proposition 3.2 to our setting, take $J = \frac{1}{2}\nabla^2 g(\theta_\star)$. The result in Proposition 3.2 rules out well-behaved estimation of any quadratic form of the shift parameter, not just those associated with the Hessian of $g(\cdot)$.

Theorem 3.1(a) also has implications for the construction of efficient estimators of $g(\theta)$ in local regions of $\theta_\star$. Standard notions of efficiency are tied to comparing the asymptotic risk of regular estimators. Theorem 3.1(a) shows that, after properly scaling, such regular estimators are not available in local regions of degeneracy. Consequently, alternative notions of efficiency must be considered in these settings and standard estimators may not be optimal in these regions. As an example, one can show that estimators of $g(\theta)$ that are efficient under a standard asymptotic regime can be dominated by alternative estimators in local asymptotic mean squared error around points of degeneracy.

Remark 3.2. 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 3.1 requires some regularity conditions on the system of limiting laws $\{\mathcal{L}_h\}_{h\in\mathbb{R}^d}$.[3] The regularity condition in Theorem 3.1(b) implies that, if a locally asymptotically α-quantile unbiased estimator were to exist, its associated limiting laws $\mathcal{L}_h$ must be able to be made *arbitrarily* 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. As an example, suppose that $\{\mathcal{L}_h\}_{h\in\mathbb{R}^d}$ is a family of Gaussian distributions on $\mathbb{R}$ associated with a locally asymptotically α-quantile unbiased estimator. Then, the variance of these Gaussian distributions must be able to become arbitrarily large as h ranges over $\mathbb{R}^d$. □

3.3 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 3.1. 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

[3] Let $F_0(\cdot)$ be the CDF associated with $\mathcal{L}_0$. Hirano and Porter (2012) show that, for any α-quantile unbiased estimate, it must be the case that either $F_0(\cdot)$ is not differentiable at zero or must satisfy $F_h'(0) = 0$.

Theorem 3.1 to infinite dimensional, i.e, semiparametric or nonparametric, models.

3.3.1 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 $g(\theta_\star)$, i.e, $r_n^2(g(\theta_{n,h}) - g(\theta_\star)) \to 0$;

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

We study the behavior of asymptotically similar tests in local regions of degeneracy. In this setup, an asymptotically similar test is a statistic $\Xi_n : \mathcal{X}^n \to \{0, 1\}$ such that

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

Equivalently, the test is similar if $\limsup_{n\to\infty} P_{\theta_{n,h}}(1 - \Xi_n \leqslant 0) = \alpha$. Letting $\Psi_n = 1 - \Xi_n$ it is apparent that this is a nearly identical requirement to that of local α-quantile unbiasedness in Definition 3.1, with the key difference being that the requirement only needs to hold for local parameters $h \in \mathcal{H}_\star$ rather than for all $h \in \mathbb{R}^d$.

However, unlike quantile unbiased estimation, which is ruled out in Theorem 3.1, asymptotically similar tests can exist — one can imagine constructing a similar test by flipping a weighted coin. Such a test, though, may not be powerful against local alternatives close to $\theta_\star$. Our main result in this subsection establishes this formally: if such test exists then its local asymptotic power curve 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

$$\mathcal{P}(h) = \limsup_{n\to\infty} P_{\theta_{n,h}}(\Xi_n = 1) \tag{3.6}$$

Proposition 3.3. *Let Ξ_n be an asymptotically similar test such that $\mathcal{P}(h)$ is differentiable at $h = 0$. Then, the directional derivative of the local asymptotic power curve,*

$\mathcal{P}(h)$, in directions $h \in \mathcal{H}_\star$ is equal to zero:

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

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

Remark 3.3 (Differentiability of $\mathcal{P}$)**.** A common strategy in hypothesis testing is to compare a test statistic Ψ_n° to a possibly data-dependent critical value $\hat{c}_n$, rejecting when the former exceeds the latter. Let $\Psi_n = \Psi_n^\circ - \hat{c}_n$, so the rejection rule can be written as $\Xi_n = \mathbf{1}\{\Psi_n \geqslant 0\}$. If $\Psi_n \overset{h}{\rightsquigarrow} \mathcal{L}_h$ for each $h \in \mathbb{R}^d$, and if the CDF of $\mathcal{L}_0$ is continuous at zero, then the resulting local asymptotic power curve $\mathcal{P}$ is differentiable at 0; see Lemma A.3. This is a milder version of the regularity condition imposed on quantile unbiased estimators in Hirano and Porter (2012). □

The first statement in Proposition 3.3 follows immediately from the definition of similarity along with the fact that $\mathcal{H}_\star$ is a cone: because $\mathcal{P}(h)$ is constant on $\mathcal{H}_\star$, its directional derivatives in directions $h \in \mathcal{H}_\star$ must be zero. The force of the result, however, lies in the structure of $\mathcal{H}^\star$ near points of degeneracy. In standard inference problems, i.e, when the true parameter is well separated from points of degeneracy, the parameter of interest in the limit experiment is a linear function of the shift parameter h, so $\mathcal{H}^\star = \{0\}$ and the zero-derivative condition carries no information about the shape of the power curve. By contrast, when g exhibits first-order degeneracy $\mathcal{H}_\star$, can be a non-trivial cone — that is, it may contain directions other than zero — and the constraint $D_h \mathcal{P}(0) = 0$ for $h \in \mathcal{H}_\star$ becomes substantive. When $\nabla^2 g(\theta_\star)$ is indefinite, $\mathcal{H}_\star$ spans $\mathbb{R}^d$ and the entire gradient of the local asymptotic power curve vanishes at the origin, implying that power cannot increase at a linear rate in any direction away from $\theta_\star$. The following examples illustrate the strength of this restriction.

Example 2.1, cont. Consider again the mediation model, where the original model is given $(P_\theta : \theta \in \Theta \subseteq \mathbb{R}^2)$. Suppose 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 $\theta_\star = 0$. We can show that $\mathcal{H}_\star = \{h \in \mathbb{R}^2 : h_1 h_2 = 0\}$. Since $\mathcal{H}_\star$ is the union of the two coordinate axes, $\operatorname{span}(\mathcal{H}_\star) = \mathbb{R}^2$. Thus, we have that $\nabla \mathcal{P}(0) = 0$ for any asymptotically similar test.

We can equivalently show that $\mathcal{H}_\star$ must span $\mathbb{R}^2$ by noting that the second derivative matrix of $g(\cdot)$ at $\theta_\star$ is given by

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

which is indefinite. $\square$

Example 3.3 (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 $\nabla^2 g(\theta_\star)$ is positive definite, $\mathrm{span}(\mathcal{H}_\star) = \{0\}$. Thus, the results of Proposition 3.3 do 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 (2019a). $\square$

Example 3.4 (Standard Inference)**.** Suppose that the primitive parameter is univariate $\theta_{n,h} = \theta_0 + h/r_n \in \mathbb{R}$, and local to a point θ_0 such that $g'(\theta_0) > 0$, that is we are well separated from points of degeneracy. In this setting, the researcher typically has access to an estimator Ψ_n that satisfies $r_n(\Psi_n - g(\theta_{n,h})) \overset{h}{\rightsquigarrow} N(0, \sigma^2)$. This estimator is regular and thus α-quantile-unbiased for all $\alpha \in (0,1)$. Based on this estimator, an asymptotically similar one sided test for the null hypothesis, $H_0 : g(\theta) = g(\theta_0)$, can be constructed with local asymptotic power curve $\mathcal{P}(h) = 1 - \Phi(c_{1-\alpha} - g'(\theta_0)h/\sigma)$, where $\Phi(\cdot)$ is the standard normal CDF and $c_{1-\alpha}$ is its $1-\alpha$ quantile. Here, $\frac{\partial}{\partial h}\mathcal{P}(h)\big|_{h=0} = g'(\theta_0)\phi(c_{1-\alpha})/\sigma > 0$.

Remark. Recent papers by van Garderen and van Giersbergen (2024) and Dufour et al. (2025) also study tests of 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 $W_n = \frac{g(\hat{\theta}) - g(\theta_\star)}{\nabla g(\hat{\theta})'\Sigma\nabla g(\hat{\theta})}$, where Σ represents the asymptotic variance of $\hat{\theta}$. The authors show that, when θ is close to $\theta_\star$, the behavior 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. The results in Proposition 3.3 are narrower in their conclusion — we establish only that similar tests must have flat power at $\theta^\star$ — but broader in their scope, since they apply to any testing procedure that depends on the data, rather than only to procedures based on a specific initial estimator $\hat{\theta}$. $\square$

3.3.2 Infinite Dimensional Models

In many settings 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 semiparametric 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 3.1 and 3.2.

Formally, 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 $\mathcal{P}$ is 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 3.1 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 3.1 can be extended in a straightforward fashion to infinite dimensional models.

Remark 3.4 (Semiparametric Models with i.i.d Data)**.** In the literature on semiparametric estimation with i.i.d data, where the researcher observes repeated observations drawn independently from a probability distribution P on $\mathcal{X}$ belonging to a model $\mathcal{P}$ (Bickel et al., 1993), one can associate the entire sequence of probability measures $\{\otimes_{i=1}^n P : n \in \mathbb{N}\}$ with the underlying common distribution P. With this association, one can consider the model $\mathcal{P}$ described above as a collection of probability measures on $\mathcal{X}$ rather than a collection of sequences of probability measures $\{P_n\}$ where each P_n is defined on $\mathcal{X}^n$. The parameter, in turn, can be defined as a function of the underlying distribution P rather than as a function of the entire sequence by $g[P] \equiv g[\{\otimes_{i=1}^n P : n \in \mathbb{N}\}]$. However, when dealing with time-series or network data, it may not be possible to define the parameter as a function of some representative underlying distribution and is instead a property of the sequence $\{P_n\}$. $\square$

Corollary 3.1 (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 3.2 with $\{P_{n,\theta_\star}\} = \{P_{0,n}\}$.*

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

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

2. *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 $\mathcal{P}_f$,*

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

4 Minimum Distance Based Inference

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

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

We focus on the settings where the standard first order approximation of $g(\theta)$ fails at $\theta_\star$, but the second order derivative is nondegenerate; that is, $\frac{\partial^2 g}{\partial\theta\partial\theta'}(\theta_\star) = H$ with H full rank.

In Section 4.1, we discuss a simple case where $\Theta \subseteq \mathbb{R}^2$ 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 4.2, we propose a computationally simple method for a general g, which can be generalized to cases with higher order singularity.

4.1 Two-Dimensional θ and Indefinite H

To simplify notation, consider the null hypothesis

$$H_0 : g(\theta_1, \theta_2) := (1+\rho)\theta_2^2 - (1-\rho)\theta_1^2 = \tau, \tag{4.1}$$

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 $n = r_n = 1$, and assume $\hat{\theta} - \theta \sim N(0, I_2)$. The quadratic form g and the normality of $\hat{\theta}$ can be viewed as second order approximations, with general asymptotic results provided in Theorem 4.1.

Let $X_2(\theta_1) \in \mathbb{R}_+$ be the positive solution for θ_2 such that (4.1) 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

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

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

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

Proposition 4.1. *Let $c = \sqrt{Q(\chi_1^2, 1-\alpha)}$ and $\hat{\theta} - \theta \sim N(0, I_2)$. Suppose either $\frac{1-\rho}{\sqrt{\tau(1+\rho)}} \leqslant \frac{1}{c}$ or $\rho \geqslant 0$. For all $\theta \in \mathcal{S}_0(\tau)$, it holds that*

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

Proposition 4.1 shows that the standard MD test remains valid under a curved null

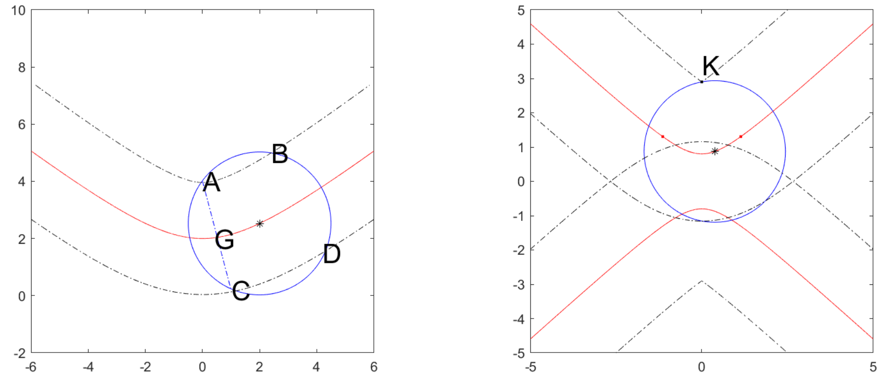


Figure 4.1: Acceptance Regions under Low (Left) and High (Right) Curvature. Red curves represent $\mathcal{S}_0^+(\tau)$ (left) and $\mathcal{S}_0(\tau)$ (right). Black dash curves represent the boundaries of $\mathcal{S}^+(\tau,c)$ (left) and $\mathcal{S}^+(\tau,c)$ and $\mathcal{S}^-(\tau,c)$ (right). "*" represents θ, and the blue curves represent $\partial B(\theta,r)$ for some $r>c$.

hypothesis when the maximum curvature of $\mathcal{S}_0^+(\tau)$, given by $\frac{1-\rho}{\sqrt{\tau(1+\rho)}}$, is sufficiently small, or when the two branches $\mathcal{S}_0^+(\tau)$ and $\mathcal{S}_0^-(\tau)$ are sufficiently close. The argument proceeds by comparing the coverage of $\mathcal{S}(\tau,c)$ with that of the auxiliary acceptance set,

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

whose coverage is exactly $1-\alpha$. Expressed in polar coordinates, the coverage depends on the fraction of each circle of radius r centered at $\theta \in \mathcal{S}_0^+(\tau)$, denoted $\partial B(\theta,r)$, that is contained in the acceptance region. Consequently, it suffices to show that, for each r, the arc length of $\partial B(\theta,r)$ contained in $\mathcal{S}(\tau,c)$ is no smaller than that of $\mathcal{S}_{aux}$.

When $r \leqslant c$, the entire circle is covered by both $\mathcal{S}_{aux}$ and $\mathcal{S}(\tau,c)$ by construction. For $r>c$, let $\mathcal{C}_u(\tau)$ and $\mathcal{C}_\ell(\tau)$ denote the upper and lower boundaries of $\mathcal{S}^+(\tau,c)$; see Figure 4.1 left panel. If $\frac{1-\rho}{\sqrt{\tau(1+\rho)}} \leqslant \frac{1}{c}$, the circle $\partial B(\theta,r)$ intersects $\mathcal{C}_u(\tau)$ at points A and B, and $\mathcal{C}_\ell(\tau)$ at points C and D. We can show that the lengths of chords $\overline{AC}$ and $\overline{BD}$ are no smaller than $2c$. Otherwise, $B(G,c) \nsubseteq \mathcal{S}^+(\tau,c)$, where G denotes the intersection of AC with $\mathcal{S}_0^+(\tau)$, contradicting the definition of $\mathcal{S}^+(\tau,c)$. Note that the arcs of $\partial B(\theta,c)$ covered by $\mathcal{S}_{aux}$ correspond to chords of length $2c$. It therefore follows that the portion of $\partial B(\theta,c)$ covered by $\mathcal{S}(\tau,c)$ is larger than that covered by $\mathcal{S}_{aux}$.

If $\frac{1-\rho}{\sqrt{\tau(1+\rho)}} > \frac{1}{c}$, $\mathcal{C}_u(\tau)$ has a kink due to the high curvature. Thus, there exist $\theta \in \mathcal{S}_0^+(\tau)$ and $r>c$ such that $\partial B(\theta,r)$ does not intersect $\mathcal{C}_u(\tau)$; see Figure 4.1 right panel. For such (θ, r), the argument based on chord lengths no longer applies. However, when $\rho \geqslant 0$, $\partial B(\theta, r)$ is sufficiently close to $\mathcal{S}_0^-(\tau)$, and we can show that $\partial B\left((\theta_1, \theta_2), r\right) \subseteq \mathcal{S}(\tau, c)$.

