Inference on Union Bounds

Xinyue Bei[*]
The University of Texas at Austin
<xinyue.bei@austin.utexas.edu>

November 5, 2024

Abstract

A union bound is a union of multiple bounds. Union bounds occur in a wide variety of empirical settings, such as difference-in-differences, regression discontinuity design, bunching, and misspecification analysis. In this paper, I propose a confidence interval for these kinds of bounds based on modified conditional inference. I show that it improves upon existing methods in a large set of data generating processes. The new procedure gives statistically significant results while the pre-existing alternatives do not in the empirical applications in Dustmann, Lindner, Schönberg, Umkehrer, and Vom Berge (2022). I implement the proposed method in the companion R package `UnionBounds` for easy use in practice.

KEYWORDS: Bound analysis, partial identification, conditional inference, moment inequalities, sensitivity analysis

[*]I thank Adam Rosen, Federico Bugni, and Matt Masten for their guidance on this project. I am also grateful to Dante Amengual, Anna Bykhovskaya, Tim Bollerslev, Désiré Kédagni, Arnaud Maurel, Andrew Patton, Michael Pollmann, Muyang Ren, Enrique Sentana, Silvia Sarpietro, Valentin Verdier, and participants of the Triangle Econometrics Conference, Université de Montréal, NYU, Bonn, LSE, CEMFI, Emory, TAMU, UTexas, UVA, Michigan, WUStL, Peking University, Cornell, Class of 2024 Conference, for their useful comments.

1 Introduction

This paper studies inference for a target object partially identified by the union of a *set* of *bounds*, namely, a union bound, and provides new procedures that significantly improve upon the existing alternatives. Union bounds commonly arise in empirical work, for example, in the assessment of the importance of the parallel trends assumption in difference-in-differences (DiD) analyses. Recent papers such as Manski and Pepper (2018) and Rambachan and Roth (2023) (RR23) study the relaxation of the classical parallel trends assumption within a DiD framework. One of their approaches is to assume that the violation of parallel trends in a post-policy period is bounded above by the maximum violation in the pre-policy periods. In this case, the identified set for the average treatment effect on the treated can be characterized as a union bound, where each *bound* is formed by the DiD estimand adding and subtracting the violation of a pre-policy year, and the *set* is all pre-policy periods. I also discuss applications to regression discontinuity designs, bunching strategies to identify the elasticity of taxable income, and misspecified models in Section 2.

In this paper, I provide a general framework for inference on union bounds. The main difficulty for inference is that the endpoints of a union bound are non-smooth functions of each single *bound*. Hirano and Porter (2012) show that there is no local asymptotic quantile unbiased estimator. Moreover, Fang and Santos (2019) show that an empirical bootstrap procedure, in the terminology of Horowitz (2019), is not uniformly valid. Similar difficulties appear in inference for moment inequalities and directionally differentiable functions, but the existing methods do not apply to union bounds because of the different restrictions on the null parameter space. So far there are two uniformly valid methods. The first one is a simple confidence interval (CI), which is the union of CIs for each *bound*. This method can be overly conservative, especially when the *bounds* are close to each other. The second one is the adjusted bootstrap procedure proposed in Ye, Keele, Hasegawa, and Small (2023) (YKHS23). This method involves a subsample so the CI converges to the identified set at a rate slower than $\sqrt{n}$, resulting in trivial power for $\sqrt{n}$ local alternatives. RR23 propose an inference procedure for their sensitivity analysis in DiD. However, the procedure relies on the specific structure of DiD and does not apply to general union bound settings.

In Section 3, I propose a modified conditional CI. Loosely speaking, I construct a conditional critical value exploiting the distribution of the maximum estimated upper *bound* (resp. the minimum estimated lower *bound*) conditional on the second largest estimated upper *bound* (resp. the second smallest estimated lower *bound*). In this way, the conditional critical value is data-adaptive and sensitive to the binding *bounds*, which leads to a shorter CI when the bounds are relatively close to each other. However, the conditional critical value is not uniformly valid, and for that reason, I propose a novel modification that truncates the conditional critical value from below to guarantee uniform coverage. The modified conditional CI converges to the identified set at a rate of $\sqrt{n}$, and thus has material power improvement upon YKHS23. I also show that under a large set of data generating processes (DGPs), the modified conditional CI is shorter than the simple CI with probability approaching one. I provide the `UnionBounds` R package to implement my method.[1]

In Section 4, I conduct simulations based on the DiD settings in RR23 and compare the performance of my modified conditional CI to the simple CI, the adjusted bootstrap in YKHS23 and the hybrid CI in RR23. The length of the median modified conditional CI is the smallest in most simulation designs and is close to being the smallest in all designs.[2] In terms of the length of the median CI, net of median point estimates of the bound, the modified conditional CI results in a decrease of up to 43% relative to alternative methods.

In Section 5, I illustrate the proposed inference procedures in an application using RR23's sensitivity analysis in Dustmann et al. (2022). Specifically, Dustmann et al. (2022) study the effects of the minimum wage introduced in Germany in 2015. The authors are interested in whether the employment effect is greater than the negative wage effect, which leads to an elasticity smaller than one. The authors conduct the analysis using DiD and relax the parallel trends assumption following RR23. Under all levels of relaxation, the modified conditional CI is shorter than the simple CI and the one provided by RR23. Under the benchmark relaxation, my 95% CI suggests that the elasticity is smaller than one, while the 95% RR23 CI and a simple 95% CI do not. My method gives a breakdown relaxation 33% to 66% larger than the RR23 and the simple CI respectively.

[1]The latest version of the R package is available at https://github.com/xinyuebei-econ/UnionBounds

[2]The median CI is the median of the endpoints of the $1 - \alpha$ CI across simulated samples.

Related Literature

Although there are many empirical examples where the identified set is a union bound, only a small number of inference approaches have been developed, which I discuss next.

First, a common practice is a simple CI constructed based on the intersection union principle discussed in Casella and Berger (2021) (ch. 8.2.3), see Conley, Hansen, and Rossi (2012), Kolesár and Rothe (2018), Hasegawa, Webster, and Small (2019), and Ban and Kedagni (2022), among others. The idea is to first construct a CI for each *bound* and then take a union over the *set*, which is intuitive and has uniformly valid coverage.[3] However, taking union over the confidence intervals inflates the coverage rate, and the simple CI can be overly conservative. I prove that the simple CI is wider and has lower local power than my proposal under a large set of DGPs.

Second, YKHS23 study the relaxation of the parallel trends assumption in DiD based on a negative correlation bracketing strategy. The resulting identified set for the average treatment effect on the treated is a union bound. To address inference, they introduce two bootstrap methods. The first one is an empirical bootstrap procedure, in the terminology of Horowitz (2019). This method is not uniformly valid and may overreject when the bounds are close to each other. The second procedure introduces an adjustment term based on a subsample so that it has uniform asymptotic coverage, but at the cost that the CI converges to the identified set at a rate slower than $\sqrt{n}$. This causes material power loss for a large set of local alternatives relative to my CI.

Third, RR23 propose an inference procedure under the specific structure of relaxation of the parallel trends assumption in DiD. The main idea is to partition the parameter space so that each element in the partition can be represented by a set of moment inequalities. RR23 first construct the CI for each element in the partition based on Andrews, Roth, and Pakes (2023). They then take a union over different elements in the partition to get a valid CI for the union bound. While

[3]The simple method relates to the statistical literature on testing whether the minimum of several elements is no greater than zero, which can be framed as a one-sided union bound problem. Cohen, Gatsonis, and Marden (1983) show that the one-sided version of the simple test is uniformly most powerful among all monotone level α tests. Berger (1989) proposes a method that strictly improves the power of the simple one, but inverting this test yields a disconnected CI, whose convex hull coincides with that of the simple CI. Additionally, the rejection region is difficult to interpret, see e.g. Perlman and Wu (1999).

the CI for each element is efficient, the efficiency may not hold after taking the union. In both the simulation and the empirical application, my CI outperforms their CI when the *bounds* are not well separated. Moreover, their method uses the specific DiD structure and does not apply to general finite union bounds.

The procedure constructed in this paper also contributes to other related literature, such as intersection bounds, directional differentiable functions, and conditional inference.

The union bound inference complements the large literature on intersection bounds and moment inequality models. Chernozhukov, Lee, and Rosen (2013) investigate inference on intersection bounds, where the target object is in the intersection of a set of bounds. A leading case of intersection bounds is inference on a parameter bounded by moment inequalities. See Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), Rosen (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010), Andrews and Shi (2013), and Bugni, Canay, and Shi (2015), among others, for different procedures. Inference for intersection bounds and union bounds share some similar challenges, but also differ in important ways: The differences between the target object and the bounds, scaled by $\sqrt{n}$, is an important element for inference, but can not be consistently estimated. With intersection bounds, the signs of the differences are known, e.g. the target object is larger than all lower bounds, while with union bounds, the sign is unclear, e.g. the target object is larger than at least one lower bound. Thus the problem of inference on union bounds is different from intersection bounds and requires a different treatment.

My method also sheds light on inference on directionally differentiable functions. A union bound can be written as the minimum of a set of lower bounds to the maximum of a set of upper bounds. The min and max operators are directionally differentiable. Fang (2018) and Ponomarev (2022) study the efficient estimation of partially differentiable functionals, but they do not consider inference. Fang and Santos (2019) propose a novel bootstrap procedure for directionally differentiable functions. However, their inference procedure requires that the null parameter space is convex, which does not hold for union bounds.[4] This paper studies a specific non-convex null space, but the modified conditional procedure is potentially applicable to more general settings.

My paper widens the use of the conditional inference technique. There is

[4]Specifically, the space of λ_ℓ and λ_u under (7) is not convex.

a growing literature on conditional inference, see, e.g. Moreira (2003), Kleibergen (2005), Andrews and Mikusheva (2016), Andrews, Kitagawa, and McCloskey (2024), Andrews, Kitagawa, and McCloskey (2021), Andrews et al. (2023), and RR23, among others. I use their insights by constructing a conditional CI that has proper coverage under a subset of DGPs, and then modifying it with a lower truncation to guarantee uniform coverage. The modification is a novel contribution that is not used in current applications of conditional inference.

2 Setup and Examples

2.1 Setup

The goal of this paper is to construct a uniformly valid confidence interval for the target object θ, whose identified set is characterized as a union bound

$$\theta \in \left[\inf_{b\in\mathcal{B}} \lambda_{\ell,b},\ \sup_{b\in\mathcal{B}} \lambda_{u,b}\right]. \tag{1}$$

In this paper, λ_ℓ and λ_u are consistently estimable with an asymptotically normal estimator. $\mathcal{B}$ is a known finite set. Below I illustrate this setting in different examples.

2.2 Examples

Example 1. (Difference in Differences). RR23 study a more credible approach to the parallel trends assumption in DiD. To illustrate, consider a panel data model $t = -\underline{T}, ..., 1$. Let $\gamma \in \mathbb{R}^{\underline{T}+1}$ be a vector of "event study" coefficients, which can be decomposed as

$$\gamma = \begin{pmatrix} \gamma^{pre} \\ \gamma^{post} \end{pmatrix} = \begin{pmatrix} \xi^{pre} \\ \theta + \xi^{post} \end{pmatrix}.$$

The target object θ is the average treatment effect on the treated, and ξ is a bias from a difference in trend. Here θ and ξ^{post} are scalar, $\xi^{pre} = \left(\xi^{pre}_{-\underline{T}}, ..., \xi^{pre}_{-1}\right)$ and $\gamma_0 = \xi^{pre}_0$ is normalized to zero. Under parallel trends, $(\xi^{pre}, \xi^{post}) = 0$ and thus θ is point identified. However, this is a strong assumption that may not hold exactly. One type of relaxation is to assume that the violation of parallel trends at time

$t = 1$ is bounded above by the maximum pre-policy trend difference

$$\left|\xi^{post} - 0\right| \leq M \max_{t=-1,...,-\underline{T}} \left|\xi^{pre}_{t+1} - \xi^{pre}_{t}\right|, \tag{2}$$

where $M \geq 0$ is the degree of relaxation specified by the researcher. Manski and Pepper (2018) implement a similar concept with a natural benchmark $M = 1$ (see their Table 3). Under (2), the identified set of θ is a union bound in (1) with $\mathcal{B} = \{1, ..., 2\underline{T}\}$,

$$\lambda_{\ell,b} = \lambda_{u,b} = \begin{cases} \gamma^{post} + M\left(\gamma^{pre}_{-b+1} - \gamma^{pre}_{-b}\right) & \text{if } b = 1, ..., \underline{T}, \\ \gamma^{post} - M\left(\gamma^{pre}_{\underline{T}-b+1} - \gamma^{pre}_{\underline{T}-b}\right) & \text{if } b = \underline{T}+1, ..., 2\underline{T}. \end{cases} \tag{3}$$

Hasegawa et al. (2019), YKHS23, and Ban and Kedagni (2022) study different types of relaxations of the parallel trends assumption where the identified set is also characterized by union bounds. □

Example 2. (Bunching and Taxable Income Elasticity). Blomquist, Newey, Kumar, and Liang (2021) study the identification of the taxable income elasticity with bunching information. Assume that the after-tax income has two linear segments with slopes $\rho_1 > \rho_2$ and a kink at K, as illustrated in the left panel of Figure 1. Assume that the preference is specified as in Saez (2010) by the isoelastic utility function:

$$U(c, y, \xi) = c - \frac{\xi}{1 + 1/\theta}(y/\xi)^{1+1/\theta}, \xi > 0, \theta > 0,$$

where y is the before-tax income with density $f(y)$, $c = y - T(y)$ is the after-tax income, θ is the income elasticity and ξ represents the unobservable heterogeneity which is continuously distributed with density $g(\xi)$. Blomquist et al. (2021) show that without the restriction on $g(\xi)$, θ is not identified, but we can learn about θ with smoothness restrictions on $g(\xi)$. Consider a bunching interval $[y_1, y_2]$ containing the kink K, as in Figure 1 right panel. Let $\xi_1 = \rho_1^{-\theta} y_1$ and $\xi_2 = \rho_2^{-\theta} y_2$ denote lower and upper end points for ξ that correspond to y_1 and y_2, respectively. Under the assumption that

$$\sigma_\ell \min\{g(\xi_1), g(\xi_2)\} \leq g(\xi) \leq \sigma_u \max\{g(\xi_1), g(\xi_2)\} \text{ for } \xi \in [\xi_1, \xi_2] \tag{4}$$

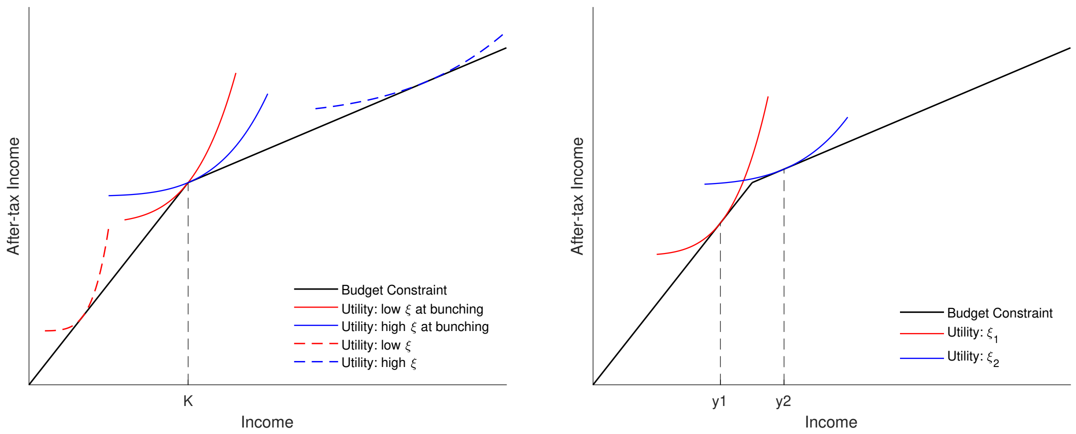


Figure 1: Left: Budget Constraint and Utility. Right: Bunching Interval.

for some $\sigma_u \geq 1 \geq \sigma_\ell > 0$, the identified set of θ is characterized by

$$\theta \in \left[\min_{b\in\mathcal{B}} \lambda_{\ell,b},\ \max_{b\in\mathcal{B}} \lambda_{u,b}\right] \cap \mathbb{R}_+$$

where $\mathcal{B} = \{1, 2\}$,

$$\lambda_{\ell,1} = \frac{\log\left(\frac{y_1}{y_2} + \frac{P(y_1 \leq Y \leq y_2)}{f^-(y_1)\sigma_u y_2}\right)}{\log \rho_1 - \log \rho_2}, \qquad \lambda_{\ell,2} = \frac{-\log\left(\frac{y_2}{y_1} - \frac{P(y_1 \leq Y \leq y_2)}{f^+(y_2)\sigma_u y_1}\right)}{\log \rho_1 - \log \rho_2},$$
$$\lambda_{u,1} = \frac{\log\left(\frac{y_1}{y_2} + \frac{P(y_1 \leq Y \leq y_2)}{f^-(y_1)\sigma_\ell y_2}\right)}{\log \rho_1 - \log \rho_2}, \qquad \lambda_{u,2} = \frac{-\log\left(\frac{y_2}{y_1} - \frac{P(y_1 \leq Y \leq y_2)}{f^+(y_2)\sigma_\ell y_1}\right)}{\log \rho_1 - \log \rho_2},$$

and $f^-(y_1) = \lim_{y\uparrow y_1} f(y)$, $\quad f^+ = \lim_{y\downarrow y_2} f(y)$.[5] The identified set of θ is restricted by $\mathbb{R}_+$, but if we have a valid CI for $\tilde{\theta}$ satisfying (1), the intersection of $\tilde{\theta}$'s CI and $\mathbb{R}_+$ is a valid CI for θ. Thus it suffices to consider inference for union bounds. Blomquist et al. (2021) focus on identification and put aside inference. □

Example 3. (Regression Discontinuity Design). Kolesár and Rothe (2018) study inference in regression discontinuity designs with a discrete running variable. Let $D = \mathbf{1}[X \geq 0]$ be a treatment indicator. Let $Y(1)$ and $Y(0)$ denote the potential outcome with and without the treatment, and $Y = DY(1) + (1-D)Y(0)$ denote the observed outcome. Let $\mu(X) = E[Y \mid X]$. The average treatment effect

[5] This is derived from Blomquist et al. (2021) equation (8), and I assume for simplicity that the terms in the log are strictly positive.

at the threshold is

$$\theta = E\left[Y(1) - Y(0)\,|X = 0\right] = \lim_{x \downarrow 0} \mu(x) - \lim_{x \uparrow 0} \mu(x).$$

A standard approach to estimate θ is to run a local OLS regression of Y on polynomial $m(X)$ with $X \in [-h, h]$ where

$$m(x) = \left(\mathbf{1}\left[x \geq 0\right], \mathbf{1}\left[x \geq 0\right] x, \ldots, \mathbf{1}\left[x \geq 0\right] x^p, 1, x, \ldots, x^p\right)'.$$

Let γ_h be the regression coefficient and $\theta_h = (1, 0, ..., 0)\gamma_h$. If X is continuous, the bias $\xi(x) = \mu(x) - m(x)'\gamma_h$ is negligible if we choose $h \to 0$ at a sufficiently fast rate as the sample size increases. However, if X is discrete, this "undersmoothing" procedure is not feasible. Kolesár and Rothe (2018) propose an honest CI under restrictions that the bias at the threshold are bounded above by the specification errors at other support points, i.e.

$$\left|\lim_{x \uparrow 0} \xi(x)\right| \leq \max_{\tilde{x} \in S_X^-} |\xi(\tilde{x})|\,, \quad \left|\lim_{x \downarrow 0} \xi(x)\right| \leq \max_{\tilde{x} \in S_X^+} |\xi(\tilde{x})|\,,$$

where $S_X^- = S_X \cap [-h, 0)$, $S_X^+ = S_X \cap [0, h]$ and S_X is the support of X. Under this restriction, the identified set of θ is characterized by (1) with

$$\begin{aligned}
&\mathcal{B} = \left\{(s_\ell, s_u, x_\ell, x_u) : s_\ell, s_u \in \{-1, 1\}, x_\ell \in S_X^-, x_u \in S_X^+\right\}, \\
&\lambda_\ell(s_\ell, s_u, x_\ell, x_u) = \lambda_u(s_\ell, s_u, x_\ell, x_u) = \theta_h + s_\ell \xi(x_\ell) + s_u \xi(x_u).
\end{aligned}$$

Kolesár and Rothe (2018) use the simple CI based on union principle for inference. □

Example 4. (Misspecification Analysis). Masten and Poirier (2021) provide a constructive way for researchers to salvage a falsified instrumental variable model. Consider the classical linear model with multiple instruments:

$$Y = X\theta + Z'\gamma + U,$$

where Y is the outcome, X is a scalar endogenous variable and Z is a $L \times 1$ vector of potentially invalid instruments. Under (i) exogeneity $\mathrm{cov}(Z, U) = 0$, (ii) exclusion $\gamma = 0$ and (iii) a proper rank condition, we can point identify θ. However, if either the exogeneity or exclusion restriction does not hold, the model may be falsified. In this context, Masten and Poirier (2021) suggest relaxing the model by

$$\Theta(\xi) = \left\{\theta \in \mathbb{R} : -\xi \mathbf{1}_{L\times 1} \leq \mathrm{var}(Z)^{-1} \left(\mathrm{cov}(Z,Y) - \mathrm{cov}(Z,X)\theta\right) \leq \xi \mathbf{1}_{L\times 1}\right\},$$

where $\xi \geq 0$ measures the level of relaxation and the inequalities hold element wise. The authors suggest reporting the falsification adaptive set $\Theta(\underline{\xi})$, where $\underline{\xi}$ is the minimum relaxation such that $\Theta(\xi)$ is non-empty. In addition, $\Theta(\underline{\xi})$ is characterized by (1), with

$$\lambda_{\ell,b} = \lambda_{u,b} = \psi_b/\pi_b,$$

ψ_b and π_b are the b-th element of $\psi = \mathrm{var}(Z)^{-1}\mathrm{cov}(Z,Y)$, $\pi = \mathrm{var}(Z)^{-1}\mathrm{cov}(Z,X)$, and $\mathcal{B} = \{b = 1, ..., L : \pi_b \neq 0\}$. In their empirical application, the authors implicitly assume that either $\pi_b = 0$ or $|\pi_b| \geq \varepsilon > 0$ for all b, so that $\mathcal{B}$ is consistently estimable, in the sense that the Hausdorff distance between $\hat{\mathcal{B}}$ and $\mathcal{B}$ converges to zero in probability. Therefore, asymptotically we can treat $\mathcal{B}$ as known. Apfel and Windmeijer (2022) propose a generalized falsification adaptive set, which also has a union bound characterization. Both papers do not consider inference.