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

Assumption 4.1. *Suppose that*

1. $\nabla^2 g(\theta_\star)$ *has full rank.*
2. *Let* BL_1 *denote the set of Lipschitz 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*

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

   where $\xi_P \sim N(0, \Sigma_P)$.
3. *Let* $\mathcal{S}$ *denote the set of matrices with eigenvalues bounded below by* $\underline{e} > 0$ *and above by* $\bar{e} \geqslant \underline{e}$. *For all* $P \in \mathcal{P}$, $\Sigma_P \in \mathcal{S}$.
4. *For all* $\varepsilon > 0$,

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

Assumption 4.1.1 assumes that the second order derivative of g at $\theta_\star$ is of full rank, so that there is no higher order degeneracy. Assumption 4.1.2 requires that the researcher has access to an estimator $\hat{\theta}$ that is uniformly asymptotically normal over the class of DGPs considered while Assumption 4.1.3 and Assumption 4.1.4 require that the asymptotic variance of this estimator is well-behaved and consistently estimable. Since degeneracy is a property of the transformation of interest, g, rather than of the primitive parameter θ, these are mild conditions that can be verified for most common estimators $\hat{\theta}$.

Theorem 4.1. *Suppose* $d = 2$, *and let* $c = \sqrt{Q(\chi^2_1, 1-\alpha)}$. *Let* $(\lambda_{P,1}, \lambda_{P,2})$ *be the eigenvalues of sign*$\left(g(\theta_P) - g(\theta_\star)\right) \Sigma_P^{1/2} H \Sigma_P^{1/2}$, *and define* $\rho_P = \frac{\lambda_{P,1}+\lambda_{P,2}}{\left|\lambda_{P,1}-\lambda_{P,2}\right|}$. *Assume* *that Assumptions 3.2 and 4.1 hold. If for some $\eta > 0$, it holds that either*

$$\mathcal{P}_n \subseteq \left\{ P \in \mathcal{P} : \frac{(1-\rho_P)\sqrt{|\lambda_{P,1} - \lambda_{P,2}|}}{2r_n\sqrt{|g(\theta_P) - g(\theta_\star)|\,(1+\rho_P)}} \leqslant \frac{1}{c},\ \rho_P \in [\eta - 1, 1-\eta] \right\} \tag{4.2}$$

or

$$\mathcal{P}_n \subseteq \left\{ P \in \mathcal{P} : \rho_P \in [0, 1-\eta] \right\}, \tag{4.3}$$

then

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

Theorem 4.1 follows from Proposition 4.1. To see this, let $\tilde{g}(h) = g\left(\theta_\star + r_n^{-1}\Sigma^{1/2}h\right)$, where r_n governs the rate at which θ is estimated and Σ adjusts for the covariance. For $h = O(1)$, $\tilde{g}(h)$ can be approximated by a hyperbola, as in (4.1). Condition (4.2) ensures that the curvature of $\tilde{g}(h)$ is not too large, while (4.3) implies that the two branches of the hyperbola are sufficiently close. The result remains uniformly valid even as $\|h\| \to \infty$.

Remark 4.1. To illustrate Theorem 4.1, consider $g(\theta) = \theta_1\theta_2$, as motivated by Example 2.1. 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 $\lambda_{P,1} = \operatorname{sign}\left(g(\theta_P)\right)(r-1)\,\sigma_1\sigma_2$, $\lambda_{P,2} = \operatorname{sign}\left(g(\theta_P)\right)(r+1)\,\sigma_1\sigma_2$, and $\rho_P = \operatorname{sign}\left(g(\theta_P)\right) r$. Therefore, (4.3) holds when $r = 0$, and the MD test with the simple critical value yields a uniformly valid confidence interval for the mediation effect. It is worth noting that, the rejection region for $\theta_1\theta_2 = 0$ is given by $\min\left\{|\hat{\theta}_1|, |\hat{\theta}_2|\right\} > Q(\chi_1^2, 1-\alpha)$, which coincides with the rejection region of the likelihood ratio test. The latter is the uniformly most powerful invariant test among information- and size-coherent tests (van Garderen and van Giersbergen (2022)). If $\operatorname{sign}\left(g(\theta_P)\right) r < 0$, then (4.2) is satisfied when $\frac{\sqrt{\sigma_1\sigma_2}(1+|r|)}{r_n\sqrt{2(1-|r|)\left|g(\theta_P)\right|}} \leqslant \frac{1}{c}$. Since σ_1, σ_2 and r can be consistently estimated, and $g(\theta_P)$ is known under H_0, conditions (4.2) and (4.3) are straightforward to verify in practice.

4.2 General Cases

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 $\theta_n = \theta_\star + h_n/r_n$. Under H_0, the test statistic is given by

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

where $\tilde{\theta}_n$ denotes the minimizer. Since $\hat{T}_n(g(\theta_n)) \leqslant r_n^2 \left\| \hat{\Sigma}^{-1/2}(\hat{\theta} - \theta_n) \right\|^2 = O_p(1)$, we have $\tilde{\theta}_n = \theta_n + O_p(\frac{1}{r_n})$. If $h_n = O(1)$, a second order Taylor expansion of $g(\tilde{\theta}_n) = g(\theta_n)$ gives

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

In addition, let $\mathbb{Z}_n = r_n \hat{\Sigma}^{-1/2}(\hat{\theta} - \theta_n)$, we can write

$$\begin{aligned} r_n \hat{\Sigma}^{-1/2}(\hat{\theta} - \tilde{\theta}_n) &= r_n \hat{\Sigma}^{-1/2} \left((\hat{\theta} - \theta_n) + (\theta_n - \theta_\star) - (\tilde{\theta}_n - \theta_\star) \right) \\ &= \mathbb{Z}_n + \hat{\Sigma}^{-1/2} \left(h_n - r_n(\tilde{\theta}_n - \theta_\star) \right). \end{aligned}$$

In sum, let $t = r_n(\tilde{\theta}_n - \theta_\star)$, given h_n, we can approximate $\hat{T}_n$ by

$$\hat{T}_n^*(h_n) = \inf_{t: t' H t = h_n' H h_n} \left\| \mathbb{Z} + \hat{\Sigma}^{-1/2}(h_n - t) \right\|^2 \tag{4.4}$$

where $\mathbb{Z} | \hat{T}_n(g(\theta_n)) \sim N(0, I_d)$. We can show that $\hat{T}_n(g(\theta_n))$ and $\hat{T}_n^*(h_n)$ have the same asymptotic distribution, regardless of whether h_n converges to $h \in \mathbb{R}$ or diverges to infinity (Lemmas B.7 and B.8). Intuitively, if $h_n \to \infty$, the restriction for the optimizer $\tilde{t}$ in (4.4) is approximately linear. In this case, both $\hat{T}_n(g(\theta_n))$ and $\hat{T}_n^*(h_n)$ are approximated χ_1^2.

Given h_n, we can easily get the quantile of $\hat{T}_n^*(h_n)$ by simulation. However, h_n is a nuisance parameter that cannot be consistently estimated. Next, we propose a two step feasible critical value. Suppose set $\mathcal{H}_z$ satisfies $P\left(N(0, I_d) \in \mathcal{H}_z\right) = 1 - \eta$.[1] In

[1] For instance, $\mathcal{H}_z = \{z \in \mathbb{R}^d : z'z \leqslant Q(\chi_d^2, 1 - \eta)\}$.

the first step, we construct a $(1-\eta)$ confidence set for h_n,

$$\mathcal{H} = r_n(\hat{\theta} - \theta_\star) - \hat{\Sigma}^{1/2}\mathcal{H}_z. \tag{4.5}$$

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

$$\hat{c} = \sup_{h \in \mathcal{H}} Q\left(\hat{T}_n^*(h)\middle| \mathbb{Z} \in \mathcal{H}_z; \frac{1-\alpha}{1-\eta}\right), \tag{4.6}$$

and reject $H_0 : g(\theta) = \tau$ if $\hat{T}_n(\tau) > \hat{c}$. In (4.6), the construction of $\hat{c}$ takes into account the first step selection, thus it is less conservative than simple Bonferroni correction.

Theorem 4.2. *Under Assumptions 3.2 and 4.1, it holds that*

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

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

$$\lim_n P_n\left(\hat{T}_n(g(\theta_{P_n})) \leqslant \hat{c}\right) \in [1-\alpha, 1-\alpha+\eta). \tag{4.7}$$

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 $Q(\chi_1^2, 1-\alpha)$. The cost is that we need to introduce an extra tuning parameter.

Remark 4.2. 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

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

$H_0 : g(\theta) = \tau$ is rejected if $\hat{T}_n(\tau) > \hat{c}$. □

Remark 4.3. Dufour et al. (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 et al. (2025) focus solely on hypothesis tests at a fixed point, i.e., testing $g(\theta) = g(\theta_\star)$, whereas this paper aims to construct uniformly valid confidence intervals. $\square$

Remark 4.4. 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 the mediation analysis problem in Examples 2.1 where one is interested in testing the null hypothesis $H_0 : \theta_1\theta_2 = \tau$. As τ approaches zero, the curvature of the null manifold can be made arbitrarily large and the critical value of Andrews and Mikusheva (2016) approaches $Q(\chi^2_2, 1-\alpha)$. However, as shown in Section 4.1 of this paper, a uniformly valid critical value in this setting is $Q(\chi^2_1, 1-\alpha)$. Indeed, even when τ is far from zero, the Andrews and Mikusheva (2016) critical value is always larger than $Q(\chi^2_1, 1-\alpha)$. $\square$

5 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 2.1, namely the construction of confidence intervals for the mediation effect. In addition to the two MD-based methods proposed in Section 4, one using the $Q(\chi^2_1, 1-\alpha)$ critical value (BN1; Section 4.1) and one using a bootstrapped critical value (BN2; Section 4.2), we consider two uniformly valid MD-based alternatives: (i) the procedure of Andrews and Mikusheva (2016) (AM),[1] and (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 $Q(\chi^2_1, 1-\alpha)$, and a naive bootstrap method. The nominal rejection rate is $\alpha = 0.05$, and the tuning parameter for BN2 is $\eta = \alpha/10$.

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

[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 procedures have nearly identical power, with the latter performing slightly worse.

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 $r = 0$ and $r = 0.5$, with $\theta_2 \in \{2, 6\}$ and $\theta_1 = [-1 : 0.2 : 1] \times \theta_2$.

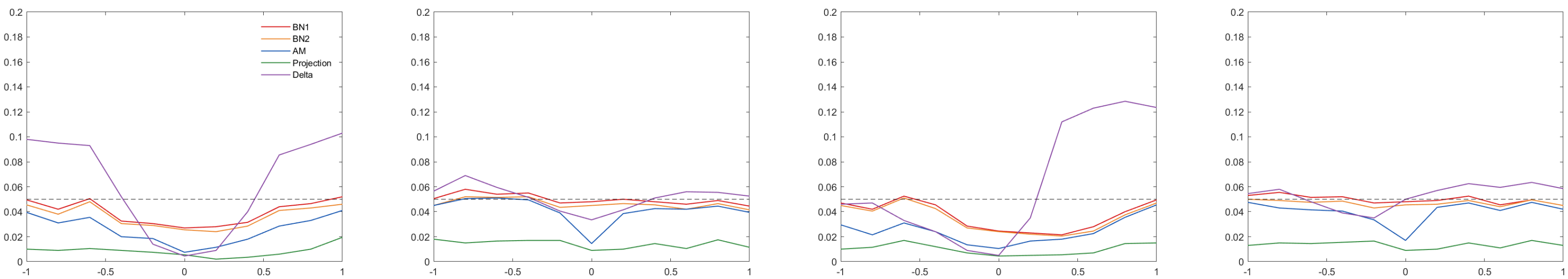
In Figure 5.1a, we plot the probability that the confidence intervals exclude the true value $g(\theta)$. The naive method does not overreject when $\theta_1\theta_2 = 0$, but its rejection probability is very low near the origin, consistent with earlier findings (e.g., Dufour et al. (2025)). Away from the origin (see, e.g., $\theta_1 = \theta_2 = 2$), the naive Wald test overrejects. According to Remark 4.1, BN1 is valid for $r = 0$. For $r = 0.5$, Theorem 4.1 does not guarantee validity when $\theta_1 < 0$, $\theta_2 = 2$, or when $\theta_1 \in [-1.4, 0)$, $\theta_2 = 6$. Nevertheless, BN1 maintains correct size across all designs, even when these conditions fail, suggesting that the condition is sufficient but not necessary. As expected, all other MD-based methods control size.

Figure 5.1b shows the probability that the confidence intervals exclude zero, i.e., the probability of obtaining a significant result. When θ is close to the origin ($\theta_2 = 2$), our methods have substantially higher power than AM, whose performance is close to that of the simple projection method. When θ is further from the origin ($\theta_2 = 6$), power curves across methods are nearly identical.

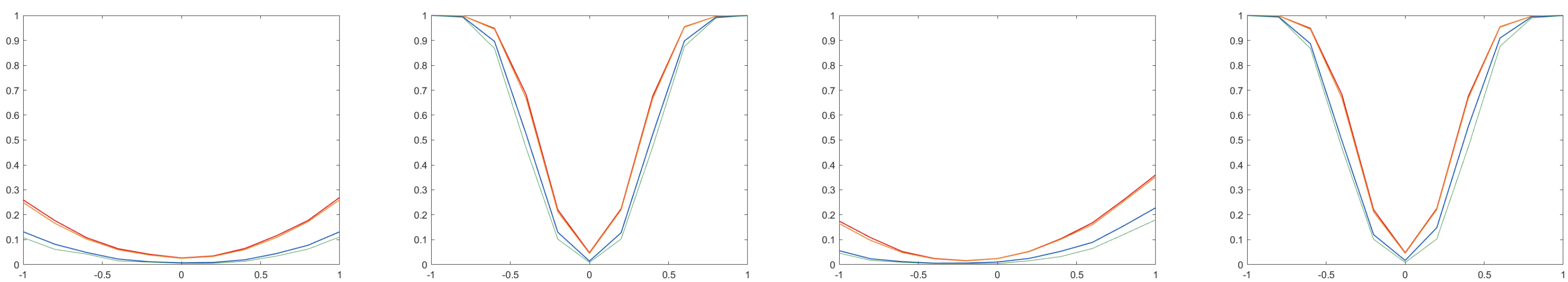
Finally, Figure 5.1c reports the median length of the confidence intervals, computed across S replications. BN1 consistently yields the shortest intervals, with BN2 close behind. The projection method is the most conservative, producing intervals 19–30% longer than BN1. AM lies between BN1 and the projection method, with median lengths 5–18% longer than those of BN1. The differences are most pronounced when θ is near the origin.

6 Empirical Application

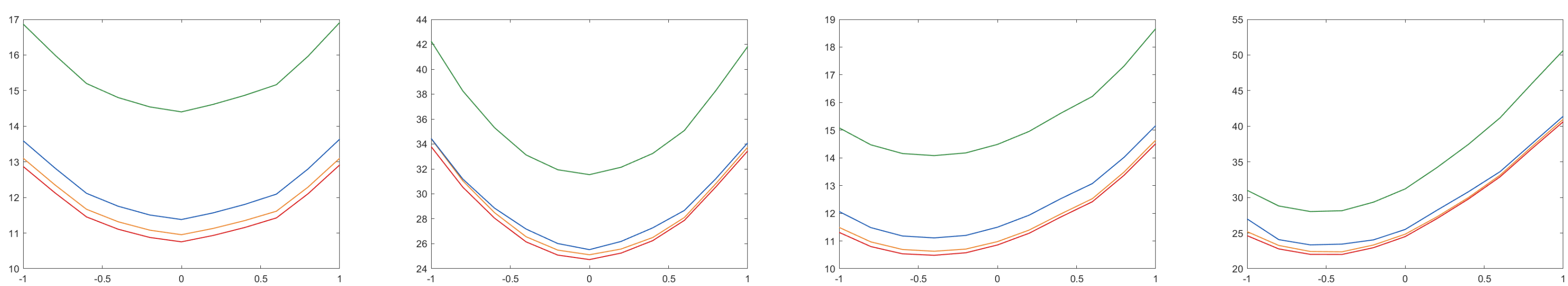
We illustrate the empirical relevance of our results using the setting analyzed by Alan et al. (2018). Their study takes advantage of a distinctive feature of the Turkish education system, in which elementary school teachers are randomly allocated across



(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 5.1: Simulation Results

The horizontal axis represents θ_1/θ_2. The first and second panels report the probability that 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.

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 the 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 traditional 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, the results of Section 3 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 4. van Garderen and van Giersbergen (2024) show that the correlation coefficient between $\hat{\theta}_1$ and $\hat{\theta}_2$ is zero; consequently, both of our methods are uniformly valid.[1]

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

Table 2: Mediation Effect in the data of Alan et al. (2018)

Confidence Intervals are generated by inverting the corresponding tests. Values are rounded to three significant figures.

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).[2] This testing procedure technically does not cover the case where we are testing the 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 4, the confidence intervals based on either the χ^2_1 critical value (BN1) or the two-step procedure (BN2) are uniformly tighter

[1] The validity of BN1 follows from Remark 4.1.

[2] In implementing the test, we follow the empirical application in the working paper version of Andrews and Mikusheva (2016) and only calculate the maximum curvature over a set "close" to the point estimate, adjusting the critical value accordingly.

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_1^2 critical value, for instance, the BN1 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 AM 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 $Q(\chi_2^2, 1-\alpha)$, which accounts for the close similarity between the two sets of intervals.

7 Conclusion

We examine inference in local regions of first-order degeneracy, meaning that the gradient of the transformation is zero or nearly zero so that first-order approximations alone do not provide reliable information and second-order terms must also be considered. In such regions of local degeneracy, we show that neither regular estimation nor quantile-unbiased procedures are feasible, paralleling impossibility results for non-differentiable functionals and ruling out standard approaches to inference. We then develop alternate inference procedures based on minimum-distance statistics that deliver uniformly valid confidence intervals. Simulation studies indicate that these procedures control size while maintaining favorable power, and the empirical application to teacher gender attitudes shows that they yield tighter confidence intervals than existing approaches.

A Proofs and Supporting Results for Section 3

Proof of Proposition 3.1.

Proof. Define $S_n = r_n^2(\Psi_n - g(\theta_\star))$. Via a second order Taylor expansion, we have

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

where in the last line we use the fact that equation (3.4) holds for any $h \in \mathbb{R}^d$ by hypothesis. Since the experiment $\{P_\theta : \theta \in \Theta^\circ\}$ satisfies Assumption 3.1 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 3.2.

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

$$\psi(s) = E_h[\exp(is(\Psi(Z, U) - h'Jh))] \tag{A.1}$$

where, by assumption, $\psi(s)$ does not depend on h. Let $\Phi_h(s) = E_h[\exp(is\Psi(Z, U))]$ and notice that (A.1) 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).

$$\lim_{r \downarrow 0} \frac{1}{r}\left[\Phi_{h_0+rh}(s) - \Phi_{h_0}(s)\right]$$

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

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

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

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

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

Again, since h is arbitrary we can write this

$$\nabla^2\Phi_{h_0}(s) = \Gamma E_{h_0}[\exp(is\Psi(Z,U))(Z-h_0)(Z-h_0)']\Gamma - \Phi_{h_0}(s)\Gamma \tag{A.2}$$

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

$$\begin{aligned}\nabla \exp(isf(h_0)) &= 2is \exp\left(isf(h_0)\right) Jh_0 \\ \nabla^2 \exp\left(isf(h_0)\right) &= \exp\left(isf(h_0)\right)\left(2isJ - 4s^2(Jh_0)(Jh_0)'\right)\end{aligned} \tag{A.3}$$

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 (A.2) and (A.3) yields, for any h_0, that

$$\begin{aligned}&\psi(s)\exp\left(isf(h_0)\right)\left(2isJ - 4s^2(Jh_0)(Jh_0)'\right) \\ &\quad= \Gamma E_{h_0}[\exp(is\Psi(Z,U))(Z-h_0)(Z-h_0)']\Gamma - \Phi_{h_0}(s)\Gamma\end{aligned} \tag{A.4}$$

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 (A.4) is bounded uniformly over $h_0 \in \mathbb{R}^d$. On the other hand, looking at the LHS of (A.4) we can see, using $\|A+B\| \geqslant \|B\| - \|A\|$, that

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

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

$$\alpha = P_{(1+r)h}(\Psi(Z,U) \leqslant ((1+r)h)'J((1+r)h))$$

In particular,

$$0 = \alpha - \alpha = P_{(1+r)h}(\Psi(Z,U) \leqslant ((1+r)h)'J((1+r)h)) - P_h(\Psi(Z,U) \leqslant h'Jh)$$

and thus

$$\begin{aligned}0 = \lim_{r\downarrow 0}\Bigg\{&\frac{1}{r}\left[P_{(1+r)h}(\Psi(Z,U) \leqslant (1+r)^2h'Jh - P_h\left(\Psi(Z,U) \leqslant (1+r)^2h'Jh\right)\right] \\ &+ \frac{1}{r}\left[P_h\left(\Psi(Z,U) \leqslant (1+r)^2h'Jh\right) - P_h\left(\Psi(Z,U) \leqslant h'Jh\right)\right]\Bigg\}\end{aligned} \tag{A.5}$$

Letting $F_h(x) = P_h(\Psi(Z,U) \leqslant x)$ and 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 A.1, we obtain for any h

$$h'\Gamma_{\theta_\star} E_h[\mathbf{1}\{\Psi(X,U) \leqslant h'Jh\}(X-h)] = 2(h'Jh)F_h'(h'Jh)$$

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 $\|E_h[\mathbf{1}\{\Psi(X,U) \leqslant h'Jh\}(X - h)]\| \lesssim 1$ by Cauchy-Schwarz. Meanwhile, $2((ch)'J(ch)) \propto c^2 F_{ch}'((ch)'J(ch))$. Recall $F_h(h'Jh) = P_h(\Psi(Z,U) - h'Jh <= 0)$ and $\Psi(Z,U) - h'Jh \sim \mathcal{L}_h$, $F_h(h'Jh)$ corresponds to the CDF of $\mathcal{L}_h$ evaluated at zero. By assumption, there exists an $\epsilon > 0$ such that $F_h'(h'Jh) \geqslant \epsilon$ for all h. Since $c^2 F_{ch}'((ch)'J(ch)) \to \infty$ as $c \to \infty$ we arrive at a contradiction. □

Proof of Theorem 3.1.

Proof. Theorem 3.1 follows directly from Proposition 3.1 along with Proposition 3.2. □

Proof of Proposition 3.3.

Proof. The first claim follows immediately from the definition of similarity as well as the fact that $\mathcal{P}(h)$ is differentiable at zero. For the second claim, it suffices to show that for an indefinite symmetric $d \times d$ real matrix J the isotropic set $\mathcal{H}_J = \{h \in \mathbb{R}^d : h'Jh = 0\}$ spans $\mathbb{R}^d$. To do so, let us diagonlize $J = Q\Lambda Q'$ where Q is orthogonal satisfying $Q'Q = I$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_d)$ is a $d \times d$ diagonal matrix containing the eigenvalues of J. Because Q is orthogonal, it suffices to show that the isotropic set of Λ, $\mathcal{H}_\Lambda = \{h \in \mathbb{R}^d : h'\Lambda h = 0\}$ spans $\mathbb{R}^d$.

Without loss of generality, let us assume that $\lambda_1 > 0$ and $\lambda_2 < 0$. Let $e_1, \ldots, e_d$ denote the standard basis vectors in $\mathbb{R}^d$. We wish to show that each $e_j \in \text{span}(\mathcal{H}_\Lambda)$ for $j = 1, \ldots, d$. If $\lambda_j = 0$ then trivially $e_j \in \mathcal{H}_\Lambda \subseteq \text{span}(\mathcal{H}_\Lambda)$. Suppose that $\lambda_j > 0$. Define $t_j = (-\lambda_j/\lambda_2)^{1/2}$. Consider $u_j^+ = e_j + t_j e_2$ and $u_j^- = e_j - t_j e_2$. Then, notice that

$$u_j^+ \Lambda u_j^+ = \lambda_j + t_j^2 \lambda_2 = 0 \quad \text{and} \quad u_j^- \Lambda u_j^- = \lambda_j + t_j^2 \lambda_2 = 0$$

so that $u_j^+ \in \mathcal{H}_\Lambda$ and $u_j^- \in \mathcal{H}_\Lambda$. Since $e_j = \frac{1}{2}(u_j^+ + u_j^-)$ it follows that $e_j \in \operatorname{span}(\mathcal{H}_\Lambda)$. The case where $\lambda_j < 0$ follows symmetrically.

The claim in Remark 3.3 follows from Lemma A.3. □

Proof of Corollary 3.1

Proof. Follows directly from Theorem 3.1. □

Lemma A.1. *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 $\alpha \in (0, 1)$,*

$$\alpha = P_h \left(\Psi(Z, U) \leqslant \frac{1}{2} h' \nabla^2 g(\theta_\star) h \right), \qquad \textit{for all } h \in \mathcal{H}.$$

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 $h'\Gamma_{\theta_\star} E_0[\mathbf{1}\{\Psi(Z, U) \leqslant 0\} Z] = 0$ for all $h \in \mathcal{H}$.

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

$$\alpha = P_{rh} \left(\Psi(Z, U) \leqslant (rh)' J (rh) \right).$$

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

$$0 = \alpha - \alpha = P_{rh} \left(\Psi(Z, U) \leqslant (rh)' J (rh) \right) - P_0(\Psi(Z, U) \leqslant 0).$$

and thus

$$\begin{aligned} 0 = \lim_{r \downarrow 0} \Bigg\{ & \frac{1}{r} \left[P_{rh}(\Psi(Z, U) \leqslant (rh)' J (rh)) - P_0 \left(\Psi(Z, U) \leqslant (rh)' J (rh) \right) \right] \\ & + \frac{1}{r} \left[P_0 \left(\Psi(Z, U) \leqslant (rh)' J (rh) \right) - P_0 \left(\Psi(Z, U) \leqslant 0 \right) \right] \Bigg\}. \end{aligned} \tag{A.6}$$

Each of the terms on the RHS of (A.6) 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

$$\begin{aligned}
&\lim_{r\downarrow 0}\frac{1}{r}\big[P_{rh}(\Psi(Z,U)\leqslant(rh)'J(rh))-P_0(\Psi(Z,U)\leqslant(rh)'J(rh))\big]\\
&\quad=\lim_{r\downarrow 0}\int_{[0,1]}\int \mathbf{1}\left\{\Psi(z,u)\leqslant(rh)'J(rh)\right\}\times\frac{1}{r}[\phi(z|rh,\Gamma_{\theta_\star}^{-1})-\phi(z|0,\Gamma_{\theta_\star}^{-1})]\,dz\,du\\
&\quad=\int_{[0,1]}\int \lim_{r\downarrow 0}\mathbf{1}\left\{\Psi(z,u)\leqslant(rh)'J(rh)\right\}\times\frac{1}{r}[\phi(z|rh,\Gamma_{\theta_\star}^{-1})-\phi(z|0,\Gamma_{\theta_\star}^{-1})]\,dz\,du\\
&\quad=\int_{[0,1]}\int \mathbf{1}\{\Psi(z,u)\leqslant 0\}\left(\frac{\partial}{\partial\tilde h}\phi(z|\tilde h,\Gamma_{\theta_\star}^{-1}\right)_{\tilde h=0} h\,dz\,du\\
&\quad=h'\Gamma_{\theta_\star}\left\{\int_{[0,1]}\int \mathbf{1}\{\Psi(z,u)\leqslant 0\}z\phi(z|0,\Gamma_{\theta_\star}^{-1})\,dz\,du\right\}
\end{aligned}$$

Since the 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 (A.6) evaluates to zero. Thus, we obtain for any $h\neq 0$ that

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

which gives the result. □

Lemma A.2. *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 $\mathcal{H}\subset\mathbb{R}^d$,*

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

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

$$\frac{\partial}{\partial r}\mathcal{P}_h(r)\big|_{r=0}=\lim_{r\downarrow 0}\frac{P_{rh}(\Psi(Z,U)\leqslant 0)-P_0(\Psi(Z,U)\leqslant 0)}{r}=0$$

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

$$
\begin{aligned}
\frac{\partial}{\partial r}\mathcal{P}_h(r)\big|_{r=0} &= \lim_{r\downarrow 0}\int_{[0,1]}\int \mathbf{1}\{\Psi(z,u)\leqslant 0\}\times\frac{1}{r}[\phi(z|rh,\Gamma_{\theta_\star}^{-1})-\phi(z|0,\Gamma_{\theta_\star}^{-1})]\,dz\,du\\
&= \int_{[0,1]}\int \mathbf{1}\{\Psi(z,u)\leqslant 0\}\times\lim_{r\downarrow 0}\frac{1}{r}[\phi(z|rh,\Gamma_{\theta_\star}^{-1})-\phi(z|0,\Gamma_{\theta_\star}^{-1})]\,dz\,du\\
&= h'\Gamma_{\theta_\star}\int_{[0,1]}\int \mathbf{1}\{\Psi(z,u)\leqslant 0\}z\phi(z|0,\Gamma_{\theta_\star}^{-1})\,dz\,du\\
&= h'\Gamma_{\theta_\star}E_0[\mathbf{1}\{\Psi(Z,U)\leqslant 0\}Z]\\
&= 0
\end{aligned}
$$

where the final equality comes from Lemma A.1. □

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

Lemma A.3. *Suppose Assumption 3.1 holds at $\theta_\star$ with Fisher information $\Gamma_{\theta_\star}$. Let Ψ_n be a real-valued statistic and consider the test $\Xi_n=\mathbf{1}\{\Psi_n\geqslant 0\}$. Let $\mathcal{P}$ be the local asymptotic power curve defined in (3.6). Suppose that $\Psi_n \overset{h}{\rightsquigarrow} \mathcal{L}_h$ for each $h\in\mathbb{R}^d$, and let F_0 be the CDF of $\mathcal{L}_0$. If F_0 is continuous at zero, then $\mathcal{P}(h)=\lim_{n\to\infty}P_{\theta_{n,h}}(\Xi_n=1)$ for each h, $\mathcal{P}$ is differentiable at 0, and*

$$\nabla\mathcal{P}(0)=\Gamma_{\theta_\star}E_0\big[\mathbf{1}\{\Psi(Z,U)\geqslant 0\}Z\big],$$

where $Z\sim N(0,\Gamma_{\theta_\star}^{-1})$, $U\sim\mathrm{Unif}(0,1)$ independently of Z, and $\Psi(Z,U)$ is a randomized statistic in the Gaussian shift experiment $\{N(h,\Gamma_{\theta_\star}^{-1}):h\in\mathbb{R}^d\}$ such that $\Psi(Z,U)\sim\mathcal{L}_h$ for all h.

Proof. By Theorem 7.10 in van der Vaart (1998), there exists a randomized statistic $\Psi(Z,U)$ in the Gaussian shift experiment $\{N(h,\Gamma_{\theta_\star}^{-1}):h\in\mathbb{R}^d\}$ such that $\Psi(Z,U)\sim\mathcal{L}_h$ for all h. Since F_0 is continuous at zero, $P_0(\Psi(Z,U)=0)=0$, and by mutual absolute continuity of $N(h,\Gamma_{\theta_\star}^{-1})$ in h we also have $P_h(\Psi(Z,U)=0)=0$ for all h.