Stoye (2020) studies inference for interval identified parameters under misspecification. The identified set for θ is $[\theta_L, \theta_U]$, and this set is empty when $\theta_L > \theta_U$. Stoye (2020) suggests reporting the misspecification robust identified set

$$[\theta_L, \theta_U] \cup \left\{\frac{\sigma_U\theta_L + \sigma_L\theta_U}{\sigma_L + \sigma_U}\right\}, \tag{5}$$

where σ_L and σ_U are the asymptotic standard deviations for estimators $\hat{\theta}_L$ and $\hat{\theta}_U$. In this case, the identified set is a union bound in (1) with $\mathcal{B} = \{1, 2\}$,

$$\lambda_{\ell,1} = \theta_L, \quad \lambda_{u,1} = \theta_U, \quad \lambda_{\ell,2} = \lambda_{u,2} = \frac{\sigma_U\theta_L + \sigma_L\theta_U}{\sigma_L + \sigma_U}.$$

Stoye (2020) proposes a CI for (5), but it does not apply to general union bounds.

□

3 Inference Procedure

In this section, I study inference on θ in (1). I illustrate with a normally distributed estimator $\hat{\lambda}_n = \left(\hat{\lambda}_\ell, \hat{\lambda}_u\right)$ such that

$$\begin{pmatrix} \hat{\lambda}_\ell \\ \hat{\lambda}_u \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \lambda_\ell \\ \lambda_u \end{pmatrix}, \Sigma_n\right),\ \Sigma_n = \begin{bmatrix} \Sigma_{\ell,n} & \Sigma'_{\ell u,n} \\ \Sigma_{\ell u,n} & \Sigma_{u,n} \end{bmatrix} \tag{6}$$

with Σ_n known, where $\Sigma_{\ell,n}$, $\Sigma_{u,n}$ and $\Sigma_{\ell u,n}$ are $|\mathcal{B}| \times |\mathcal{B}|$ matrices. The true value $\lambda = (\lambda_\ell, \lambda_u) \in \Lambda$ and Λ can be a lower dimensional subspace of $\mathbb{R}^{2|\mathcal{B}|}$, e.g. as in Example 1. In general, the normality holds asymptotically with appropriate scaling, and the asymptotic variance can be consistently estimated.

I construct a modified conditional CI by inverting the test of the null hypothesis

$$H_0 : \min_{b\in\mathcal{B}} \lambda_{\ell,b} \leq \theta \leq \max_{b\in\mathcal{B}} \lambda_{u,b}. \tag{7}$$

The test takes the form

$$\phi\left(\theta, \hat{\lambda}_n, \Sigma_n\right) = \mathbf{1}\left[\hat{T}(\theta) > \hat{c}^{\mathrm{m}}(\theta; \alpha)\right],$$

where $\hat{T}(\theta)$ is the test statistic, and θ is rejected if $\hat{T}(\theta)$ exceeds the modified conditional critical value $\hat{c}^{\mathrm{m}}(\theta; \alpha)$. Consequently, the corresponding $1-\alpha$ CI is

$$CI^{\mathrm{m}}(\hat{\lambda}_n, \Sigma_n; \alpha) = \left[\inf_{\phi(\theta,\hat{\lambda}_n,\Sigma_n)=0} \theta,\ \sup_{\phi(\theta,\hat{\lambda}_n,\Sigma_n)=0} \theta\right]. \tag{8}$$

3.1 The Test Statistic

The test statistic has a max-min form

$$\hat{T}(\theta) = \max\left\{\min_{b\in\mathcal{B}} \mathcal{Z}_{\ell,b},\ \min_{b\in\mathcal{B}} \mathcal{Z}_{u,b}\right\} \tag{9}$$

where $\sigma_{\ell,b} = \sqrt{\Sigma_{\ell,bb}}$, $\sigma_{u,b} = \sqrt{\Sigma_{u,bb}}$,

$$\mathcal{Z}_{\ell,b} = \frac{\hat{\lambda}_{\ell,b} - \theta}{\sigma_{\ell,b}}, \text{ and } \mathcal{Z}_{u,b} = \frac{\theta - \hat{\lambda}_{u,b}}{\sigma_{u,b}}. \tag{10}$$

Observing that H_0 in (7) is equivalent to

$$H_0 : \max\left\{\min_{b\in\mathcal{B}}(\lambda_{\ell,b}-\theta),\ \min_{b\in\mathcal{B}}(\theta-\lambda_{u,b})\right\} \le 0, \tag{11}$$

and the test statistic is constructed by replacing λ_ℓ and λ_u in (11) by their estimator, adjusted for the standard deviation. Put another way, the population version of $\hat{T}(\theta)$, which replaces $(\hat{\lambda}_\ell, \hat{\lambda}_u)$ with $(\lambda_\ell, \lambda_u)$, is non-positive if and only if H_0 holds.

If we use a simple critical value $c^{\text{sim}} = \Phi^{-1}(1-\frac{\alpha}{2})$, then we will get a simple CI

$$CI^{\text{sim}} = \left[\min_{b\in\mathcal{B}}\hat{\lambda}_{\ell,b} - \sigma_{\ell,b}\Phi^{-1}(1-\frac{\alpha}{2}),\ \max_{b\in\mathcal{B}}\hat{\lambda}_{u,b} + \sigma_{u,b}\Phi^{-1}(1-\frac{\alpha}{2})\right], \tag{12}$$

which is often used in current practice, see e.g. Kolesár and Rothe (2018), Hasegawa et al. (2019), Ban and Kedagni (2022). The simple CI is uniformly valid under mild conditions, see Proposition 2 in Kolesár and Rothe (2018). However, in general, it can be very conservative. To illustrate, define

$$b_\ell = \arg\min_{b\in\mathcal{B}}\lambda_{\ell,b}, \quad b_u = \arg\max_{b\in\mathcal{B}}\lambda_{u,b}. \tag{13}$$

and observe that

$$\begin{aligned}
P\left(\theta \notin CI^{\text{sim}}\right) &= P\left(\max\left\{\min_{b\in\mathcal{B}}\mathcal{Z}_{\ell,b},\ \min_{b\in\mathcal{B}}\mathcal{Z}_{u,b}\right\} > \Phi^{-1}(1-\frac{\alpha}{2})\right)\\
&\le P\left(\max\left\{\mathcal{Z}_{\ell,b_\ell},\ \mathcal{Z}_{u,b_u}\right\} > \Phi^{-1}(1-\frac{\alpha}{2})\right)\\
&\le P\left(\mathcal{Z}_{\ell,b_\ell} > \Phi^{-1}(1-\frac{\alpha}{2})\right) + P\left(\mathcal{Z}_{u,b_u} > \Phi^{-1}(1-\frac{\alpha}{2})\right) && (14)\\
&= P\left(\frac{\hat{\lambda}_{\ell,b_\ell}-\lambda_{\ell,b_\ell}}{\sigma_{\ell,b}} + \frac{\lambda_{\ell,b_\ell}-\theta}{\sigma_{\ell,b}} > \Phi^{-1}(1-\frac{\alpha}{2})\right) && (15)\\
&\quad + P\left(\frac{\lambda_{u,b_u}-\hat{\lambda}_{u,b_u}}{\sigma_{u,b}} + \frac{\theta-\lambda_{u,b_u}}{\sigma_{u,b}} > \Phi^{-1}(1-\frac{\alpha}{2})\right) && (16)\\
&\le \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.
\end{aligned}$$

Here the first inequality holds because I replace the minimum of $\mathcal{Z}_\ell$ and $\mathcal{Z}_u$ by the value at b_ℓ and b_u, which may not be the realized minimizers. The second inequality follows from $P(A \cup B) \leq P(A) + P(B)$. The final inequality holds under the null hypothesis (7).

The potential conservativeness comes mainly from the first and last inequalities. The first inequality tends to be conservative when the minimum λ_{ℓ,b_ℓ} is close to other elements in λ_ℓ. In such cases, we should consider the minimum of the vector $\mathcal{Z}_\ell$ instead of merely $\mathcal{Z}_{\ell,b_\ell}$. The same reasoning applies to the upper bound. The last inequality becomes conservative if the union bound is wide, i.e. $\lambda_{u,b_u} - \lambda_{\ell,b_\ell} \gg \max\{\sigma_{u,b_u}, \sigma_{\ell,b_\ell}\}$. In such cases, either (15) or (16) is negligible, allowing us to replace $\Phi^{-1}(1-\frac{\alpha}{2})$ with $\Phi^{-1}(1-\alpha)$. This scenario is also studied in Imbens and Manski (2004) and Stoye (2009) for a single bound where $|\mathcal{B}| = 1$. Besides the first and last inequalities, the simple CI is also conservative because (14) does not fully use the joint distribution of $(\mathcal{Z}_{\ell,b_\ell}, \mathcal{Z}_{u,b_u})$.

That said, the simple critical value is near optimal in less favorable cases, where both the minimum and maximum are well separated, and the length of the identified set is short, i.e.

$$\min_{b \in \mathcal{B} \setminus b_\ell} \frac{\lambda_{\ell,b} - \lambda_{\ell,b_\ell}}{\sigma_{\ell,b}} \gg 0, \quad \min_{b \in \mathcal{B} \setminus b_u} \frac{\lambda_{u,b_u} - \lambda_{u,b}}{\sigma_{u,b}} \gg 0, \quad \frac{\lambda_{u,b_u} - \lambda_{\ell,b_\ell}}{\min\{\sigma_{\ell,b_\ell}, \sigma_{u,b_u}\}} \approx 0. \tag{17}$$

In such scenarios, the first and last inequalities are close to equality, mitigating any significant power loss. This implies that c^{sim} is nearly optimal among constant critical values because it protects against the less favorable distributions, although at the cost of an inflated coverage rate against more favorable DGPs. Therefore, it is crucial to devise a data-dependent critical value that ensures proper coverage under case (17) but is more efficient under other DGPs.