Thus, by the Portmanteau theorem,

$$\mathcal{P}(h) = \lim_{n\to\infty} P_{\theta_{n,h}}(\Psi_n \geqslant 0) = P_h(\Psi(Z,U) \geqslant 0).$$

Fix $h \in \mathbb{R}^d$ and define $G(z,u) = \mathbf{1}\{\Psi(z,u) \geqslant 0\}$. For $r \in \mathbb{R}$, let P_{rh} denote the law of $Z \sim N(rh, \Gamma_{\theta_\star}^{-1})$. The likelihood ratio satisfies

$$\frac{dP_{rh}}{dP_0}(Z) = \exp\left(rh'\Gamma_{\theta_\star}Z - \tfrac{1}{2}r^2 h'\Gamma_{\theta_\star}h\right),$$

so $\mathcal{P}(rh) = E_0\Big[G(Z,U)\exp\left(rh'\Gamma_{\theta_\star}Z - \tfrac{1}{2}r^2h'\Gamma_{\theta_\star}h\right)\Big]$. Differentiating at $r = 0$ and invoking Lemma 1(a) of Hirano and Porter (2012) to justify exchanging limits and integrals we obtain:

$$\frac{\partial}{\partial r}\mathcal{P}(rh)\bigg|_{r=0} = h'\Gamma_{\theta_\star}E_0\big[G(Z,U)Z\big].$$

Since the right-hand side is linear in h, $\mathcal{P}$ is differentiable at 0 with gradient $\nabla\mathcal{P}(0) = \Gamma_{\theta_\star}E_0[G(Z,U)Z]$. □

B Proofs for Section 4

B.1 Proofs of the Main Theorems

We first introduce notation for the results in Section 4.1. Let $\mathcal{C}_u(c)$ and $\mathcal{C}_\ell(c)$ be the upper and lower boundaries of $\partial\mathcal{S}^+(\tau,c)$,

$$\mathcal{C}_u(c) = \Big\{\big(C_{u,1}(x_1,c), C_{u,2}(x_1,c)\big) : x_1 \in \mathbb{R}\backslash(-x_1^*, x_1^*)\Big\} \tag{B.1}$$

where

$$C_{u,1}(x_1,c) = x_1 - \frac{c(1-\rho)x_1}{\sqrt{(1+\rho)^2X_2(x_1)^2 + (1-\rho)^2x_1^2}},$$
$$C_{u,2}(x_1,c) = X_2(x_1) + \frac{c(1+\rho)X_2(x_1)}{\sqrt{(1+\rho)^2X_2(x_1)^2 + (1-\rho)^2x_1^2}},$$

$$x_1^* = \begin{cases} \frac{\sqrt{c^2(1-\rho)^2-(1+\rho)\tau}}{\sqrt{2}\sqrt{1-\rho}} & \text{if } \tau \leqslant \frac{c^2(1-\rho)^2}{1+\rho} \\ 0 & \text{otherwise} \end{cases},$$

and the lower boundary is

$$\mathcal{C}_\ell(c) = \left\{ \left(C_{\ell,1}(x_1, c), C_{\ell,2}(x_1, c) \right) : x_1 \in \mathbb{R} \right\}, \tag{B.2}$$

where

$$C_{\ell,1}(x_1, c) = x_1 + \frac{c(1-\rho)x_1}{\sqrt{(1+\rho)^2 X_2(x_1)^2 + (1-\rho)^2 x_1^2}},$$
$$C_{\ell,2}(x_1, c) = X_2(x_1) - \frac{c(1+\rho)X_2(x_1)}{\sqrt{(1+\rho)^2 X_2(x_1)^2 + (1-\rho)^2 x_1^2}}.$$

Details on the calculation of $\mathcal{C}_\ell$ and $\mathcal{C}_u$ are given in Lemma B.2.

If $\tau \leqslant \frac{c^2(1-\rho)^2}{1+\rho}$, let the kink of $\mathcal{C}_u(\tau, c)$ be

$$K = \left(0, \frac{\sqrt{2}\sqrt{c^2(1-\rho)+\tau}}{\sqrt{1-\rho^2}} \right). \tag{B.3}$$

Let $\bar{r}(\theta_1)$ denote the distance between $O = (\theta_1, X_2(\theta_1))$ and K,

$$\bar{r}(\theta_1) = d\left((\theta_1, X_2(\theta_1)), K \right). \tag{B.4}$$

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 $\overset{\frown}{AB}$ be the arc from A to B, and $\overline{AB}$ the line segment.

Proof of Proposition 4.1.

Proof. This follows from Proposition B.1 and B.2. □

Proposition B.1. *Let $\tau \geqslant \frac{c^2(1-\rho)^2}{1+\rho}$ and $|\rho| < 1$. For all $\theta = (\theta_1, \theta_2) \in \mathcal{S}_0^+(\tau)$, $(Z_1, Z_2) \sim N(0, I_2)$,*

$$P\left((Z_1 + \theta_1, Z_2 + \theta_2) \in \mathcal{S}^+(\tau, c) \right) \geqslant 1 - \alpha.$$

where

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

Proof. The proof is based on Lemma B.1, with $\bar{\mathcal{S}}=\mathcal{S}^{+}(\tau,c)$. Condition 1 of Lemma B.1 holds trivially. We now verify Condition 2 of Lemma B.1. Let $r>c$. By Lemma B.5, $\partial B\left(\theta,r\right)$ intersects $\mathcal{S}_0^{+}(\tau)$ at a minimum of two points I and J, with I to the left of J. See Figure B.1. By Lemma B.4.1, there is a point P in $\partial B\left(\theta,r\right)$ above curve $\mathcal{C}_u(\tau,c)$. Therefore, there is at least one point on $\partial B\left(\theta,r\right)$ between P and I that intersects $\mathcal{C}_u(\tau,c)$. Let the closest point (if there's more than one point) to I be point A. Similarly, define B as the point on $\partial B\left(\theta,r\right)$ between P and J that intersects $\mathcal{C}_u(\tau,c)$ and is closest to J. By Lemma B.4.2, there is a point Q on $\partial B\left(\theta,r\right)$ below curve $\mathcal{C}_\ell(\tau,c)$. Therefore, there is at least one point on $\partial B\left(\theta,r\right)$ between Q and I that intersects $\mathcal{C}_\ell(\tau,c)$. Let the closest point (if there's more than one point) to I be point C. Similarly define D between Q and J. Therefore, by construction $A\overset{\frown}{I}C\subset\mathcal{S}^{+}(\tau,c)$ and $B\overset{\frown}{J}D\subset\mathcal{S}^{+}(\tau,c)$.

To show that length $\left(A\overset{\frown}{I}C\right)$ + length $\left(B\overset{\frown}{J}D\right)\geqslant 4r\arcsin\frac{c}{r}$, 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 $B(G,c)\nsubseteq\mathcal{S}^{+}(\tau,c)$, which contradicts the definition of $\mathcal{S}^{+}(\tau,c)$. Therefore, Condition 1 and 2 of Lemma B.1 hold, and

$$P\left((Z_1+\theta_1,Z_2+\theta_2)\in\mathcal{S}^{+}(\tau,c)\right)\geqslant 1-\alpha. \tag{B.5}$$

This completes the proof. ∎

Proposition B.2. *Let* $0<\tau<\frac{(1-\rho)^2c^2}{1+\rho}$ *and* $\rho\in[0,1)$. *For all* $\theta=(\theta_1,\theta_2)\in\mathcal{S}_0(\tau)$, $(Z_1,Z_2)\sim N(0,I_2)$,

$$P\left((Z_1+\theta_1,Z_2+\theta_2)\in\mathcal{S}(\tau,c)\right)\geqslant 1-\alpha.$$

Proof. The proof is based on Lemma B.1 with $\bar{\mathcal{S}}=\mathcal{S}(\tau,c)$. WLOG, let $\theta\in\mathcal{S}_0^{+}(\tau)$.

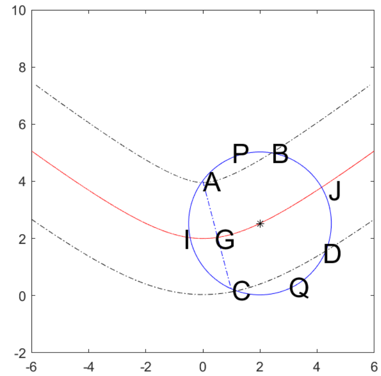


Figure B.1: Illustration of Proposition B.1.

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

By Lemma B.4.1 and B.4.2, for all $|\theta_1| \geqslant x_1^*$, the coverage is at least $1 - \alpha$ with the same argument as in Proposition B.1.

For $\theta_1 \in (-x_1^*, x_1^*)$. If $r > \bar{r}(\theta_1)$, where $\bar{r}(\theta_1)$ is defined in (B.4), $\partial B\left(O, r\right)$ intersects with $\mathcal{C}_u$ at least two points, since (i) by Lemma B.4.3,

$$\inf_{x \in \mathcal{C}_u} d\left(x, \theta\right) = d\left(K, \theta\right) = \bar{r}(\theta_1) < r,$$

(ii)

$$\sup_{x \in \mathcal{C}_u, x_1 > 0} d\left(x, \theta\right) = \sup_{x \in \mathcal{C}_u, x_1 < 0} d\left(x, \theta\right) = \infty.$$

Therefore

$$\left|\text{length}\left(\partial B\left(\theta, r\right) \cap \mathcal{S}^+(\tau, c)\right)\right| \geqslant 4r \arcsin\frac{c}{r}$$

following from the same argument in Proposition B.1.

Then we show that for $r \in (c, r(\theta_1)]$,

$$\left|\text{length}\left(\partial B\left(\theta, r\right) \cap \bar{\mathcal{S}}\right)\right| = 2\pi r \geqslant 4r \arcsin\frac{c}{r}. \tag{B.6}$$

By Lemma B.4.4, $\partial B\left(\theta, r\right) \cap \mathcal{C}_\ell(c) \cap \left\{(0, x) : x \geqslant 0\right\} = \varnothing$. By Lemma B.4.3, $\partial B\left(\theta, r\right) \cap \mathcal{C}_u(c) = \varnothing$. Therefore, $B\left(\theta, r\right) \cap \left\{(0, x) : x \geqslant 0\right\} \subseteq \mathcal{S}^+(\tau, c)$. Case 1: if

$$B\left(\theta, r\right) \cap \left\{(0, x_2) : x_2 \leqslant 0\right\} = \varnothing,$$

then (B.6) holds. Case 2: if $\exists(x_1, 0) \in \partial B(\theta, r)$, by Lemma B.6.2,

$$B(\theta, r) \cap \left\{(0, x) : x \leqslant 0\right\} \subseteq \mathcal{S}^-(\tau, c).$$

In sum, $\partial B(\theta, r) \subseteq \mathcal{S}(\tau, c)$. □

Proof of Theorem 4.1.

Proof. There exists a subsequence $P_{n_j} \in \mathcal{P}_n$ such that

$$\liminf_n \inf_{P \in \mathcal{P}_n} P\left(\hat{T}_n(g(\theta_P)) \leqslant c^2\right) = \lim_{j \to \infty} P_{n_j}\left(\hat{T}_n(g(\theta_{n_j})) \leqslant c^2\right) \text{ where } \theta_n = \theta_{P_n}.$$

Since Θ is compact, the sequence $\{\theta_{n_j}\}$ is bounded and thus has a further subsequence $n_{j'}$ such that $\lim_{n_{j'}} \theta_{n_{j'}} = \theta_\infty \in \Theta$, $\lim_{n_{j'}} r_{n_{j'}}(\theta_{n_{j'}} - \theta_\star) = h \in \mathbb{R}^2_{[\pm\infty]}$, $\lim_{n_{j'}} R_{P_{n_j}} = R$, and $\rho_{P_n} \to \rho$. With slight abuse of notation, we will refer to this convergent subsequence as $\{\theta_n\}$ from here on. In addition, we use n instead of P_n for subscript in λ, R, and ρ.

Case 1. If $\lim_n \theta_n = \theta_\infty \neq \theta_\star$, standard minimum distance arguments (see, for example Section 9.1 in Newey and McFadden (1994)) imply that

$$\lim_n P_n\left(\hat{T}_n(g(\theta_n)) \leqslant Q(\chi^2_1, 1 - \alpha)\right) \geqslant 1 - \alpha.$$

Case 2. Suppose $\theta_\infty = \theta_\star$ and $\lim_n r_n(\theta_n - \theta_\star) = h \in \mathbb{R}^2$. We first normalize the problem to match the notation in Proposition 4.1. WLOG, assume $\lambda_{n,2} \geqslant \lambda_{n,1}$. There exists an orthogonal matrix R_n such that

$$\operatorname{sign}(g(\theta_n) - g(\theta_\star))\Sigma_n^{1/2} H \Sigma_n^{1/2} = R_n' \begin{bmatrix} \lambda_{n,1} & 0 \\ 0 & \lambda_{n,2} \end{bmatrix} R_n = \frac{\lambda_{n,2} - \lambda_{n,1}}{2} R_n' \begin{bmatrix} \rho_n - 1 & 0 \\ 0 & \rho_n + 1 \end{bmatrix} R_n.$$

Define

$$\vartheta_n = R_n \Sigma_n^{-1/2} \theta_n,\ \vartheta_{\star,n} = R_n \Sigma_n^{-1/2} \theta, \quad \tilde{g}_n(\vartheta) = \frac{\operatorname{sign}(g(\theta_n) - g(\theta_\star))}{\lambda_{n,2} - \lambda_{n,1}} g\left(\Sigma_n^{1/2} R_n' \vartheta\right).$$

By construction,

$$\frac{\partial \tilde{g}(\vartheta_{\star,n})}{\partial \vartheta} = 0, \; \frac{\partial^2 \tilde{g}(\vartheta_{\star,n})}{\partial \vartheta \, \partial \vartheta'} = \frac{\operatorname{sign}(g(\theta_n) - g(\theta_\star))}{\lambda_{n,2} - \lambda_{n,1}} R_n \Sigma_n^{1/2} H \Sigma_n^{1/2} R_n' \quad = \frac{1}{2} \begin{bmatrix} \rho_n - 1 & 0 \\ 0 & \rho_n + 1 \end{bmatrix}.$$

Let $\hat{\vartheta}_n = R_n \Sigma^{-1/2} \hat{\theta}_n$. Under Assumption 4.1, it holds that $r_n(\hat{\vartheta}_n - \vartheta_n) \xrightarrow{d} N(0, I_2)$ uniformly when $\lim_n r_n(\theta_n - \theta_\star) = h \in \mathbb{R}^2$. Let $\tilde{H} = \begin{bmatrix} \rho - 1 & 0 \\ 0 & \rho + 1 \end{bmatrix}$.

Under Assumption 4.1, by the almost sure representation theorem, there exists a probability space with random variables $\mathbb{Z}_n$ and $\mathbb{Z}$ defined on it such that (i) $\mathbb{Z}_n$ has the same distribution as $r_n(\hat{\vartheta}_n - \vartheta_n)$, (ii) $\mathbb{Z} \sim N(0, I_2)$, and (iii) $\mathbb{Z}_n \xrightarrow{\text{a.s.}} \mathbb{Z}$. Define $\tilde{h}_n = r_n(\vartheta_n - \vartheta_{\star,n})$,

$$\begin{aligned}
\hat{T}_n(g(\theta_n)) &= \inf_{\theta: g(\theta) = g(\theta_n)} \left\| r_n R_n \hat{\Sigma}^{-1/2} (\hat{\theta} - \theta) \right\|^2 \\
&= \inf_{\theta: g(\theta) = g(\theta_n)} \left\| r_n R_n \Sigma^{-1/2} (\hat{\theta} - \theta) \right\|^2 + o_p(1) \\
&= \inf_{\vartheta: \tilde{g}_n(\vartheta_{\star,n} + \vartheta - \vartheta_{\star,n}) = \tilde{g}_n(\vartheta_n)} \left\| r_n(\hat{\vartheta} - \vartheta_n) + \tilde{h}_n - r_n(\vartheta - \vartheta_{\star,n}) \right\|^2 + o_p(1) \\
&= \inf_{\vartheta: \tilde{g}_n(\vartheta_{\star,n} + r_n^{-1} x) = \tilde{g}_n(\vartheta_n)} \left\| r_n(\hat{\vartheta} - \vartheta_n) - (x - \tilde{h}_n) \right\|^2 + o_p(1).
\end{aligned}$$

The second line follows from $r_n(\hat{\theta} - \theta) = O_p(1)$, with θ denoting the optimizer, and from the consistency of $\hat{\Sigma}$. The third and fourth lines follow from rearranging terms.