3.2 Conditional Critical Value

Following from the previous discussion, I now construct a data dependent critical value that is valid under less favorable DGPs and more efficient otherwise. To do so, note that under less favorable DGPs in (17),

$$P(E_\ell \cup E_u) \approx 1 \tag{18}$$

where[6]

$$
\begin{aligned}
E_\ell &= \left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,\hat{b}_\ell}\right\} \cap \left\{\lambda_{\ell,\hat{b}_\ell} \le \theta\right\}, \\
E_u &= \left\{\hat{T}(\theta) = \mathcal{Z}_{u,\hat{b}_u}\right\} \cap \left\{\lambda_{u,\hat{b}_u} \ge \theta\right\}, \\
\hat{b}_\ell &= \arg\min_{b\in\mathcal{B}} \mathcal{Z}_{\ell,b},\ \hat{b}_u = \arg\min_{b\in\mathcal{B}} \mathcal{Z}_{u,b}.
\end{aligned} \tag{19}
$$

If the critical value $\hat{c}(\theta)$ satisfies

$$
P\left(\hat{T}(\theta) > \hat{c}(\theta)\,|E_\ell \cup E_u\right) \le \alpha^{\mathrm{c}} < \alpha, \tag{20}
$$

the unconditional rejection rate is bounded above by α following from (18). Therefore, I construct a conditional critical value based on the conditional distribution

$$
\hat{T}(\theta)\left|\hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\right. \text{ and } \hat{T}(\theta)\left|\hat{T}(\theta) = \mathcal{Z}_{u,b_2}\right.
$$

for b_1, b_2 satisfying $\lambda_{\ell,b_1} \le \theta \le \lambda_{u,b_2}$.

Lemma 1. *Under H_0, assume that (6) holds and*

$$
P\left(\mathcal{Z}_{\ell,\hat{b}_\ell} = \mathcal{Z}_{u,\hat{b}_u}\right) = 0. \tag{21}
$$

Let b_1, b_2 satisfy $\lambda_{\ell,b_1} \le \theta \le \lambda_{u,b_2}$, then

$$
\begin{aligned}
\frac{\Phi\left(\hat{T}(\theta)\right) - \Phi\left(t_{\ell,1}(\theta,b_1)\right)}{\Phi\left(t_{\ell,2}(\theta,b_1)\right) - \Phi\left(t_{\ell,1}(\theta,b_1)\right)}\left|\left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\right\}\right. &\overset{\text{FOSD}}{\preceq} \text{Unif}(0,1) \\
\frac{\Phi\left(\hat{T}(\theta)\right) - \Phi\left(t_{u,1}(\theta,b_1)\right)}{\Phi\left(t_{u,2}(\theta,b_1)\right) - \Phi\left(t_{u,1}(\theta,b_1)\right)}\left|\left\{\hat{T}(\theta) = \mathcal{Z}_{u,b_2}\right\}\right. &\overset{\text{FOSD}}{\preceq} \text{Unif}(0,1)
\end{aligned} \tag{22}
$$

where

$$
t_{\ell,1}(\theta,b) = \begin{cases} \min_{\tilde{b}\in\mathcal{B}} \left(1+\rho_{\ell u}(b,\tilde{b})\right)^{-1}\left(\mathcal{Z}_{u,\tilde{b}} + \rho_{\ell u}(b,\tilde{b})\mathcal{Z}_{\ell,b}\right), & \textit{if } \min_{\tilde{b}\in\mathcal{B}} \rho_{\ell u}(b,\tilde{b}) > -1 \\ -\infty & \textit{otherwise} \end{cases}
$$

[6] If the minimizer of $\mathcal{Z}_\ell$ is not unique, define $\hat{b}_\ell$ as the smallest element of the minimizers, with an analogous definition for $\hat{b}_u$.

$$t_{u,1}(\theta, b) = \begin{cases} \min_{\tilde{b}\in\mathcal{B}} \left(1 + \rho_{\ell u}(\tilde{b}, b)\right)^{-1} \left(\mathcal{Z}_{\ell,\tilde{b}} + \rho_{\ell u}(\tilde{b}, b)\mathcal{Z}_{u,b}\right), & \text{if } \min_{\tilde{b}\in\mathcal{B}} \rho_{\ell u}(\tilde{b}, b) > -1 \\ -\infty & \text{otherwise} \end{cases}$$

$$t_{\ell,2}(\theta, b) = \begin{cases} \min_{\tilde{b}\in\mathcal{B}:\rho_\ell(b,\tilde{b})<1} \left(1 - \rho_\ell(b, \tilde{b})\right)^{-1} \left(\mathcal{Z}_{\ell,\tilde{b}} - \rho_\ell(b, \tilde{b})\mathcal{Z}_{\ell,b}\right) & \text{if } \min_{\tilde{b}\in\mathcal{B}} \rho_\ell(b, \tilde{b}) < 1 \\ +\infty & \text{otherwise} \end{cases}$$

$$t_{u,2}(\theta, b) = \begin{cases} \min_{\tilde{b}\in\mathcal{B}:\rho_u(\tilde{b},b)<1} \left(1 - \rho_u(\tilde{b}, b)\right)^{-1} \left(\mathcal{Z}_{u,\tilde{b}} - \rho_u(\tilde{b}, b)\mathcal{Z}_{u,b}\right) & \text{if } \min_{\tilde{b}\in\mathcal{B}} \rho_u(b, \tilde{b}) < 1 \\ +\infty & \text{otherwise} \end{cases}$$

$$\rho_\ell(b_1, b_2) = \frac{\Sigma_{\ell,b_1b_2}}{\sigma_{\ell,b_1}\sigma_{\ell,b_2}}, \quad \rho_u(b_1, b_2) = \frac{\Sigma_{u,b_1b_2}}{\sigma_{u,b_1}\sigma_{u,b_2}}, \quad \rho_{\ell u}(b_1, b_2) = \frac{\Sigma_{\ell u,b_1b_2}}{\sigma_{\ell,b_1}\sigma_{u,b_2}},$$

and $\mathcal{Z}_\ell$, $\mathcal{Z}_u$ are defined in (10).

Loosely speaking, Lemma 1 implies that the distribution of $\hat{T}(\theta)$ conditional on $\hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}$ is first order stochastically dominated by a truncated normal distribution $\mathcal{TN}(0, 1, [t_{\ell,1}(\theta, b_1), t_{\ell,2}(\theta, b_1)])$, where $\mathcal{TN}(\mu, \sigma^2, [t_1, t_2])$ is a normal distribution $\mathcal{N}(\mu, \sigma^2)$ truncated at $[t_1, t_2]$. Hence, we can guarantee conditional coverage by using the $1-\alpha^c$ quantile of $\mathcal{TN}(0, 1, [t_{\ell,1}(\theta, b), t_{\ell,2}(\theta, b)])$ with $\alpha^c < \alpha$. Condition (21) holds in most examples previously discussed and is assumed in Proposition 1 for simplicity.

Define the conditional critical value $\hat{c}^c(\theta, \alpha^c)$ as:

$$\hat{c}^c(\theta, \alpha^c) = \begin{cases} \Phi^{-1}\Big(\alpha^c\Phi\big(t_{\ell,1}(\theta, \hat{b}_\ell)\big) + (1 - \alpha^c)\Phi\big(t_{\ell,2}(\theta, \hat{b}_\ell)\big)\Big) & \text{if } \mathcal{Z}_{\ell,\hat{b}_\ell} \geq \mathcal{Z}_{u,\hat{b}_u} \\ \Phi^{-1}\Big(\alpha^c\Phi\big(t_{u,1}(\theta, \hat{b}_u)\big) + (1 - \alpha^c)\Phi\big(t_{u,2}(\theta, \hat{b}_u)\big)\Big) & \text{if } \mathcal{Z}_{\ell,\hat{b}_\ell} < \mathcal{Z}_{u,\hat{b}_u} \end{cases} \tag{23}$$

where $\alpha^c \in (\frac{1}{2}\alpha, \alpha)$ is a user chosen tuning parameter, with a suggested rule of thumb value $\frac{4}{5}\alpha$. As we will see later, α^c trades off the power under more and less favorable DGPs.

Proposition 1. *Under H_0, (6) and (21), it holds that*

$$P\left(\hat{T}(\theta) > \hat{c}^c(\theta, \alpha^c) \,\middle|\, E_\ell \cup E_u\right) \leq \alpha^c. \tag{24}$$

Under (21), $E_\ell \cup E_u$ can be partitioned into $\{\hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\}$ and $\{\hat{T}(\theta) = \mathcal{Z}_{u,b_2}\}$ for b_1, b_2 satisfying $\lambda_{\ell,b_1} \leq \theta \leq \lambda_{u,b_2}$. Hence (24) follows directly from Lemma 1. Under more favorable DGPs diverging from (17), $\hat{c}^c(\theta, \alpha^c)$ can be significantly smaller than $\Phi^{-1}(1-\frac{\alpha}{2})$. To see this, let $\theta = \lambda_{\ell,b_\ell}$ be the lower bound of the identified set and assume that $\hat{T}(\theta) = \mathcal{Z}_{\ell,\hat{b}_\ell}$. If the identified set is very large relative to the standard deviation, we have

$$\begin{aligned} t_{\ell,1}(\theta,\hat{b}_\ell) &\leq \left(1+\rho_{\ell u}(\hat{b}_\ell,b_u)\right)^{-1}\left(\mathcal{Z}_{u,b_u}+\rho_{\ell u}(\hat{b}_\ell,b_u)\mathcal{Z}_{\ell,\hat{b}_\ell}\right) \\ &= \left(1+\rho_{\ell u}(\hat{b}_\ell,b_u)\right)^{-1}\left(\frac{\hat{\lambda}_{\ell,\hat{b}_\ell}-\hat{\lambda}_{u,b_u}}{\sigma_{u,b_u}}+\left(\rho_{\ell u}(\hat{b}_\ell,b_u)-\frac{\sigma_{\ell,\hat{b}_\ell}}{\sigma_{u,b_u}}\right)\mathcal{Z}_{\ell,\hat{b}_\ell}\right)\approx-\infty, \end{aligned} \tag{25}$$

where the approximation $\approx$ follows from $\frac{\hat{\lambda}_{\ell,\hat{b}_\ell}-\hat{\lambda}_{u,b_u}}{\sigma_{u,b_u}}\approx-\infty$. In this case,

$$\hat{c}^{\mathrm{c}}(\theta,\alpha^{\mathrm{c}})\approx\Phi^{-1}\left((1-\alpha^{\mathrm{c}})\Phi\left(t_{\ell,2}(\theta,\hat{b}_\ell)\right)\right)\leq\Phi^{-1}\left(1-\alpha^{\mathrm{c}}\right)<\Phi^{-1}(1-\frac{\alpha}{2}).$$

Moreover, if the minimum $\lambda_{\ell,-b_\ell}$ is not well separated from λ_{ℓ,b_ℓ}, then the upper bound $t_{\ell,2}(\theta,\hat{b}_\ell)$ will be the minimum of several random variables, which will further reduce the critical value. I next illustrate the conditional critical value using a simple example.