It follows that $\hat{T}_n \sim T_n + o_p(1)$, where

$$T_n = \inf_{\vartheta: \tilde{g}_n(\vartheta_{\star,n} + r_n^{-1} x) = \tilde{g}_n(\vartheta_n)} \left\| \mathbb{Z}_n - (x - \tilde{h}_n) \right\|^2.$$

Let

$$T = \inf_{x: x' \tilde{H} x = \tilde{h}' \tilde{H} \tilde{h}} \left\| \mathbb{Z} - (x - h) \right\|^2. \tag{B.7}$$

By Lemma B.7, $T_n = T + o_p(1)$. Moreover, by Lemma B.11, T is continuously distributed. Hence

$$\lim_n P_n(T_n \leqslant c^2) = P(T \leqslant c^2) \geqslant 1 - \alpha,$$

where the inequality follows from Proposition 4.1 with condition either

$$\begin{aligned}\tilde{h}'\tilde{H}\tilde{h} &= \lim_n 4r_n^2 \left(\tilde{g}(\vartheta_n) - \tilde{g}(\vartheta_\star)\right) \\ &= \lim_n \frac{4\mathrm{sign}(g(\theta_n) - g(\theta_\star))}{\lambda_{n,2} - \lambda_{n,1}} r_n^2 \left(g(\theta_n) - g(\theta_\star)\right) \geqslant \frac{c^2(1-\rho)^2}{1+\rho}\end{aligned}$$

or $\rho \in [0, 1-\eta]$.

Case 3. Suppose $\theta_n \to \theta_\star$ and $\lim_n r_n(\theta_n - \theta_\star) \to \infty$. Define $s_n = \frac{1}{||\theta_n - \theta_\star||} \ll r_n$. By construction, $\left\|s_n(\theta_n - \theta_\star)\right\| = 1$, so there exists a subsequence such that $\lim s_n(\theta_n - \theta_\star) = \lim h_n = h \in \mathbb{R}^d \backslash \{0_d\}$. Similar to Case 2, let $\mathbb{Z}_n$ has the same distribution as $\hat{\Sigma}^{-1/2} r_n(\hat{\theta}_n - \theta_n)$,

$$T_n = \inf_{x: g(\theta_n + r_n^{-1}x) = g(\theta_n)} \left\| \mathbb{Z}_n - \hat{\Sigma}^{-1/2} x \right\|^2.$$

Note that $\hat{T}_n \sim T_n$. By Lemma B.8, $T_n = T + o_p(1)$, where

$$T = \inf_{x: h'Hx = 0} \left\| \mathbb{Z} - \Sigma_n^{-1/2} x \right\|^2. \tag{B.8}$$

Since $h'H \neq 0$, $T \sim \chi_1^2$, it holds that

$$\lim_n P_n(T_n \leqslant c^2) = P(T \leqslant c^2) = 1 - \alpha.$$

□

Proof of Theorem 4.2.

Proof. There exists a subsequence $P_{n_j} \in \mathcal{P}_{n_j}$ such that

$$\liminf_n \inf_{P \in \mathcal{P}_n} P\left(\hat{T}_n(g(\theta_n)) \leqslant \hat{c}\right) = \lim_j P_{n_j}\left(\hat{T}_n(g(\theta_{n_j})) \leqslant \hat{c}\right).$$

Since Θ and $\mathcal{S}$ are compact, the sequences $\{\theta_{n_j}\}$ and Σ_{n_j} have further subsequences n_k such that $\lim_k \theta_{n_k} = \theta_\infty \in \Theta$, $\lim_{n_k} r_{n_k}(\theta_{n_k} - \theta_\star) \to h \in \mathbb{R}^d_{[\pm\infty]}$, and $\lim_k \Sigma_{n_k} \to \Sigma \in \mathcal{S}$.

With slight abuse of notation, we will refer to this convergent subsequence as $\{n\}$ from here on.

Case 1. If $\lim_n \theta_n = \theta_\infty \neq \theta_\star$, standard minimum distance arguments apply and will show that $\hat{T}_n(g(\theta_n)) \overset{P_n}{\rightsquigarrow} \chi_1^2$. Let $z \in \mathcal{H}_z$. By construction, $\hat{h}_n = r_n(\hat{\theta} - \theta_\star) - \hat{\Sigma}_n^{1/2} z \in \mathcal{H}$. In addition, $\hat{h}_n / r_n \xrightarrow{p} (\theta_\infty - \theta_\star) \neq 0$, and since H has full rank, $\frac{\hat{h}_n' H}{||\hat{h}_n||} \xrightarrow{p} \frac{(\theta_\infty - \theta_\star)' H}{||\theta_\infty - \theta_\star||} \neq 0$. By Lemma B.9,

$$\hat{T}_n^*(\hat{h}_n) = \inf_{h'x=0} \left\| \mathbb{Z} - \Sigma^{-1/2} x \right\|^2 + o_p(1). \tag{B.9}$$

The critical value

$$\begin{aligned}\hat{c} &\geqslant Q\left(\hat{T}_n^*(\hat{h}_n) \middle| \mathbb{Z} \in \mathcal{H}_z; \frac{1-\alpha}{1-\eta} \right) \\ &= Q\left(\inf_{h'x=0} \left\| \mathbb{Z} - \Sigma^{-1/2} x \right\|^2 \middle| \mathbb{Z} \in \mathcal{H}_z; \frac{1-\alpha}{1-\eta} \right) + o_p(1)\end{aligned}$$

where the inequality follows from $\hat{h}_n \in \mathcal{H}$ and the equality follows from (B.9) and the continuity of $\inf_{h'x=0} \left\| \mathbb{Z} - \Sigma^{-1/2} x \right\|^2$. By Lemma B.12,

$$Q\left(\inf_{h'x=0} \left\| \mathbb{Z} - \Sigma^{-1/2} x \right\|^2 \middle| \mathbb{Z} \in \mathcal{H}_z; \frac{1-\alpha}{1-\eta} \right) \in \left[Q(\chi_1^2, 1-\alpha), q_{\chi_1^2, 1-\alpha+\eta} \right).$$

By the continuity of the limit distribution of $\hat{T}_n$, it holds that

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

Case 2. $\theta_\infty = \theta_\star$ and $\lim_n r_n(\theta_n - \theta_\star) = h \in \mathbb{R}^d$. Similar to the proof of Theorem 4.1 Case 2, $\hat{T}_n(g(\theta_n)) \sim T + o_p(1)$, $\hat{T}_n^*(h_n) \sim T + o_p(1)$, where $h_n = r_n(\theta_n - \theta_\star)$,

$$T = \inf_{x: x'Hx = h'Hh} \left\| \mathbb{Z} - \Sigma^{-1/2}(x - h) \right\|^2.$$

By Lemma B.11, T is continuously distributed, thus $\hat{c} \xrightarrow{p} Q(T | \mathbb{Z} \in \mathcal{H}_z, \frac{1-\alpha}{1-\eta})$. Note that $h_n \in \mathcal{H}$ is equivalent to $r_n(\hat{\theta}_n - \theta_n) \in \mathcal{H}_z$. To see the coverage rate,

$$P\left(\hat{T}_n(g(\theta_n)) \leqslant \hat{c} \right) \geqslant P\left(\hat{T}_n(g(\theta_n)) \leqslant \hat{c}, h_n \in \mathcal{H} \right)$$

$$
\begin{aligned}
&= P\left(\hat{T}_n(g(\theta_n)) \leqslant c, \mathbb{Z}_n \in \mathcal{H}_z\right) + o(1)\\
&= P\left(T \leqslant c \middle| \mathbb{Z} \in \mathcal{H}_z\right) P\left(\mathbb{Z} \in \mathcal{H}_z\right) + o(1)\\
&= \frac{1-\alpha}{1-\eta}(1-\eta) + o(1).
\end{aligned}
$$

Case 3. If $\theta_n \to \theta_\star$ and $\lim_n r_n(\theta_n - \theta_\star) \to \infty$. Similar to the proof of Theorem 4.1 Case 3, $\hat{T}_n(g(\theta_n)) \sim T + o_p(1)$, $\hat{T}_n^*(h_n) \sim T + o_p(1)$, where T is defined in (B.8), which is χ_1^2. The same argument as in Case 1 applies here, together with Lemma B.12, it holds that

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

The lower bound is binding when $k = 1$. □

B.2 Lemmas

Lemma B.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 $\hat{\theta} - \theta \sim N(0, I_2)$. Then

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

Proof. The key idea of Lemma B.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 $\hat{\theta} = (\theta_1 + r\cos\omega, \theta_2 + r\sin\omega)$,

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

$$= \int_{r=0}^{c} \exp(-\frac{r^2}{2}) r dr + \int_{r=c}^{+\infty} \frac{4 \arcsin \frac{c}{r}}{2\pi} \exp(-\frac{r^2}{2}) r dr. \tag{B.10}$$

To see (B.10), note that if $r \leqslant c$, then $(r \sin \omega)^2 \leqslant c^2$ 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,

$$\begin{aligned}
P\left(\hat{\theta} \in \bar{\mathcal{S}}\right) &= \frac{1}{2\pi} \int_{r=0}^{+\infty} \frac{1}{r} \left| \text{length} \left(\partial B(\theta, r) \cap \bar{\mathcal{S}}\right) \right| \exp(-\frac{r^2}{2}) r dr \\
&\geqslant \int_{r=0}^{c} \exp(-\frac{r^2}{2}) r dr + \int_{r=c}^{+\infty} \frac{4 \arcsin \frac{c}{r}}{2\pi} \exp(-\frac{r^2}{2}) r dr. \qquad \text{(B.11)}
\end{aligned}$$

This lower bound matches the expression in (B.10), which completes the proof. An illustration of Lemma B.1 is provided in Figure B.2. □

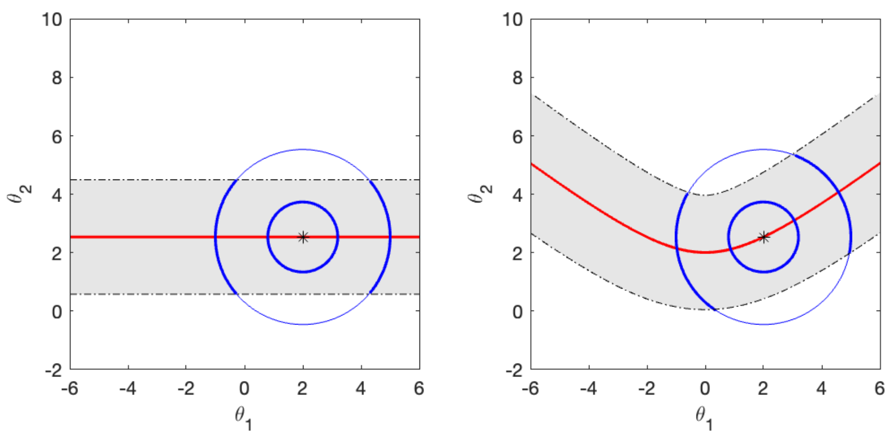


Figure B.2: Lemma B.1: Acceptance Region of Linear and Curved Null.

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, for 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 $1 - \alpha$.

Lemma B.2. *The boundary $\partial\mathcal{S}^+(\tau, c)$ can be characterized by two curves. The upper one is $\mathcal{C}_u(c)$ defined in (B.1) and the lower one is $\mathcal{C}_\ell(c)$ defined in (B.2).*

Proof. $\mathcal{C}_u(c)$ and $\mathcal{C}_\ell(c)$ are obtained by shifting $\mathcal{S}_0^+(\tau)$ a distance c along its normal direction. Note that with $\tau \in (0, \frac{c^2(1-\rho)^2}{1+\rho}]$, for all $x \in (-x_1^*, x_1^*)$, $\big(C_{u,1}(x_1, c), C_{u,2}(x_1, c)\big)$ is an interior point of $\mathcal{S}^+(\tau, c)$, thus not included in $\mathcal{C}_u(c)$. By construction, $\mathcal{C}_u(c) \cup \mathcal{C}_\ell(c) \subseteq \mathcal{S}(\tau, c)$.

Then we show that $\mathcal{C}_u$ and $\mathcal{C}_\ell$ are the boundaries of $\mathcal{S}^+(\tau, c)$. First, we show that $\mathcal{C}_u$ is convex. Let $C'_{u,1}$ and $C''_{u,1}$ be the first and second order derivatives of $C_{u,1}$ with respect to x,

$$
\begin{aligned}
\frac{\mathrm{d}^2 C_{u,2}}{\mathrm{d}C_{u,1}^2} &= \frac{C''_{u,2}C'_{u,1} - C'_{u,2}C''_{u,1}}{\left(C'_{u,1}\right)^3} \\
&= \frac{(1-\rho)\tau\left((\rho+1)\tau + 2(1-\rho)x_1^2\right)^{3/2}}{\sqrt{1+\rho}\left((1-\rho)x_1^2+\tau\right)^{3/2}}\left(\left((1+\rho)\tau + 2(1-\rho)x_1^2\right)^{3/2} - c\left(1-\rho^2\right)\tau\right)^{-1}.
\end{aligned}
$$

The sign of $\frac{\mathrm{d}^2 C_{u,2}}{\mathrm{d}C_{u,1}^2}$ is the same as $\left((1+\rho)\tau + 2(1-\rho)x_1^2\right)^{3/2} - c\left(1-\rho^2\right)\tau$. If $\tau \geqslant \frac{c^2(1-\rho)^2}{1+\rho}$, then

$$
\begin{aligned}
&\left((1+\rho)\tau + 2(1-\rho)x_1^2\right)^{3/2} - c\left(1-\rho^2\right)\tau \geqslant \left((1+\rho)\tau + 0\right)^{3/2} - c\left(1-\rho^2\right)\tau \\
&\quad = \tau(1+\rho)\left((1+\rho)^{1/2}\tau^{1/2} - c\left(1-\rho\right)\right) \\
&\quad \geqslant \tau(1+\rho)\left((1+\rho)^{1/2}\frac{c(1-\rho)}{\sqrt{1+\rho}} - c\left(1-\rho\right)\right) = 0.
\end{aligned}
$$

If $\tau \in \left[0, \frac{c^2(1-\rho)^2}{1+\rho}\right]$,

$$
\begin{aligned}
&\left((1+\rho)\tau + 2(1-\rho)x_1^2\right)^{3/2} - c\left(1-\rho^2\right)\tau \geqslant \left((1+\rho)\tau + 2(1-\rho)x_1^{*2}\right)^{3/2} - c\left(1-\rho^2\right)\tau \\
&\quad = c^3(1-\rho)^3 - c\left(1-\rho^2\right)\tau \geqslant 0.
\end{aligned}
$$

Thus $\frac{\mathrm{d}^2 C_{u,2}}{\mathrm{d}C_{u,1}^2} \geqslant 0$ and $\mathcal{C}_u$ is convex.

Next, by Lemma B.3, the connecting line of $\big(C_{u,1}(x_1, c), C_{u,2}(x_1, c)\big)$ and $(x_1, X_2(x_1))$ is orthogonal to the tangent line of $\mathcal{C}_u$ at $\big(C_{u,1}(x_1, c), C_{u,2}(x_1, c)\big)$. In addition, since

$\mathcal{C}_u$ is convex, $\mathcal{C}_u$ is above its tangent line. This implies that $\forall x_1$

$$d\left((x_1, X_2(x_1)), \mathcal{C}_u\right) = d\Big(\big(x_1, X_2(x_1)\big), \big(C_{u,1}(x_1,c), C_{u,2}(x_1,c)\big)\Big) = c. \tag{B.12}$$

This also implies that for all $(C_{u,1}(x_1,c), C_{u,2}(x_1,c)) \in \mathcal{C}_u$,

$$c \geqslant d\left((C_{u,1}(x_1,c), C_{u,2}(x_1,c)), \mathcal{S}_0^+(\tau)\right) \geqslant c,$$

where the first equality follows from the fact that

$$d\left((x_1, X_2(x_1)), (C_{u,1}(x_1,c), C_{u,2}(x_1,c))\right) = c$$

and the second inequality follows from (B.12), i.e.

$$d\Big(\big(C_{u,1}(x_1,c), C_{u,2}(x_1,c)\big), \mathcal{S}_0^+(\tau)\Big) \geqslant d\left(\mathcal{C}_u, \mathcal{S}_0^+(\tau)\right) = \inf_{x_1} d\left(\mathcal{C}_u, (x_1, X_2(x_1))\right) = c.$$

Therefore, $\mathcal{C}_u$ is the upper part of the boundary of $\mathcal{S}^+(\tau, c)$.

To show that $\mathcal{C}_\ell$ is on the boundary of $\mathcal{S}_0^+(\tau)$, note that by Lemma B.3, the connecting line of $\big(C_{\ell,1}(x_1,c), C_{\ell,2}(x_1,c)\big)$ and $(x_1, X_2(x_1))$ is orthogonal to the tangent line of $\mathcal{S}_0^+(\tau)$ at $\big(x_1, X_2(x_1)\big)$. In addition, since

$$\frac{\mathrm{d}^2 X_2(x_1)}{\mathrm{d}x_1^2} = \frac{(1-\rho)\tau}{\sqrt{1+\rho}\left(\tau + (1-\rho)x_1^2\right)^{3/2}} \geqslant 0,$$

$\mathcal{S}_0^+(\tau)$ is convex. Thus $\mathcal{S}_0^+(\tau)$ is above its tangent line. This implies that

$$d\Big(\big(C_{\ell,1}(x_1,c), C_{\ell,2}(x_1,c)\big), \mathcal{S}_0^+(\tau)\Big) = c.$$

Thus $\mathcal{C}_\ell$ is the lower part of the boundary of $\mathcal{S}^+(\tau, c)$. □

Lemma B.3. *Let $x_1 \in \mathbb{R}\backslash(-x_1^*, x_1^*)$.*

1. *The perpendicular bisector of $\big(C_{\ell,1}(x_1,c), C_{\ell,2}(x_1,c)\big)$ and $\big(C_{u,1}(x_1,c), C_{u,2}(x_1,c)\big)$ is tangent to $\mathcal{S}_0^+(\tau)$ at $(x_1, X_2(x_1))$.*
2. *The connecting line of $\big(C_{u,1}(x_1,c), C_{u,2}(x_1,c)\big)$ and $\big(C_{\ell,1}(x_1,c), C_{\ell,2}(x_1,c)\big)$ is orthogonal to the tangent line of $\mathcal{C}_u$ and $\mathcal{C}_\ell$ at these points.*

Proof. To show 1. It is easy to verify that

$$\frac{1}{2}\big(C_{\ell,1}(x_1,c)+C_{u,1}(x_1,c)\big)=x_1,$$
$$\frac{1}{2}\big(C_{\ell,2}(x_1,c)+C_{u,2}(x_1,c)\big)=X_2(x_1).$$

In addition,

$$\frac{C_{u,2}(x_1,c)-C_{\ell,2}(x_1,c)}{C_{u,1}(x_1,c)-C_{\ell,1}(x_1,c)}=-\frac{(1+\rho)X_2(x_1)}{(1-\rho)x_1},\quad \frac{\mathrm{d}X_2(x_1)}{\mathrm{d}x_1}=\frac{(1-\rho)x_1}{(1+\rho)X_2(x_1)}$$
$$\Rightarrow\frac{C_{u,2}(x_1,c)-C_{\ell,2}(x_1,c)}{C_{u,1}(x_1,c)-C_{\ell,1}(x_1,c)}\frac{\mathrm{d}X_2(x_1)}{\mathrm{d}x_1}=-1.$$

This completes the proof.

To show 2, straightforward calculation shows that

$$\frac{\mathrm{d}C_{u,2}}{\mathrm{d}C_{u,1}}=\frac{C'_{u,2}}{C'_{u,1}}=\frac{\mathrm{d}X_2(x_1)}{\mathrm{d}x_1},$$
$$\frac{\mathrm{d}C_{\ell,2}}{\mathrm{d}C_{\ell,1}}=\frac{C'_{\ell,2}}{C'_{\ell,1}}=\frac{\mathrm{d}X_2(x_1)}{\mathrm{d}x_1}.$$

By the first part, the connecting line of $\big(C_{u,1}(x_1,c),C_{u,2}(x_1,c)\big)$ and $\big(C_{\ell,1}(x_1,c),C_{\ell,2}(x_1,c)\big)$ is orthogonal to the tangent line of $\mathcal{S}_0^+(\tau)$ at $(x_1,X_2(x_1))$. Therefore, it is also orthogonal to the tangent line of $\mathcal{C}_i(c)$ at $\big(C_{i,1}(x_1,c),C_{i,2}(x_1,c)\big)$ with $i=u,\ell$. □

Lemma B.4. *Let $r>c$, $\theta=(\theta_1,\theta_2)\in\mathcal{S}_0^+(\tau)$.*

1. *If $\theta_1\notin(-x_1^*,x_1^*)$, there exists $(x_1,x_2^{(1)})\in\partial B\big((\theta_1,\theta_2),r\big)$, $(x_1,x_2^{(2)})\in\mathcal{C}_u(\tau,c)$ such that $x_2^{(1)}>x_2^{(2)}$.*
2. *There exists $(x_1,x_2^{(1)})\in\partial B\big((\theta_1,\theta_2),r\big)$, $(x_1,x_2^{(2)})\in\mathcal{C}_\ell(\tau,c)$ such that $x_2^{(1)}<x_2^{(2)}$.*
3. *If $\tau\leqslant\frac{c^2(1-\rho)^2}{1+\rho}$, $\theta_1\in(-x_1^*,x_1^*)$, then*

$$d\big(\theta,\mathcal{C}_u(c)\big)=d(\theta,K),\ \text{with } K \text{ in } (\text{B.3}).$$

4. *Suppose $\rho\geqslant 0$, $\tau\leqslant\frac{c^2(1-\rho)^2}{1+\rho}$, $\theta_1\in(-x_1^*,x_1^*)$, and $r\in\big(c,\bar{r}(\theta_1)\big)$. If*

$$\big(C_{\ell,1}(x_1,c),C_{\ell,2}(x_1,c)\big)\in\partial B\left(\theta,r\right),$$

then $C_{\ell,2}(x_1, c) < 0$.

Proof. Prove 1. Let $x_1 = C_{u,1}(\theta_1, c)$ and $x_2^{(2)} = C_{u,2}(\theta_1, c)$. By construction, $(x_1, x_2^{(2)}) \in \mathcal{C}_u(\tau, c)$ and $d\left(\theta, (x_1, x_2^{(2)})\right) = c < r$. In addition, there exists x_2 large enough such that $d\left(\theta, (x_1, x_2)\right) > r$. By continuity, there exists $x_2^{(1)} \in (x_2^{(2)}, x_2)$ such that $d\left(\theta, (x_1, x_2^{(2)})\right) = r$.

The proof of 2 is an analog of 1.

To prove 3, let $x_1 \notin (-x_1^*, x_1^*)$. The distance between θ and $\left(C_{u,1}(x_1, c), C_{u,2}(x_1, c)\right) \in \mathcal{C}_u(c)$ is

$$h(x_1) = \left(C_{u,1}(x_1, c) - \theta_1\right)^2 + \left(C_{u,2}(x_1, c) - X_2(\theta_1)\right)^2.$$

Taking the first order derivative

$$\begin{aligned}\frac{\mathrm{d}h(x_1)}{\mathrm{d}x_1} &= (x_1 - \theta_1)\left(\left((1+\rho)\tau + 2(1-\rho)x_1^2\right)^{3/2} - c(1-\rho^2)\tau\right) \\ &\quad \times \frac{2\left((1-\rho)^2 x_1(x_1+\theta_1) + (1-\rho^2)x_1^2 + (1+\rho)\tau + (1+\rho)\sqrt{(1-\rho)x_1^2+\tau}\sqrt{\theta_1^2(1-\rho)+\tau}\right)}{(1+\rho)\sqrt{\tau+(1-\rho)x_1^2}\left((1+\rho)\tau+2(1-\rho)x_1^2\right)^{3/2}\left(\sqrt{\tau+(1-\rho)x_1^2}+\sqrt{\theta_1^2(1-\rho)+\tau}\right)}.\end{aligned}$$

Note that (i) $|x_1| \geqslant x_1^* \geqslant |\theta_1|$ thus $x_1(x_1 + \theta_1) \geqslant 0$; (ii) $\tau \leqslant \frac{c^2(1-\rho)^2}{1+\rho}$ thus

$$\left((1+\rho)\tau + 2(1-\rho)x_1^2\right)^{3/2} - c(1-\rho^2)\tau \geqslant \left((1+\rho)\tau + 2(1-\rho)x_1^{*2}\right)^{3/2} - c(1-\rho^2)\tau = 0.$$

Therefore, the sign of $\frac{\mathrm{d}h(x_1)}{\mathrm{d}x_1}$ is the same as $(x_1 - \theta_1)$, and thus $\frac{\mathrm{d}h(x)}{\mathrm{d}x} < 0$ for $x < -x_1^* < 0$, and $\frac{\mathrm{d}h(x)}{\mathrm{d}x} > 0$ for $x > x_1^* > 0$. Hence, $h(x)$ is minimized at x_1^*, i.e. point H.

To prove 4, by contradiction, assume that there exists $A = (C_{\ell,1}(x_1), C_{\ell,2}(x_1)) \in \partial B\left((\theta_1, \theta_2), r\right)$ and $C_{\ell,2}(x_1) \geqslant 0$. WLOG, assume that $C_{\ell,1}(x_1) \geqslant 0$. Since $C_{\ell,2}(x_1)$ is increasing in x_1, and $C_{\ell,2}\left(\frac{\sqrt{c^2(\rho+1)^2-(\rho+1)\tau}}{\sqrt{2}\sqrt{1-\rho}}\right) = 0$, we have

$$x_1 > \frac{\sqrt{c^2(1+\rho)^2 - (1+\rho)\tau}}{\sqrt{2}\sqrt{1-\rho}} \geqslant x_1^*.$$

Let $A' = (C_{u,1}(x_1), C_{u,2}(x_1))$. By Lemma B.3, the perpendicular bisector of AA' is tangent of $\mathcal{S}_0^+(\tau)$ at $(x_1, X_2(x_1))$. Since $\mathcal{S}_0^+(\tau)$ is convex, θ is above the perpendicular bisector. This further implies that

$$r = d\,(\theta, A) > d\left(\theta, A'\right).$$

However, by Lemma B.4.3, we have

$$d\left(\theta, A'\right) \geqslant d\,(\theta, K) = \bar{r}(\theta_1),$$

which is a contradiction. Therefore, such x_1 does not exist. $\square$

Lemma B.5. *Let $r > 0$. $\mathcal{S}_0^+(\tau)$ intersects $\partial B\left((\theta_1, X_2(\theta_1)), r\right)$ at a minimum of two points.*

Proof. Let $h(x_1)$ be the distance between $(x_1, X_2(x_1)) \in \mathcal{S}_0^+(\tau)$ and the center of the circle $(\theta_1, \theta_2) = (\theta_1, X_2(\theta_1))$, i.e.

$$\begin{aligned} h(x_1) &= (x_1 - \theta_1)^2 + \left(X_2(x_1) - X_2(\theta_1)\right)^2 \\ &= \frac{\left(\sqrt{\tau + (1-\rho)x_1^2} - \sqrt{\tau + (1-\rho)\theta_1^2}\right)^2}{1+\rho} + (x_1 - \theta_1)^2. \end{aligned}$$

It is easy to see that $h(\theta_1) = 0$, $h(-\infty) = \infty$ and $h(+\infty) = \infty$. Since $h(x_1)$ is a continuous function, there exists $x_1^{(1)} < \theta_1 < x_1^{(2)}$ such that

$$h(x_1^{(1)}) = h(x_1^{(2)}) = r^2.$$

Thus $\mathcal{S}_0^+(\tau)$ intersects $\partial B\left((\theta_1, \theta_2), r\right)$ at $\left(x_1^{(1)}, X_2(x_1^{(1)})\right)$ and $\left(x_1^{(2)}, X_2(x_1^{(2)})\right)$. $\square$

Lemma B.6. *Let $\tau \leqslant \frac{(1-\rho)^2 c^2}{1+\rho}$, with $\rho \geqslant 0$. Suppose $|x_1| < x_1^*$ and let $O = (x_1, X_2(x_1)) \in \mathcal{S}_0(\tau)$. Define $\mathcal{C}_j^- = \left\{(x_1, x_2) : (x_1, -x_2) \in \mathcal{C}_j(c)\right\}$ where $j = \ell, u$, $\mathcal{C}_j$ is defined as in Lemma B.2. For all $r < \bar{r}(x_1)$, let $(x_1^{(1)}, 0), (x_1^{(2)}, 0) \in \partial B(O, r)$. Then*

1. $\max\left\{|x_1^{(1)}|, |x_1^{(2)}|\right\} \leqslant \bar{x}_1 := \frac{\sqrt{2}\sqrt{c^2(\rho+1) - \tau}}{\sqrt{1-\rho^2}}$.
2. If $(x_1, x_2) \in \partial B(O, r) \cap \mathcal{C}_\ell^-$, then $x_2 \geqslant 0$. Moreover, $\partial B(O, r) \cap \mathcal{C}_u^- = \varnothing$.