Example 5. (Simple Union Bounds) Consider a simple union bound

$$\theta\in\left[\min\left\{\lambda_1,\lambda_2\right\},\ \max\left\{\lambda_1,\lambda_2\right\}\right]$$

and the estimator satisfies

$$\left(\hat{\lambda}_1-\lambda_1,\hat{\lambda}_2-\lambda_2\right)\sim\mathcal{N}\left(0,\mathcal{I}_2\right).$$

The test statistic has the form

$$\hat{T}(\theta)=\max\left\{\min\left\{\hat{\lambda}_1-\theta,\hat{\lambda}_2-\theta\right\},\min\left\{\theta-\hat{\lambda}_1,\theta-\hat{\lambda}_2\right\}\right\}.$$

Assume that $\hat{T}(\theta)=\hat{\lambda}_1-\theta$. In this case, the conditional critical value is

$$\hat{c}^{\mathrm{c}}(\theta;\alpha^{\mathrm{c}})=\Phi^{-1}\left((1-\alpha^{\mathrm{c}})\Phi\left(\hat{\lambda}_2-\theta\right)+\alpha^{\mathrm{c}}\Phi\left(\theta-\hat{\lambda}_2\right)\right)<\Phi^{-1}\left(1-\alpha^{\mathrm{c}}\right).$$

If the minimizer and maximizer are well separated, e.g. $\lambda_1=\theta$ and $\lambda_2\to\infty$, the efficient critical value is $\Phi^{-1}(1-\alpha)$, as discussed in Imbens and Manski (2004). In this case, $\hat{\lambda}_2$ will be large and $\hat{c}^{\mathrm{c}}(\theta;\alpha^{\mathrm{c}})\to\Phi^{-1}(1-\alpha^{\mathrm{c}})$, which is slightly conser-

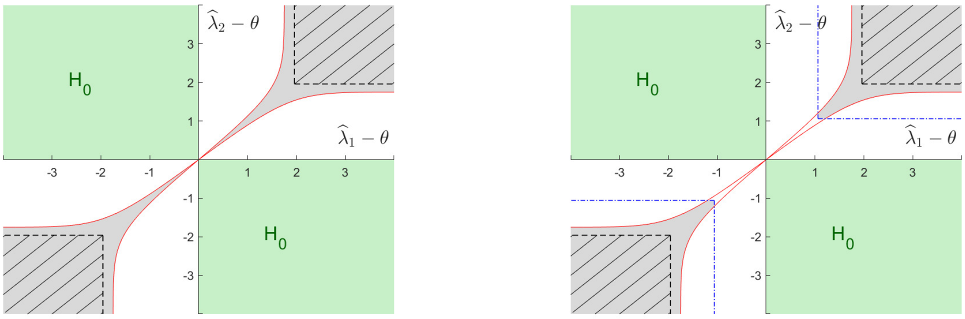


Figure 2: Left: Conditional Critical Value. Right: Modified Conditional Critical Value.

vative. This follows from the fact that the conditional critical value is designed to correct for the case when all elements, except for b_ℓ and b_u, are far away from binding. On the other hand, if $\hat{\lambda}_2$ is relatively small, then the critical value is smaller. In Figure 2 left panel, I plot the rejection region for the simple and conditional critical values with $\alpha = 0.05$ and $\alpha^{\mathrm{c}} = 0.04$. The green region denotes the null parameter space.[7] The red curve is the boundary corresponding to the conditional critical value, and the grey region is the rejection region for the conditional critical value. Finally, the two square regions filled with lines are the rejection region of the simple test. The rejection region of the conditional test is strictly larger than the simple test. □

It is important to note that $\hat{c}^{\mathrm{c}}(\theta, \alpha^{\mathrm{c}})$ may not serve as a valid critical value, because $P\left(E_\ell \cup E_u\right)$ can be much smaller than one when moving away from (17). For that reason, next, I show how to construct a uniformly valid modified conditional critical value that retains favorable power properties relative to the simple critical value.

3.3 The Modified Conditional Critical Value

I introduce a novel modification to the conditional critical value:

$$\hat{c}^{\mathrm{m}}(\theta; \alpha) = \tilde{c}^{\mathrm{m}}(\theta, \hat{c}^{\mathrm{t}}; \alpha) = \max\left\{\hat{c}^{\mathrm{c}}(\theta, \alpha^{\mathrm{c}}), \hat{c}^{\mathrm{t}}\right\} \tag{26}$$

[7] The null parameter space mirrors that of testing for sign agreement, see e.g. Miller, Molinari, and Stoye (2024) and Kim (2024). However, with more than two variables, these two testing problems do not nest within each other.

where $\hat{c}^{\mathrm{t}}$ is defined later in (30).

To illustrate the construction of the truncation $\hat{c}^{\mathrm{t}}$, let $\widetilde{CI}^{\mathrm{m}}(c)$ be the CI based on (8) with $\hat{c}^{\mathrm{m}}(\theta;\alpha)$ replaced by $\tilde{c}^{\mathrm{m}}(\theta,c;\alpha)$. Given a potential true value λ, the rejection rate at θ is

$$p(c;\theta,\lambda,\Sigma) = P\left(\theta \notin \widetilde{CI}^{\mathrm{m}}(c); \mathcal{N}(\lambda,\Sigma)\right),$$

where $P(\cdot;\mathcal{N}(\lambda,\Sigma))$ is the probability under (6). It suffices to define the lower truncation as the minimum value that achieves uniform size control, i.e.

$$c^{\mathrm{t}}(\theta) = \inf\left\{c \in \mathbb{R}_+ : \sup_{\lambda\in\Lambda_0(\theta)} p(c;\theta,\lambda,\Sigma) \le \alpha\right\}, \tag{27}$$

where Λ_0 is the set of feasible λ satisfying H_0:

$$\Lambda_0(\theta) = \left\{(\lambda_\ell,\lambda_u) \in \Lambda : \min_{b\in\mathcal{B}} \lambda_{\ell,b} \le \theta \le \max_{b\in\mathcal{B}} \lambda_{u,b}\right\}.$$

Note that $c^{\mathrm{t}}(\theta) \le c^{\mathrm{sim}} = \Phi^{-1}(1-\frac{\alpha}{2})$ because $p(c^{\mathrm{sim}};\theta,\lambda,\Sigma) \le \alpha$ from the discussion in Section 3.1. In fact $\hat{c}^{\mathrm{t}}(\theta)$ is usually significantly smaller than c^{sim}. The intuition is that by virtue of Lemma 1, truncation is unnecessary for DGPs such that

$$P\left(E_\ell \cup E_u; \mathcal{N}(\lambda,\Sigma)\right) \ge \frac{1-\alpha}{1-\alpha^c} \tag{28}$$

with $\alpha^{\mathrm{c}} < \alpha$, where E_ℓ, E_u are defined in (19). Thus we only need to consider truncation in more favorable DGPs deviating from (17), i.e. when the minimizer or maximizer is not well separated, in which case a smaller critical value suffices. Given (θ,λ), we can calculate $p(c;\theta,\lambda,\Sigma)$ by simulation. Nevertheless, calculating $\widetilde{CI}^{\mathrm{m}}(c^{\mathrm{t}})$ can be time consuming because (i) we need to calculate $c^{\mathrm{t}}(\theta)$ for a grid of θ to get the CI and (ii) $\Lambda_0(\theta)$ is an unbounded set, which slows the computation down.

To improve computational efficiency, I propose a lower truncation that does not depend on θ. First, note that for given λ, either $\theta \in [\theta_\ell,\theta_m]$ or $\theta \in [\theta_m,\theta_u]$, where $\theta_\ell = \lambda_{\ell,b_\ell}$, $\theta_u = \lambda_{u,b_u}$ and $\theta_m = (\theta_\ell+\theta_u)/2$. As a result, we can bound $p(c;\theta,\lambda,\Sigma)$ by

$$p(c;\theta,\lambda,\Sigma) \le \max\left\{P\left([\theta_\ell,\theta_m] \not\subseteq \widetilde{CI}^{\mathrm{m}}(c);\mathcal{N}(\lambda,\Sigma)\right), P\left([\theta_m,\theta_u] \not\subseteq \widetilde{CI}^{\mathrm{m}}(c);\mathcal{N}(\lambda,\Sigma)\right)\right\}$$

$$\leq \max \left\{ P\left(\hat{T}(\theta_\ell) > \tilde{c}^{\mathrm{m}}(\theta_\ell, c) \text{ or } \left\{ \hat{T}(\theta_m) > \tilde{c}^{\mathrm{m}}(\theta_m, c) \text{ and } \hat{T}(\theta_u) > \tilde{c}^{\mathrm{m}}(\theta_u, c) \right\}; \mathcal{N}(\lambda, \Sigma) \right), \right.$$
$$\left. P\left(\hat{T}(\theta_u) > \tilde{c}^{\mathrm{m}}(\theta_u, c) \text{ or } \left\{ \hat{T}(\theta_m) > \tilde{c}^{\mathrm{m}}(\theta_m, c) \text{ and } \hat{T}(\theta_\ell) > \tilde{c}^{\mathrm{m}}(\theta_\ell, c) \right\}; \mathcal{N}(\lambda, \Sigma) \right) \right\}$$
$$=: \bar{p}(c, \lambda, \Sigma). \tag{29}$$

Therefore, it is valid, but conservative, to replace $p(c; \theta, \lambda, \Sigma)$ in (27) with $\bar{p}(c, \lambda, \Sigma)$. In addition, to avoid maximization over an unbounded set, let $\hat{\Lambda}$ be a $1 - \eta$ compact confidence set of λ, with suggested value $\eta = 0.001$. In sum, it suffices to use

$$\hat{c}^{\mathrm{t}} = \inf_c \left\{ c \geq 0 : \sup_{\lambda \in \hat{\Lambda}_\eta} \bar{p}(c, \lambda, \Sigma) + \eta \leq \alpha \right\}, \tag{30}$$

and $\bar{p}(c, \lambda, \Sigma)$ is defined in (29). In terms of computation, $\bar{p}(c, \lambda, \Sigma)$ can be conveniently calculated via simulation, and we only need to calculate the maximization over a bounded set once rather than for a grid of θ. In general, with η small enough, $\hat{c}^{\mathrm{t}}$ is much smaller than $\Phi^{-1}(1 - \frac{\alpha}{2})$ by the intuition explained around (28). Moreover, in many examples, the feasible space Λ is a lower dimensional subspace of $\mathbb{R}^{2|\mathcal{B}|}$, so that the supremum is taken over a space much smaller than $\mathbb{R}^{2|\mathcal{B}|}$, which reduces the computational cost.