Proof. Proof of Part 1. WLOG, assume $x_1^{(1)} \geqslant 0$. Let $k = \frac{\sqrt{2}\sqrt{c^2(1-\rho)+\tau}}{\sqrt{1-\rho^2}}$ be the vertical axis of K. Let

$$\begin{aligned} h(x_1) &= \left\|O - (\bar{x}_1, 0)\right\|^2 - \|O - K\|^2 \\ &= (\bar{x}_1 - x_1)^2 + X_2(x_1)^2 - x_1^2 - \left(k - X_2(x_1)\right)^2. \end{aligned}$$

It suffices to show that $h(x_1^{(1)}) \geqslant 0$. Note that

$$\begin{aligned} \frac{\partial h(x_1)}{\partial x_1} &= \frac{2\sqrt{2}\left((1+\rho)^2(1-\rho)\left(c^2(1-\rho)+\tau\right)\left(\tau+(1-\rho)x_1^2\right)\right)^{-1/2}}{\left(c^2(1-\rho)+\tau\right)x_1(1-\rho)+\sqrt{(1+\rho)\left(c^4(1-\rho^2)+2c^2\rho\tau-\tau^2\right)}\sqrt{\tau+(1-\rho)x_1^2}}\,\tilde{h}(x_1) \\ \tilde{h}(x_1) &= 2(1-\rho)\left(c^2\tau(1-3\rho)+\tau^2-2c^4(1-\rho)\rho\right)x_1^2 - (1+\rho)\left(c^4(1-\rho^2)+2c^2\rho\tau-\tau^2\right)\tau. \end{aligned}$$

The sign of $\frac{\partial h(x_1)}{\partial x_1}$ depends on $\tilde{h}(x_1)$. Observe that

$$\tilde{h}(0) = -(1+\rho)\left(c^4 - (\tau - \rho c^2)^2\right)\tau \leqslant 0$$

where the inequality follows from $\tau \leqslant \frac{(1-\rho)^2 c^2}{1+\rho} \leqslant (1+\rho)c^2$. Moreover

$$\tilde{h}(x_1^*) = -2c^2(1-\rho)\rho\left(\tau - c^2(\rho - 1)\right)^2 \leqslant 0.$$

Since $\tilde{h}(x_1)$ is monotone in x_1 for $x_1 > 0$, we have

$$\tilde{h}(x_1) \leqslant \max\left\{\tilde{h}(0),\ \tilde{h}(x_1^*)\right\} \leqslant 0.$$

Thus $h(x_1)$ decreases in x_1, and $h(x_1^{(1)}) \geqslant 0$ follows from

$$\begin{aligned} &h(x_1) \geqslant h(x_1^*) \\ &= \frac{8c^4\rho^2(1-\rho^2)^{-1}}{c^2(1+\rho^2) - (1+\rho)\tau + \sqrt{(\rho+1)\left(c^2(1+\rho)-\tau\right)}\sqrt{c^2(1-\rho)^2 - (\rho+1)\tau}} \geqslant 0. \end{aligned}$$

Proof of Part 2. We can verify that $(\bar{x}_1, 0) \in \mathcal{C}_\ell^-$. Moreover, for all $(x_1, x_2) \in \mathcal{C}_\ell^-$, if $x_2 < 0$, then $|x_1| > \bar{x}_1$. By Part 1, such points cannot lie on $\partial B(O, r)$. Finally, $\partial B(O, r) \cap \mathcal{C}_u^-$ because

$$d(O, \mathcal{C}_u^-) > d(O, \mathcal{C}_u) = d(O, K) = \bar{r}(x_1) > r.$$

The inequality from the symmetry of $\mathcal{C}_u$ and $\mathcal{C}_u^-$ about the x_1-axis, combined with $X_2(x_1) > 0$. The equality follows from Lemma B.4.3. □

Lemma B.7. *Suppose $\vartheta_{\star,n} \to \vartheta_\star$, and $\tilde{g}_n$ satisfies (i) $\frac{\partial \tilde{g}_n(\vartheta_{\star,n})}{\partial \vartheta} = 0$, (ii) $\frac{\partial^2 \tilde{g}_n(\vartheta_{\star,n})}{\partial \vartheta \partial \vartheta'} \to H$ with a full rank H, (iii) $\frac{\partial^2 \tilde{g}_n(\vartheta)}{\partial \vartheta \partial \vartheta'}$ is Lipschitz continuous in ϑ with Lipschitz coefficient $M \in \mathbb{R}_+$ for all $\vartheta \in B(\vartheta_\star, \epsilon)$ where $\epsilon > 0$. Let $\mathbb{Z} \sim N(0, I_d)$ and $\mathbb{Z}_n = \mathbb{Z} + o_p(1)$. Let $h_n = r_n(\vartheta_n - \vartheta_{\star,n})$. If $\lim_n h_n = h \in \mathbb{R}^d$, then*

$$T_n^{(1)} = T^{(1)} + o_p(1),$$

where

$$T_n^{(1)} = \inf_{\tilde{g}_n(\vartheta_{\star,n} + r_n^{-1} x) = \tilde{g}_n(\vartheta_n)} \left\| \mathbb{Z}_n - (x - h_n) \right\|^2, \quad T^{(1)} = \inf_{x' H x = h' H h} \left\| \mathbb{Z} - (x - h) \right\|^2.$$

Proof. Let $x_n^*, x^* \in \mathbb{R}^k$ be minimizers of $T_n^{(1)}$ and $T^{(1)}$, respectively. Standard quadratic arguments imply $x_n^*, x^* = O_p(1)$.

Step 1. Prove $T^{(1)} \leqslant T_n^{(1)} + o_p(1)$. By feasibility of x_n^*,

$$\begin{aligned}
&\tilde{g}_n(\vartheta_{\star,n} + r_n^{-1} x_n^*) = \tilde{g}_n(\vartheta_n) = \tilde{g}_n(\vartheta_{\star,n} + r_n^{-1} h_n) \\
&\tilde{g}_n(\vartheta_{\star,n}) + r_n^{-1} \tfrac{\partial \tilde{g}_n(\vartheta_{\star,n})}{\partial \vartheta'} x_n^* + r_n^{-2} x_n^{*\prime} \tfrac{\partial^2 \tilde{g}_n(\bar{\vartheta})}{\partial \vartheta \partial \vartheta'} x_n^* = \tilde{g}_n(\vartheta_{\star,n}) + r_n^{-1} \tfrac{\partial \tilde{g}_n(\vartheta_{\star,n})}{\partial \vartheta'} h_n + r_n^{-2} h_n' \tfrac{\partial^2 \tilde{g}_n(\bar{\bar{\vartheta}})}{\partial \vartheta \partial \vartheta'} h_n \\
&\Rightarrow \; x_n^{*\prime} \tfrac{\partial^2 \tilde{g}_n(\bar{\vartheta})}{\partial \vartheta \partial \vartheta'} x_n^* = h_n' \tfrac{\partial^2 \tilde{g}_n(\bar{\bar{\vartheta}})}{\partial \vartheta \partial \vartheta'} h_n
\end{aligned}$$

where $\bar{\vartheta}$ is between $\vartheta_{\star,n}$ and $\vartheta_{\star,n} + r_n^{-1} x_n^*$, and $\bar{\bar{\vartheta}}$ is between $\vartheta_{\star,n}$ and $\vartheta_{\star,n} + r_n^{-1} h_n$. By Assumption 4.1.1,

$$\begin{aligned}
x_n^{*\prime} \left[H + o_p(1) \right] x_n^* &= \left(h + o(1) \right)' \left[H + o_p(1) \right] \left(h + o(1) \right) \\
\Rightarrow x_n^{*\prime} H x_n^* &= h' H h + o_p(1).
\end{aligned} \tag{B.13}$$

Case 1. $h' H h = 0$ and H is positive/ negative definite. (B.13) implies $x_n^* = o_p(1)$. Hence

$$T^{(1)} \leqslant \left\| \mathbb{Z} + h \right\|^2 = T_n^{(1)} + o_p(1).$$

For Case 2 and Case 3 below, we construct $\tilde{x}_n^* = x_n^* + o_p(1)$ such that $\tilde{x}_n^{*\prime} H \tilde{x}_n^* = h' H h$.

It then follows that

$$T^{(1)} \leqslant \left\|\mathbb{Z} - (\tilde{x}_n^* - h)\right\|^2 = \left\|\mathbb{Z}_n - (x_n^* - h)\right\|^2 + o_p(1) = T_n^{(1)} + o_p(1).$$

Case 2. $h'Hh = 0$ and H is indefinite. By Lemma B.13, there is y_n such that

$$x_n^{*\prime} H y_n = 0, \quad y_n' H y_n = -\text{sign}(x_n^{*\prime} H x_n^*).$$

Let $\tilde{x}_n^* = x_n^* + \xi_n y_n$, where $\xi_n = \sqrt{|x_n^{*\prime} H x_n^*|}$. By (B.13), $\xi_n = o_p(1)$. In addition,

$$\tilde{x}_n^{*\prime} H \tilde{x}_n^* = (x_n^* + \xi_n y_n)' \, H \, (x_n^* + \xi_n y_n) = x_n^{*\prime} H x_n^* - \text{sign}(x_n^{*\prime} H x_n^*)\xi_n^2 = 0.$$

Case 3. $h'Hh \neq 0$. Let $(1+\xi_n)^2 = \frac{h'Hh}{x_n^{*\prime} H x_n^*}$. By (B.13), $\xi_n = o_p(1)$. Let $\tilde{x}_n^* = x_n^* + \xi_n x_n^*$. By construction, $\tilde{x}_n^{*\prime} H \tilde{x}_n^* = h'Hh$.

Step 2. Prove $T_n^{(1)} \leqslant T^{(1)} + o_p(1)$. Case 1. $h'Hh = 0$ and H is positive or negative definite. Here $x^* = 0$ and $T = \mathbb{Z}'\mathbb{Z}$. Thus,

$$T_n^{(1)} \leqslant \|\mathbb{Z}_n\|^2 + o_p(1) = T^{(1)} + o_p(1).$$

For Case 2 and Case 3 below, we show that there exists $\eta_n = o_p(1)$ such that

$$\tilde{g}_n(\vartheta_{\star,n} + r_n^{-1}(x^* + \eta_n) = \tilde{g}_n(\vartheta_n).$$

Then the conclusion follows from

$$T_n^{(1)} \leqslant \left\|\mathbb{Z}_n - (x^* + \eta_n - h_n)\right\|^2 = \left\|\mathbb{Z} - (x^* - h)\right\|^2 + o_p(1) = T^{(1)} + o_p(1).$$

Case 2. $h'Hh = 0$ and H is indefinite. Assume $H = \text{diag}\{\lambda_1, ..., \lambda_m, -\lambda_{m+1}, ..., -\lambda_d\}$ with $\lambda_1, ..., \lambda_d > 0$. If H is not diagonal, write $H = P'\Lambda P$ with diagonal Λ and transform x^* by Px^*. Define

$$y^* = -\text{sign}\left(\tilde{g}_n(\vartheta_{\star,n} + r_n^{-1}x^*) - \tilde{g}_n(\vartheta_n)\right)(x_1^*, ..., x_m^*, -x_{m+1}^*, ..., -x_d^*).$$

Then

$$y^{*\prime}Hy^{*}=x^{*\prime}Hx^{*}=0,\ y^{*\prime}Hx^{*}=-\text{sign}\left(\tilde{g}_{n}(\vartheta_{\star,n}+r_{n}^{-1}x^{*})-\tilde{g}_{n}(\vartheta_{n})\right)\sum_{i=1}^{d}\lambda_{i}x_{i}^{*2}.$$

Let

$$u_{n}(\xi)=r_{n}^{2}\left(\tilde{g}_{n}\left(\vartheta_{\star,n}+r_{n}^{-1}(x^{*}+\xi y^{*})\right)-\tilde{g}_{n}(\vartheta_{n})\right).$$

Define

$$\xi_{n}=\arg\min_{\xi}|\xi|\text{ s.t. }u_{n}(\xi)=0.$$

We show $\xi_{n}=o_{p}(1)$. Note that

$$u_{n}(0)=r_{n}^{2}\left(\tilde{g}_{n}\left(\vartheta_{\star,n}+r_{n}^{-1}x^{*}\right)-\tilde{g}_{n}(\vartheta_{n})\right),$$

and for all $\epsilon>0$,

$$u_{n}(\epsilon)=2\epsilon y^{*\prime}Hx^{*}+o_{p}(1)=-2\epsilon\text{sign}\left(u_{n}(0)\right)\sum_{i=1}^{d}\lambda_{i}x_{i}^{*2}+o_{p}(1).$$

By Lemma B.10, there is N such that for all $n\geqslant N$,

$$P(|\xi_{n}|\leqslant\epsilon)\geqslant P\left(u_{n}(\epsilon)u_{n}(0)\leqslant 0\right)\geqslant 1-\epsilon,$$

which implies that $\xi_{n}=o_{p}(1)$. Let $\eta_{n}=\xi_{n}y^{*}$, the conclusion follows.

Case 3. $h'Hh\neq 0$. Define

$$\xi_{n}=\arg\min_{\xi}|\xi|\text{ s.t. }r_{n}^{2}\left(\tilde{g}_{n}(\vartheta_{\star,n}+(1+\xi)r_{n}^{-1}x^{*})-\tilde{g}_{n}(\vartheta_{n})\right)=0,$$

and $\xi_{n}=\infty$ if no solution exists. We show $\xi_{n}=o_{p}(1)$, i.e. for all $\epsilon>0$, there is N such that for all $n>N$ $P\left(|\xi_{n}|>\epsilon\right)<\epsilon$. To see this, let

$$u_{n}(\xi)=r_{n}^{2}\left(\tilde{g}_{n}(\vartheta_{\star,n}+r_{n}^{-1}(1+\xi)x^{*})-\tilde{g}_{n}(\vartheta_{n})\right).$$

Then

$$u_{n}(\epsilon)=\left((1+\epsilon)^{2}-1\right)h'Hh+o_{p}(1),\ u_{n}(-\epsilon)=\left((1-\epsilon)^{2}-1\right)h'Hh+o_{p}(1).$$

Therefore, there is N such that for all $n > N$,

$$P\left(u_n(\epsilon)u_n(-\epsilon) < 0\right) \geqslant 1 - \epsilon.$$

By the continuity of u_n, $\left\{u_n(\epsilon) > 0 \text{ and } u_n(-\epsilon) < 0\right\}$ implies $|\xi_n| \leqslant \epsilon$. Therefore, for all $n > N$,

$$P(|\xi_n| \leqslant \epsilon) \geqslant P\left(u_n(\epsilon) > 0 \text{ and } u_n(-\epsilon) < 0\right) \geqslant 1 - \epsilon.$$

The conclusion follows from $\eta_n = \xi_n x^*$. $\square$

Lemma B.8. *Suppose Assumption 3.2 holds. Let $\mathbb{Z} \sim N(0, I_d)$ and $\mathbb{Z}_n = \mathbb{Z} + o_p(1)$. Suppose $\Sigma_n = \Sigma + o_p(1)$ with $\Sigma \in \mathcal{S}$, where $\mathcal{S}$ is defined in Assumption 4.1.3. If $\lim s_n(\theta_n - \theta_\star) = h \in \mathbb{R}^d$ with $h'H \neq 0$, for some sequence $s_n \to \infty$ with $s_n/r_n \to 0$, then*

$$T_n^{(2)} = T^{(2)} + o_p(1)$$

where

$$T_n^{(2)} = \inf_{g(\theta_n + r_n^{-1}x) = g(\theta_n)} \left\|\mathbb{Z}_n - \Sigma_n^{-1/2}x\right\|^2, \quad T^{(2)} = \inf_{h'Hx=0} \left\|\mathbb{Z} - \Sigma^{-1/2}x\right\|^2. \tag{B.14}$$

Proof. Let x_n^* and x^* denote the optimizers of $T_n^{(2)}$ and $T^{(2)}$, respectively. It is easy to verify that $x_n^*, x^* = O_p(1)$.