The lower truncation $\hat{c}^{\mathrm{t}}$ is more likely to bind under more favorable DGPs, and it increases in the tuning parameter α^{c}. Hence, α^{c} trades off the power between more and less favorable DGPs. A larger α^{c} leads to higher power under less favorable DGPs, while a smaller α^{c} leads to higher power under more favorable DGPs. It is possible to choose an optimal α^{c} by, e.g., maximizing the weighted average power. I leave this to future research.

Remark 1. The relaxation in (29) to get $\bar{p}(c, \lambda, \Sigma)$ is not overly conservative. To see this, if the identified set is large, θ_m will be covered by the modified conditional confidence interval with probability close to 1, so the conservativeness introduced by this relaxation is negligible.[8] Conversely, if the identified set is very small, then the set's coverage will be similar to the coverage of a point. Moreover, we can reduce conservativeness by increasing the number of elements in the partition at the cost of increased computational difficulty.

Example 5. (Simple Union Bounds, Cont.) For simplicity, in this example I let $\eta = 0$. With $\alpha = 0.05$ and $\alpha^{\mathrm{c}} = 0.04$, we can calculate that $\hat{c}^{\mathrm{t}} = 1.06$. In Figure

[8] By construction $\hat{c}^{\mathrm{m}} \geq 0$, thus if $\sqrt{n}\,(\lambda_{u,b_u} - \lambda_{\ell,b_\ell}) \to \infty$, we have $P\left(\hat{\lambda}_{\ell,\hat{b}_\ell} \leq \theta_m \leq \hat{\lambda}_{u,\hat{b}_u}\right) \to 1$.

2 right panel, I plot the rejection region of the modified conditional test. The blue dotted curve is the boundary corresponding to the lower truncation $\hat{c}^{\mathrm{t}}$, and the grey region is the rejection region for the modified conditional test. The rest are the same as in Figure 2 left panel. As we can see, the rejection region of the modified conditional test is strictly larger than the simple test, leading to power improvements. Compared to the simple rejection region, the conditional test also rejects if both $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are small. The intuition is that if both λ_1 and λ_2 are close to zero, then there are multiple approximate minimizers and maximizers, so we only need a small critical value. The lower truncation $\hat{c}^{\mathrm{t}}$ removes some counter-intuitive values from the rejection region close to H_0, e.g. $\left(\hat{\lambda}_1 - \theta, \hat{\lambda}_2 - \theta\right) = (\varepsilon, \varepsilon) \approx (0, 0)$. □

3.4 Size and Power Properties

I now present the conditions under which the modified conditional CI has asymptotic uniform validity.

Assumption 1. *(Known Singularity). A_ℓ, A_u are known $|\mathcal{B}| \times J$ matrices such that*

$$\begin{aligned} \lambda_\ell &= A_\ell \delta_P, \ \lambda_u = A_u \delta_P \\ \hat{\lambda}_\ell &= A_\ell \hat{\delta}_n, \ \hat{\lambda}_u = A_u \hat{\delta}_n \\ \hat{\Sigma}_n &= \begin{bmatrix} A'_\ell & A'_u \end{bmatrix}' \hat{\Omega}_n \begin{bmatrix} A'_\ell & A'_u \end{bmatrix} \end{aligned} \tag{31}$$

for some $(\delta_P, \hat{\delta}_n, \hat{\Omega}_n)$.

Assumption 2. *(Asymptotic Normality). 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*

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

where $\xi_P \sim \mathcal{N}(0, \Omega_P)$.

Assumption 3. *(Full Rank). Let $\mathcal{S}$ denote the set of matrices with eigenvalues bounded below by $\underline{e} > 0$ and above by $\bar{e} \geq \underline{e}$. For all $P \in \mathcal{P}$, $\Omega_P \in \mathcal{S}$.*

Assumption 4. *(Consistent Covariance Estimator). For all $\varepsilon > 0$,*

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

Assumption 5. *(Confidence Set of λ). For all $\eta \in [0, \frac{\alpha}{4})$, the confidence set $\hat{\Lambda}_\eta$ satisfies*

$$\liminf_{n} \inf_{P\in\mathcal{P}} P\left((\lambda_\ell, \lambda_u) \in \hat{\Lambda}_\eta\right) \geq 1 - \eta.$$

Assumptions 1, 2, 3 and 4 imply that $\sqrt{n}\left(\hat{\lambda}_n - \lambda_P\right)$ is asymptotically normal with a consistently estimable variance. The asymptotic variance is allowed to be singular, but the source of the singularity, i.e. A_ℓ and A_u, is known to the researcher. Given this, we only need to verify whether $A_{\ell,b_1} = -aA_{u,b_2}$ for some $a > 0$ to know whether $\rho_{\ell u}(b_1, b_2)$ is at the boundary -1, which simplifies the construction of $\hat{c}^c(\theta, \alpha)$. These assumptions hold for the examples in Section 2.2 with finite $\mathcal{B}$ under mild conditions, and I give detailed illustration based on RR23 in Appendix C. Assumption 5 requires that $\hat{\Lambda}$ is a uniformly valid $1-\eta$ confidence set of $(\lambda_\ell, \lambda_u)$, e.g.

$$\hat{\Lambda} = \left\{(A_\ell\delta,\ A_u\delta) \in \Lambda : \delta \in \hat{\Delta}\right\}, \quad \hat{\Delta} = \left\{\delta : \sqrt{n}|\hat{\delta}_j - \delta_j| \leq \hat{\Omega}_{jj}^{1/2}\hat{c}_\eta, \forall j = 1, ..., J\right\} \tag{32}$$

$$\hat{c}_\eta = Q\left(\max_{j=1,\ldots,J} |Z_j^*|, 1-\eta\right), \quad Z^* \sim \mathcal{N}\left(0, \text{diag}(\hat{\Omega})^{-\frac{1}{2}}\hat{\Omega}\text{diag}(\hat{\Omega})^{-\frac{1}{2}}\right),$$

where $Q(X, 1-\eta)$ is the $1-\eta$ quantile of X.

Theorem 1. *(Uniform Coverage) Suppose Assumptions 1, 2, 3, 4, and 5 hold. Let $\alpha \in (0, 1/2)$, $\alpha^c \in (\frac{\alpha}{2}, \alpha)$, $\eta \in [0, \frac{\alpha-\alpha^c}{2})$. It holds that*

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \sup_{\theta\in\left[\lambda_{\ell,b_\ell}, \lambda_{u,b_u}\right]} P\left(\theta \notin CI^m\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) \leq \alpha.$$

Next, I compare my method to two existing approaches which are also uniformly valid: (i) the simple CI given in (12), and (ii) the adjusted bootstrap CI proposed in YKHS23.

Theorem 2. *(Power Comparison with Simple CI) Suppose Assumptions 1, 2, 3, and 4 hold. $\hat{\Lambda}_\eta$ is defined as in (32). Let $\alpha \in (0, \frac{1}{2})$, $\alpha^c \in (\frac{\alpha}{2}, \alpha)$, $\eta \in [0, \frac{\alpha-\alpha^c}{2})$. If one of the following two conditions hold*

1. *(Symmetric Bounds)* $A_\ell = A_u$, *and* $\mathcal{P}_n$ *satisfies*

$$\limsup_{P \in \mathcal{P}_n} \max_{b_1 \in \mathcal{B}} \min_{b_2 \in \mathcal{B}} \rho_\ell(b_1, b_2) < \rho_1^*(\alpha, \alpha^c), \tag{33}$$

$$\limsup_{P \in \mathcal{P}_n} \rho_\ell(b_\ell, b_u) < \rho_2^*(\alpha, \eta), \tag{34}$$

where $\rho_1^*(\alpha, \alpha^c)$ *and* $\rho_2^*(\alpha, \eta)$ *are defined in Lemma 10 and Lemma 7, respectively.*[9]

2. *(Large Bounds) Let* $\kappa_n = o(\sqrt{n})$ *and* $\kappa_n \to \infty$, *and*

$$\mathcal{P}_n = \left\{ P \in \mathcal{P} : \lambda_{u,b_u} - \lambda_{\ell,b_\ell} \geq \frac{\kappa_n}{\sqrt{n}} \right\}. \tag{35}$$

It holds that

1. *Modified conditional CI is shorter: there is* $\alpha' > \alpha$ *such that*

$$\liminf_n \inf_{P \in \mathcal{P}_n} P\left(CI^m\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right) \subseteq CI^{sim}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha'\right)\right) = 1. \tag{36}$$

2. *Modified conditional CI has higher power: for all* $P_n \in \mathcal{P}_n$, *there is a subsequence* P_{a_n} *and* $\kappa \in (0, +\infty)$ *thus that*

$$\liminf_{a_n \to \infty} P_{a_n}\left(\theta_{a_n} \notin CI^m\left(\hat{\lambda}_{a_n}, \hat{\Sigma}_{a_n}/a_n; \alpha\right)\right) - P_{a_n}\left(\theta_{a_n} \notin CI^{sim}\left(\hat{\lambda}_{a_n}, \hat{\Sigma}_{a_n}/a_n; \alpha\right)\right) > 0 \tag{37}$$

for $\theta_n = \theta_\ell - \frac{\kappa}{\sqrt{n}}$. The same applies to the upper bound.

The first part of Theorem 2 considers the case where the upper and lower bounds are symmetric, as in Kolesár and Rothe (2018), Masten and Poirier (2021), and RR23. If the correlation coefficients among $\hat{\lambda}_\ell$ are not too large, the modified conditional CI is strictly shorter than the simple CI. The upper bounds $\rho_1^*(\alpha, \alpha^c)$ and $\rho_2^*(\alpha, \eta)$ can be easily solved for numerically, and I list the value for a few combinations:

$$\rho_1^*(0.05, 0.04) = 0.84, \qquad \rho_1^*(0.10, 0.08) = 0.83,$$
$$\rho_2^*(0.05, 0.001) \approx 1, \qquad \rho_2^*(0.10, 0.001) \approx 1.$$

[9]Here $\rho_\ell(b_1, b_2) = \rho_u(b_1, b_2) = \rho_{\ell u}(b_1, b_2)$, so I only impose restrictions on ρ_ℓ.

The values are large and thus the restriction (33) is not binding in most applications.

The second part of Theorem 2 compares the modified conditional CI with the simple CI in a different set of DGPs. It shows that if the identified set is relatively large compared to the standard deviation of the estimators, which is $O(\frac{1}{\sqrt{n}})$, the modified conditional CI is shorter than the simple CI with probability approaching one. The intuition follows from the discussion around (25).