Step 1. Prove $T^{(2)} \leqslant T_n^{(2)} + o_p(1)$. Let $\theta_n^* = \theta_n + r_n^{-1}x_n^*$. By the feasibility constraint,

$$0 = s_n r_n \left(g(\theta_n^*) - g(\theta_n)\right).$$

Expanding $g(\theta_n^*)$ around θ_n using a second-order Taylor expansion, and linearizing $\nabla g(\theta_n)$ around $\theta_\star$, we obtain

$$\begin{aligned}
0 =& s_n r_n \left(\frac{\partial g(\theta_n)}{\partial \theta'}(\theta_n^* - \theta_n) + (\theta_n^* - \theta_n)'\frac{\partial g(\bar{\theta})}{\partial\theta\partial\theta'}(\theta_n^* - \theta_n)\right) \\
=& s_n(\theta_n - \theta_\star)\frac{\partial g(\bar{\bar{\theta}})}{\partial\theta\partial\theta'} r_n(\theta_n^* - \theta_n) + \frac{s_n}{r_n} r_n(\theta_n^* - \theta_n)'\frac{\partial g(\bar{\theta})}{\partial\theta\partial\theta'} r_n(\theta_n^* - \theta_n) \\
=& h_n'\frac{\partial g(\bar{\bar{\theta}})}{\partial\theta\partial\theta'} x_n^* + \frac{s_n}{r_n} x_n^{*\prime}\frac{\partial g(\bar{\theta})}{\partial\theta\partial\theta'} x_n^* \\
=& h'Hx_n^* + o_p(1).
\end{aligned} \tag{B.15}$$

where $\bar{\theta}$ is between θ_n and θ_n^* and $\bar{\bar{\theta}}$ is between θ_n and $\theta_\star$. Since $h'H \neq 0$, there is ι such that $h'H\iota = 1$, and $a_n = o_p(1)$ such that

$$h'H(x_n^* + a_n\iota) = 0.$$

Thus $x_n^* + a_n\iota$ is feasible for the problem in $T^{(2)}$, which implies that

$$T^{(2)} \leqslant \left\| \mathbb{Z} - \Sigma_n^{-1/2}(x_n^* + a_n\iota) \right\|^2 = T_n^{(2)} + o_p(1).$$

Step 2, we show that $T_n^{(2)} \leqslant T^{(2)} + o_p(1)$. By the same algebra in (B.15), for $b = O_p(1)$,

$$s_n r_n \left(g(\theta_n + r_n^{-1}(x^* + b\iota)) - g(\theta_n) \right) = h'H(x^* + b\iota) + o_p(1) = b + o_p(1),$$

where the last equality follows from $h'Hx^* = 0$. Define b_n as the solution for

$$b_n = \arg\min_b |b| \text{ s.t. } g(\theta_n + r_n^{-1}(x^* + b\iota)) - g(\theta_n) = 0,$$

and $b_n = \infty$ if there is no solution for $g(\theta_n + r_n^{-1}(x^* + b\iota)) - g(\theta_n) = 0$. Next, we show that $b_n = o_p(1)$, i.e. for all $\epsilon > 0$, there is N such that for all $n > N$, $P(|b_n| > \epsilon) < \epsilon$. Let

$$u_n(b) = s_n r_n \left(g(\theta_n + r_n^{-1}(x^* + b\iota)) - g(\theta_n) \right).$$

Then

$$u_n(\epsilon) = \epsilon + o_p(1), \; u_n(-\epsilon) = -\epsilon + o_p(1).$$

Therefore, there is N such that for all $n > N$,

$$P\left(u_n(\epsilon) u_n(-\epsilon) < 0 \right) \geqslant 1 - \epsilon.$$

This implies that

$$T_n^{(2)} \leqslant \left\| \mathbb{Z}_n - \Sigma_n^{-1/2}(x^* + b_n\iota) \right\|^2 = \left\| \mathbb{Z} - \Sigma^{-1/2}x^* \right\|^2 + o_p(1) = T^{(2)} + o_p(1).$$

Step 1 and 2 complete the proof. □

Lemma B.9. *Suppose Assumption 4.1.3 and 4.14 hold. Let $\hat{h}_n$ and s_n be sequences* *such that* $s_n \to \infty$, $||\hat{h}_n||/s_n \xrightarrow{p} b \neq 0$, *and* $\frac{\hat{h}_n'}{||\hat{h}_n||}H \xrightarrow{p} h'$. *For* $\hat{T}_n^*(h)$ *defined in (4.4) with* $\mathbb{Z} \sim N(0, I_d)$, *it holds that*

$$\hat{T}_n^*(\hat{h}_n) = \inf_{h'x=0} \left\|\mathbb{Z} - \Sigma^{-1/2}x\right\|^2 + o_p(1).$$

Proof. Let $x = t - h_n$, we can rewrite $\hat{T}_n^*(\hat{h}_n)$ as

$$\hat{T}_n^*(\hat{h}_n) = \inf_{x: \frac{\hat{h}_n' H}{||\hat{h}_n||}x = -\frac{x'Hx}{2||\hat{h}_n||}} \left\|\mathbb{Z} - (\hat{\Sigma})^{-1/2}x\right\|^2. \tag{B.16}$$

Note that the optimizer $x_n^* = O_p(1)$, thus

$$h'x_n^* = \left(\frac{\hat{h}_n' H}{||\hat{h}_n||} + o(1)\right) x_n^* = -\frac{x_n^{*\prime} H x_n^*}{2||\hat{h}_n||} + o_p(1) = o_p(1)$$

where the last equality follows from $\hat{h}_n/s_n \xrightarrow{p} b \neq 0$. The remainder follows by the same continuity and perturbation arguments as in Lemma B.8. □

Lemma B.10. *Let* $\varepsilon_n = o_p(1)$, $\mathbb{X} \sim N(h, I_d)$ *with* $h \in \mathbb{R}^d$, *and let* x^* *be the solution to*

$$\inf_{x'Hx=0} T(x), \text{ where } T(x) = \|\mathbb{X} - x\|^2,$$

where $H = diag\{\lambda_1, ..., \lambda_m, -\lambda_{m+1}, ..., -\lambda_d\}$ *with* $\lambda_i > 0$. *Then for all* $\epsilon > 0$, *there is* N *such that for all* $n > N$,

$$P\left(\epsilon \sum_{i=1}^d \lambda_i x_i^{*2} + \varepsilon_n > 0\right) \geqslant 1 - \epsilon.$$

Proof. Step 1. Let $\ell_1 = \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_1 + \lambda_d}}$, $\ell_2 = \sqrt{1 - \ell_1^2}$. Then

$$\inf_{x'Hx=0} T(x) \leqslant \inf_{\ell_1 x_1 - \ell_2 x_d = 0,\ x_{2:d-1}=0} T(x) = T(0) - (\ell_2 \mathbb{X}_1 + \ell_1 \mathbb{X}_d)^2.$$

By the continuity of $T(x)$, for all $\epsilon > 0$, there is $C_\epsilon > 0$ such that

$$P\left(\sum_{i=1}^d \lambda_i x_i^{*2} < C_\epsilon\right) \leqslant P\left(T(0) - \inf_{x'Hx=0} T(x) < \left(\Phi^{-1}(\frac{1}{2} + \frac{\epsilon}{4})\right)^2\right)$$

$$\leqslant P\left((\ell_2\mathbb{X}_1+\ell_1\mathbb{X}_d)^2<\left(\Phi^{-1}(\frac{1}{2}+\frac{\epsilon}{4})\right)^2\right).$$

Since $\ell_2\mathbb{X}_1+\ell_1\mathbb{X}_d \sim N\left(\ell_2 h_1+\ell_1 h_d, 1\right)$, the squared term follows a noncentral χ_1^2, so the probability is bounded by $\epsilon/2$.

Step 2. Since $\varepsilon_n = o_p(1)$, there is N such that for all $n \geqslant N$,

$$P\left(\varepsilon_n \geqslant -\frac{C_\epsilon \epsilon}{2}\right) \geqslant 1-\frac{\epsilon}{2}.$$

Combining with Step 1,

$$\begin{aligned}
P\left(\epsilon\sum_{i=1}^d \lambda_i x_i^{*2}+\varepsilon_n>0\right) &\geqslant P\left(\sum_{i=1}^d \lambda_i x_i^{*2} \geqslant C_\epsilon, \varepsilon_n > -\frac{C_\epsilon\epsilon}{2}\right)\\
&\geqslant P\left(\sum_{i=1}^d \lambda_i x_i^{*2} \geqslant C_\epsilon\right)+P\left(\varepsilon_n > -\frac{C_\epsilon\epsilon}{2}\right)-1\\
&\geqslant 1-\frac{\epsilon}{2}+1-\frac{\epsilon}{2}-1=1-\epsilon.
\end{aligned}$$

□

Lemma B.11. *Let $D(Y)=\inf_{x:x'Hx=c}\|Y-x\|$ where Y is continuously distributed and $c \geqslant 0$. Then $D(Y)$ is continuously distributed.*

Proof. It suffices to show that for all $a \geqslant 0$, $P\left(D(Y)=a\right)=0$. Let $S=\{x : x'Hx=c\}$.

Case 1. $a=0$. Here $P(D(Y)=0)=P(Y\in S)$. Since H is indefinite, S is a variety of dimension at most $d-1$, and hence S has Lebesgue measure zero. Because Y has a continuous distribution, so $P(Y\in S)=0$.

Case 2. $a>0$. Suppose, by contradiction, that $P\left(D(Y)=a\right)>0$. Then the set $S_a=\left\{y : D(y)=a\right\}$ must have positive Lebesgue measure. Since $D(y)$ is continuous in y, S_a is a closed set. Hence, there exists a ball $B(o,r)$ with $a>r>0$ such that $B(o,r)\subseteq S_a$. Let o' be the projection of o onto S, i.e. $o'\in S$ and $||o-o'||=a$. Choose a point $k\in\partial B(o,r)\cap\overline{oo'}$. Such k exists since o lies inside $\partial B(o,r)$ while o'

lies outside. By construction,

$$||k - o'|| = ||o - o'|| - ||o - k|| = a - r < a$$

which contradicts the fact that $k \in B(o, r) \subseteq S_a$. Therefore, no such $a > 0$ can exist. □

Lemma B.12. *Let $\mathbb{Z} \sim N(0, I_d)$, and $\mathcal{H}_z$ be a set satisfying $P(\mathbb{Z} \in \mathcal{H}_z) \geqslant 1 - \eta$. Then*

$$Q\left(\inf_{h'x=0} \|\mathbb{Z} - x\|^2 \middle| \mathbb{Z} \in \mathcal{H}_z; \frac{1-\alpha}{1-\eta}\right) \in \left[Q(\chi_1^2, 1-\alpha),\ Q(\chi_1^2, 1-\alpha+\eta)\right).$$

Proof. To show the upper bound,

$$\begin{aligned}
&P\left(\inf_{h'x=0} \|\mathbb{Z} - x\|^2 \leqslant Q(\chi_1^2, 1-\alpha+\eta) \middle| \mathbb{Z} \in \mathcal{H}_z\right) \\
=&\frac{P\left(\inf_{h'x=0} \|\mathbb{Z} - x\|^2 \leqslant Q(\chi_1^2, 1-\alpha+\eta), \mathbb{Z} \in \mathcal{H}_z\right)}{P(\mathbb{Z} \in \mathcal{H}_z)} \\
>&\frac{P\left(\inf_{h'x=0} \|\mathbb{Z} - x\|^2 \leqslant Q(\chi_1^2, 1-\alpha+\eta)\right) + P(\mathbb{Z} \in \mathcal{H}_z) - 1}{P(\mathbb{Z} \in \mathcal{H}_z)} \\
=&\frac{1-\alpha+\eta+1-\eta-1}{1-\eta} = \frac{1-\alpha}{1-\eta}.
\end{aligned}$$

To show the lower bound,

$$\begin{aligned}
&P\left(\inf_{h'x=0} \|\mathbb{Z} - x\|^2 \leqslant Q(\chi_1^2, 1-\alpha) \middle| \mathbb{Z} \in \mathcal{H}_z\right) \\
=&\frac{P\left(\inf_{h'x=0} \|\mathbb{Z} - x\|^2 \leqslant Q(\chi_1^2, 1-\alpha), \mathbb{Z} \in \mathcal{H}_z\right)}{P(\mathbb{Z} \in \mathcal{H}_z)} \\
\leqslant&\frac{P\left(\inf_{h'x=0} \|\mathbb{Z} - x\|^2 \leqslant Q(\chi_1^2, 1-\alpha)\right)}{P(\mathbb{Z} \in \mathcal{H}_z)} = \frac{1-\alpha}{1-\eta}.
\end{aligned}$$

□

Lemma B.13. *Let H be a $d \times d$ full rank indefinite matrix. For all $d \times 1$ vector x,* *there exists* $d \times 1$ *vector* y *such that* $y'Hx = 0$ *and* $y'Hy = -sign(x'Hx)$.

Proof. If $x'Hx = 0$, the conclusion holds trivially with $y = 0_d$. WLOG, assume that $x'Hx > 0$. First assume that H is diagonal. If not, write $H = P'\Lambda P$, where Λ is diagonal, $\tilde{x} = Px$ and $\tilde{y} = Py$; the same argument then applies. Without loss of generality, let

$$H = \text{diag}(\lambda_1, ..., \lambda_m, -\lambda_{m+1}, ..., -\lambda_d), \quad \lambda_1, ..., \lambda_d > 0.$$

Case 1. If $x_{m+1}, ..., x_d = 0$, let $y = (0_{d-1}, 1)$. It is easy to verify that $y'Hx = 0$ and $y'Hy = -\lambda_d < 0$. Case 2. Suppose $x_d \neq 0$, let $y = (x_1, ..., x_m, 0_{d-m-1}, y_d)$, $y_d = \frac{\sum_{i=1}^m \lambda_i x_i^2}{\lambda_d x_d}$. Then

$$y'Hx = \sum_{i=1}^m \lambda_i x_i^2 - \lambda_d x_d y_d = \sum_{i=1}^m \lambda_i x_i^2 - \lambda_d x_d \frac{\sum_{i=1}^m \lambda_i x_i^2}{\lambda_d x_d} = 0,$$

$$\begin{aligned} y'Hy &= \sum_{i=1}^m \lambda_i x_i^2 - \lambda_d y_d^2 = -\frac{\sum_{i=1}^m \lambda_i x_i^2}{\lambda_d x_d^2} \left(\sum_{i=1}^m \lambda_i x_i^2 - \lambda_d x_d^2 \right) \\ &\leqslant -\frac{\sum_{i=1}^m \lambda_i x_i^2}{\lambda_d x_d^2} x'Hx < 0. \end{aligned}$$

Then conclusion holds with a simple normalization of y. □

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). Gender stereotypes in the classroom and effects on achievement. *The Review of Economics and Statistics 100*(5), 876–890.

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

Andrews, I. and A. Mikusheva (2022). Gmm is inadmissible under weak identification. *arXiv preprint arXiv:2204.12462*.

Benkwitz, A., M. H. Neumann, and H. Lütekpohl (2000). Problems related to confidence intervals for impulse responses of autoregressive processes. *Econometric Reviews 19*(1), 69–103.

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

Chen, Q. and Z. Fang (2019b). Inference on functionals under first order degeneracy. *Journal of Econometrics 210*(2), 459–481.

Cohen, J., P. Cohen, S. G. West, and L. S. Aiken (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

Drton, M. and H. Xiao (2016). Wald tests of singular hypotheses. *Bernoulli 22*(1), 38–59.

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.

Dufour, J.-M. and P. Valery (2016). Rank-robust regularized wald-type tests. Technical report, Working paper.

Fang, Z. and A. Santos (2019). Inference on directionally differentiable functions. *The Review of Economic Studies 86*(1), 377–412.

Gaffke, N., B. Heiligers, and R. Offinger (2002). On the asymptotic null-distribution of the wald statistic at singular parameter points. *Statistics & Risk Modeling 20*(1-4), 379–398.

Gaffke, N., R. Steyer, and A. A. v. Davier (1999). On the asymptotic null-distribution of the wald statistic at singular parameter points. *Statistics & Risk Modeling 17*(4), 339–358.

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.

Hawinkel, S., W. Waegeman, and S. Maere (2024). Out-of-sample r 2: estimation and inference. *The American Statistician 78*(1), 15–25.

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.

Kaji, T. (2021). Theory of weak identification in semiparametric models. *Econometrica 89*(2), 733–763.

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.

Lütkepohl, H. (2013). *Introduction to multiple time series analysis*. Springer Science & Business Media.

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

Miller, D. L., F. Molinari, and J. Stoye (2025). Testing sign congruence between two parameters. *Journal of Applied Econometrics*.

Ober-Reynolds, D. (2026). Robustness to missing data: breakdown point analysis. *Journal of Econometrics 253*, 106151.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological methodology 13*, 290–312.

Staiger, D. O. and J. H. Stock (1994). Instrumental variables regression with weak instruments.

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

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

van Garderen, K. J. and N. van Giersbergen (2022). On the optimality of the lr test for mediation. *Symmetry 14*(1), 178.

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