Theorem 3. *(Power Comparison with YKHS23) Let* $CI^{YKHS}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)$ *be the adjusted bootstrap procedure proposed in Hasegawa et al. (2019) Algorithm 1 equation (15) with tuning parameter* $m = \frac{n}{\kappa_n}$*, where* $\kappa_n \to \infty$ *and* $\kappa_n = o(n)$*. Let* $a > 0$*,* $\kappa'_n = o(\sqrt{\kappa_n})$*,* $\kappa'_n \to \infty$*. Define local alternatives*

$$\theta_n = \min_{b \in \mathcal{B}} \lambda_{\ell,b} - \frac{\kappa'_n}{\sqrt{n}} a, \quad \text{or } \theta_n = \max_{b \in \mathcal{B}} \lambda_{u,b} + \frac{\kappa'_n}{\sqrt{n}} a.$$

Then

$$\liminf_{n \to \infty} \inf_{P \in \mathcal{P}} P\left(\theta_n \notin CI^m\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) = 1,$$
$$\limsup_{n \to \infty} \sup_{P \in \mathcal{P}} P\left(\theta_n \notin CI^{YKHS}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) \leq \alpha.$$

YKHS23 method relies on a random draw of a subsample with size $m = \frac{n}{\kappa_n}$, and thus the convergence rate of the CI to the identified set is $\sqrt{m}$, slower than $\sqrt{n}$. Theorem 3 follows from the convergence rate of CI^{YKHS} and CI^{m}. The sequence of θ_n is rejected by the modified conditional CI with probability approaching one following from Lemma 2, while it is rejected by CI^{YKHS} with probability bounded above by α. Hence, CI^{m} has large power improvement upon CI^{YKHS}.

4 Simulation

In this section, I study the size and power properties of the proposed procedures and compare them to several alternatives. I conduct simulations in the context of Example 1, i.e. relaxation of the parallel trends assumption as in RR23. Besides the modified conditional CI proposed in Section 3, I consider two existing procedures for union bounds: (i) the adjusted bootstrap in YKHS23, (ii) the simple CI in (12), and (iii) the inference procedure in RR23.[10] All three methods are uniformly valid. All tuning parameters are set at the values in the papers in which they are proposed.

Each sample $\{Y_i\}_{i=1}^n$ and estimator is generated by

$$Y_i \sim \mathcal{N}(\gamma, n\Sigma), \quad \hat{\gamma} = \frac{1}{n}\sum Y_i \sim \mathcal{N}(\gamma, \Sigma).$$

The inference is conducted using the pair $(\hat{\gamma}, \Sigma)$. The covariance matrix Σ is calibrated from the empirical results reported in (i) Dustmann et al. (2022) Figure 7(c), (ii) Benzarti and Carloni (2019) Figure 2(E), (iii) Lovenheim and Willén (2019) Figure 3(A), and (iv) Christensen, Keiser, and Lade (2023) Figure 5(b). Specifically, Σ is set to be the estimated covariance matrix for $t = -\underline{T}, ..., -1, 1$, where $\underline{T}$ is reported under Figure 3. For each Σ, I considered three true values for γ: (i) the parallel trends assumption holds, i.e. $\gamma^{pre} = 0_{\underline{T}}$; (ii) there is a small violation of the parallel trends, where γ^{pre} is calibrated from the same source as Σ; (iii) there is a large violation, where $\gamma^{pre} = (10\sigma_M, 0_{\underline{T}-1})$, $\sigma_M = \max_{b\in\mathcal{B}}\{\sigma_{\ell,b}\}$. Without loss of generality, I normalize $\gamma^{post} = 0$. In sum, I consider $4 \times 3 = 12$ empirically motivated DGPs. Note that the simulation results of the modified conditional CI, the RR23 CI, and the simple CI are invariant to n, while YKHS depends on n because of a subsampling step. I set $n = 5000$ and $S = 1000$ sample draws.

In Figure 3, I plot the rejection rate near the upper bound. The lower bound is similar and thus omitted. The horizontal axis is the value of θ, while the vertical axis is the rate that θ is not included in the CI. The asterisks represent the identified set, and the nominal rejection rate is 10%. The modified conditional CI is the red curve and it has proper size control in all simulation designs. The simple CI is the black dotted curve and it has significantly lower power than modified conditional CI in all designs.

RR23 CI is plotted in blue dashed curves. The performance of RR23 varies with the DGPs, and the power is usually between the modified conditional CI and simple CI, see e.g. Figure 3a, 3b, and 3d. In some DGPs, RR23 may perform

[10] YKHS23 propose two CIs in Algorithm 1 equation (15): one with the tuning parameter $m/N \to 0, m \to \infty$ and the other with $m = N$. The second one is not uniformly valid, and thus I only consider the first one with $m = N/\log(\log(N))$ as suggested in their Section S1.4. For RR23, I use their hybrid conditional CI with tuning parameter $\eta = \frac{\alpha}{10}$, which is the default choice in their code. For my CI, $\alpha^c = 0.8\alpha, \eta = 0.001$.

worse than the simple CI, e.g. in Figure 3g. When there is only one large violation, for example in Figure 3i-3l, the minimum and maximum of the union bound are well-separated from other bounds, and RR23 is near optimal by their Corollary 3.1. In this case, the modified conditional CI has a slightly smaller rejection rate and is close to optimal.

YKHS is plotted in green circled curves. YKHS has slightly higher power than the modified conditional CI for points very close to the identified set but often suffers from large power loss for points farther away, see e.g. all designs except Figure 3c and 3g. This is consistent with the slower convergence rate of the YKHS CI to the identified set.

In Table 1, I report the median CIs.[11] I compare the differences between the length of median CIs and the length of the union bound estimates, as a measure of efficiency.[12] The difference of the modified conditional CI is the shortest, or slightly larger than the shortest, in all DGPs. It significantly reduces the value of RR23 (resp. YKHS, simple CI) by a proportion up to 43% (resp. 32%, 37%), see third panel with small violation (resp. fourth panel with parallel trends, fourth panel with parallel trends).

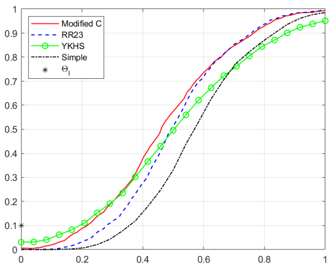
The power of modified conditional CI is not sensitive to the tuning parameter α^{c}. When α^{c} varies from 0.06 to 0.09, the largest change in the rejection rate across all DGPs is 0.12, and the average change in the length of the median CI, net of the length of the union bound estimate, is 13%. The average computing times on a standard PC without parallelization for $\underline{T} = 3, 4, 9, 15$ are 80s, 80s, 100s, and 260s, respectively.

5 Empirical Illustration

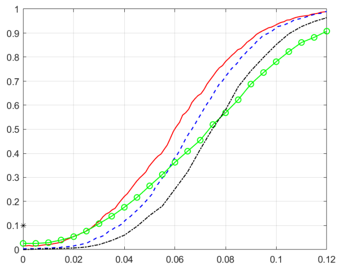
In this section, I apply the modified conditional CI to the sensitivity analysis in Dustmann et al. (2022). The authors study the labor market effects of the minimum wage implemented by the German government in January 2015, impacting approximately 15% of the workforce. The minimum wage policy remains a subject of considerable controversy within the labor market, as it simultaneously addresses wage inequality while potentially leading to disemployment. One main conclusion

[11] A median CI is the median lower bound of the $1-\alpha$ CI to the median upper bound, and the median is taken over S samples.

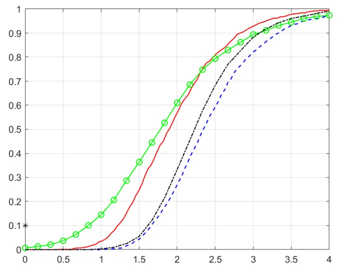
[12] The consistent union bound estimate is $\left[\min_{b\in\mathcal{B}} \hat{\lambda}_{\ell,b},\ \max_{b\in\mathcal{B}} \hat{\lambda}_{u,b}\right]$.



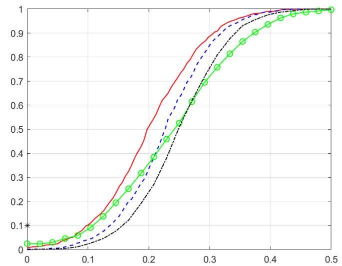
(a) Parallel Trends with Σ_1



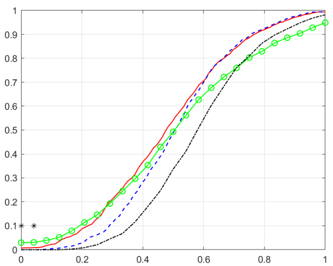
(b) Parallel Trends with Σ_2



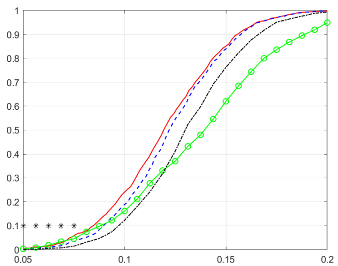
(c) Parallel Trends with Σ_3



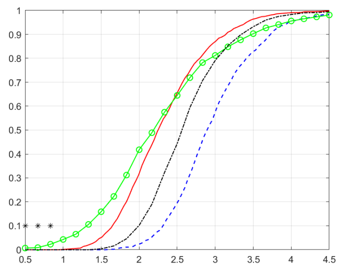
(d) Parallel Trends with Σ_4



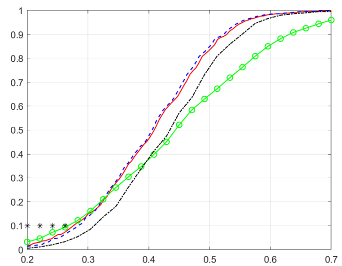
(e) Small Violation with (γ_1, Σ_1)



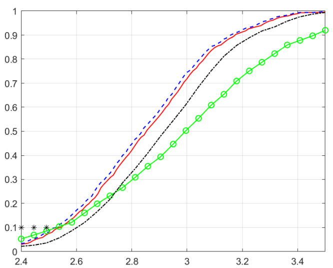
(f) Small Violation with (γ_2, Σ_2)



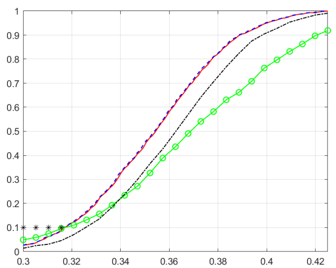
(g) Small Violation with (γ_3, Σ_3)



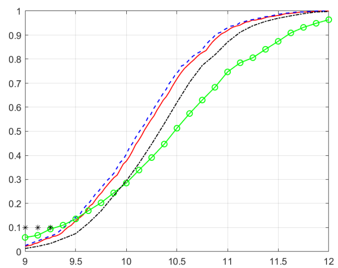
(h) Small Violation with (γ_4, Σ_4)



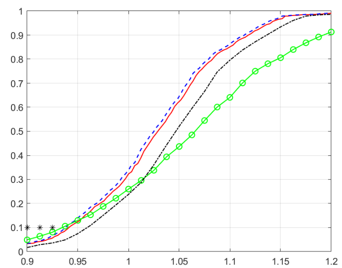
(i) Large Violation with Σ_1



(j) Large Violation with Σ_2



(k) Large Violation with Σ_3



(l) Large Violation with Σ_4

Figure 3: (μ_1, Σ_1) is calibrated from Dustmann et al. (2022) with $\underline{T} = 3$, (μ_2, Σ_2) from Benzarti and Carloni (2019) with $\underline{T} = 4$, (μ_3, Σ_3) from Lovenheim and Willén (2019) with $\underline{T} = 9$, and (μ_4, Σ_4) from Christensen et al. (2023) with $\underline{T} = 15$.

Table 1: Simulation Results - Median CI

		Point	Modi. Con.	RR23	YKHS	Simple
		Dustmann et al. (2022) $\underline{T} = 3$				
Parallel	CI	[-0.188, 0.188]	[-0.470, 0.456]	[-0.505, 0.492]	[-0.486, 0.502]	[-0.578,0.565]
$[0, 0]$	Diff.		0.550	0.621	0.612	0.768
Small Vio.	CI	[-0.196, 0.195]	[-0.467, 0.475]	[-0.504, 0.505]	[-0.492, 0.504]	[-0.576,0.581]
$[-0.080, 0.080]$	Diff.		0.551	0.619	0.605	0.766
Large Vio.	CI	[-2.508, 2.503]	[-2.857, 2.852]	[-2.840, 2.836]	[-2.955, 2.995]	[-2.920, 2.916]
$[-0.316, 0.316]$	Diff.		0.698	0.666	0.939	0.825
		Benzarti and Carloni (2019) $\underline{T} = 4$				
Parallel	CI	[-0.028, 0.029]	[-0.058, 0.059]	[-0.067, 0.067]	[-0.070, 0.074]	[-0.075, 0.075]
$[0, 0]$	Diff.		0.061	0.077	0.088	0.093
Small Vio.	CI	[-0.085, 0.085]	[-0.120, 0.119]	[-0.121, 0.122]	[-0.130, 0.140]	[-0.130, 0.130]
$[-0.080, 0.080]$	Diff.		0.069	0.073	0.101	0.090
Large Vio.	CI	[-0.316, 0.317]	[-0.359, 0.354]	[-0.358, 0.354]	[-0.374, 0.368]	[-0.368, 0.363]
$[-0.316, 0.316]$	Diff.		0.080	0.079	0.109	0.098
		Lovenheim and Willén (2019) $\underline{T} = 9$				
Parallel	CI	[-0.909, 0.884]	[-1.886, 1.867]	[-2.341, 2.343]	[-1.709, 1.800]	[-2.235, 2.236]
$\theta \in [0, 0]$	Diff.		1.960	2.891	1.715	2.678
Small Vio.	CI	[-1.360, 1.354]	[-2.261, 2.225]	[-2.927, 2.893]	[-2.193, 2.194]	[-2.590, 2.567]
$[-0.993, 0.993]$	Diff.		1.772	3.106	1.673	2.442
Large Vio.	CI	[-9.366, 9.332]	[-10.034, 10.174]	[-9.999, 10.128]	[-10.201, 10.483]	[-10.153, 10.323]
$[-9.350, 9.350]$	Diff.		1.509	1.428	1.985	1.778
		Christensen et al. (2023) $\underline{T} = 15$				
Parallel	CI	[-0.108, 0.108]	[-0.197, 0.195]	[-0.225, 0.227]	[-0.233, 0.242]	[-0.247, 0.249]
$[0, 0]$	Diff.		0.176	0.236	0.259	0.280
Small Vio.	CI	[-0.279, 0.281]	[-0.391, 0.409]	[-0.391, 0.405]	[-0.431, 0.445]	[-0.416, 0.434]
$[-0.276, 0.276]$	Diff.		0.240	0.236	0.316	0.290
Large Vio.	CI	[-0.932, 0.933]	[-1.040, 1.029]	[-1.036, 1.025]	[-1.084, 1.064]	[-1.062, 1.047]
$[-0.934, 0.934]$	Diff.		0.204	0.196	0.283	0.243

of Dustmann et al. (2022) is that the minimum wage increase resulted in higher wages without causing a decline in employment levels.

To study the employment effect, the authors estimate the DiD design

$$\log(\text{emp}_{rt}) = \sum_{\tau=2011,\tau\neq 2014}^{2016} \gamma_\tau \overline{GAP}_r \mathbf{1}\left[\tau = t\right] + \alpha_r + \xi_t + \varepsilon_{rt} \tag{38}$$

where $\log(\text{emp}_{rt})$ is the log employment in district r, time t; $\overline{GAP}_r$ is a measure of the exposure to the minimum wage; α_r and ξ_t are district and year fixed effects. The parameter vector γ is the event study coefficients with γ_{2014} normalized to zero. Figure 4 left panel shows the estimated coefficients $\{\hat{\gamma}_\tau\}$ from specification (38). Under the parallel trends assumption, the high and barely exposed districts evolved at the same rate in the absence of the minimum wage policy. In this context, the coefficients γ_{2015} and γ_{2016} in the post-policy years serve as measures for the employment effects of the minimum wage policy. However, Figure 4 indicates that the coefficients $\gamma_{2011}, \gamma_{2012}$ and γ_{2013} in the pre-policy years are not statistically or economically indistinguishable from zero. Hence, it is evident that the parallel trends assumption does not hold. Consequently, the authors conduct sensitivity analysis using RR23, as detailed in their Appendix A.14.

In particular, the authors conduct the sensitivity analysis using the second differences relative magnitudes (SDRM) relaxation. This approach assumes that

$$\left|(\xi_{2015} - \gamma_{2014}) - (\gamma_{2014} - \gamma_{2013})\right| \leq M \max_{s=2014,2013} \left|(\gamma_s - \gamma_{s-1}) - (\gamma_{s-1} - \gamma_{s-2})\right|,$$

where ξ_{2015} represents the potential differential trend without the minimum wage policy. Essentially, without the minimum wage policy, the slope change at $t = 2015$ is bounded above by a factor of M times the previous slope changes. M measures the level of relaxation. This aligns with the approximately linear pretrend observed in Figure 4. The employment effect of interest is quantified as $\gamma_{2015} - \xi_{2015}$.[13] That is, with one unit increase in GAP and other covariates fixed, the employment rate will increase by $100(\gamma_{2015} - \xi_{2015})\%$ in expectation.

In Figure 4 right panel, I report the 95% CI for different values of M constructed based on three different methods: the modified conditional CI proposed in Section 3, the hybrid CI in RR23, and the simple CI in (12).[14] We can clearly

[13] The identified set is given in Appendix C.

[14] The estimated coefficient and covariance are available but the data for regression is confi-

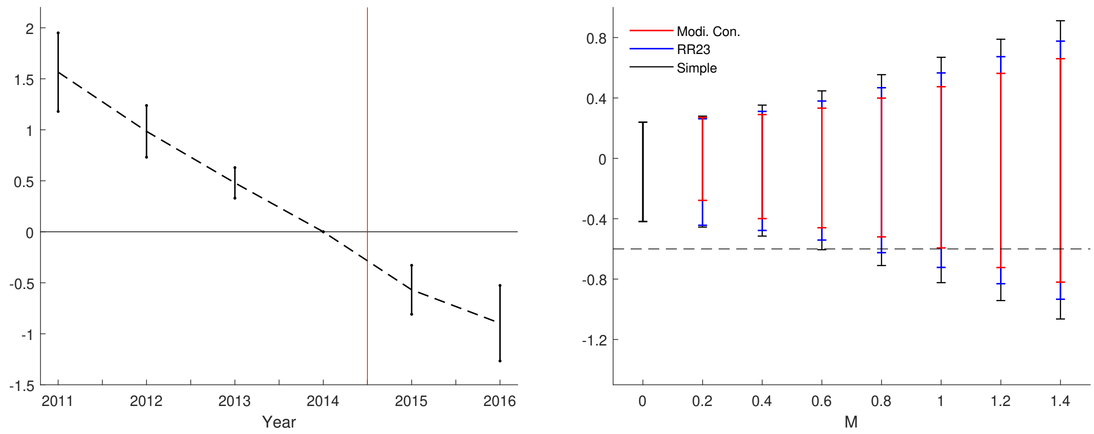


Figure 4: Left: Estimated Employment Effect. Right: Sensitivity Analysis under SDRM.

see that the modified conditional CI is the shortest and the simple CI is the widest for all M, and the improvement of the modified conditional CI upon the simple CI doubles the improvement of RR23 upon the simple CI.

The authors compare the minimum wage induced disemployment effects and wage effects. To do so, they estimate the wage effect using the same DiD design as (38) with regressor $\log(\text{wage}_{rt})$. After adjusting the linear pretrend, the point estimate of the wage effect at $t = 2015$ is 0.6, represented by a dashed line (with an inverse sign) in Figure 4 right panel. The authors are interested in whether the employment effect is robustly higher than -0.6, leading to an employment elasticity with respect to own wage less than 1 in absolute value. When using the natural benchmark $M = 1$, only the modified conditional CI is above the negative wage effect. It is also informative to report the "breakdown" relaxation at which the wage effect is no longer larger than the (negative) employment effect. In this case, the breakdown M for the hybrid CI is around 0.75, while the one for the simple CI is around 0.6. Remarkably, the breakdown relaxation M of my method is 33% to 66% larger than the other two. The average computing time of my method is 50s per confidence interval.

dential, thus I can not implement the YKHS23 bootstrap procedure.

6 Conclusion

In this paper, I propose inference procedures for a target object whose identified set is a union of bounds. When the union is taken over a finite set, I introduce a novel modified conditional CI based on a modified conditional critical value, which significantly improves upon existing procedures over a large set of DGPs. There are a few directions for future work. First, the important tuning parameter α^{c} trades off the rejection rate between less and more favorable DGPs, and the suggested rule of thumb value is $\frac{4}{5}\alpha$. It would be useful to consider a choice of α^{c} that optimizes some objective function, for example, weighted average power. In addition, the idea of modified conditional inference could potentially apply to other non-standard inference problems like directionally differentiable functions. This idea does not impose shape restrictions, e.g. convexity, on the null space. Lastly, my inference procedures assume a correct specification that the union bound is non-empty. If the model is misspecified, the CI can be an empty set or spuriously short. It would be interesting to consider misspecification robust inference for general union bounds, in the spirit of Stoye (2020) and Andrews and Kwon (2023).

A Proofs for Theorems and Propositions

For $P \in \mathcal{P}$, let δ_P denote the true value of δ, $\lambda_{P,\ell} = A_\ell \delta_P$, $\lambda_{P,u} = A_u \delta_P$,

$$\theta_{P,\ell} = \min_{b \in \mathcal{B}} \lambda_{P,\ell,b},\ \theta_{P,u} = \max_{b \in \mathcal{B}} \lambda_{P,u,b},\ \theta_{P,m} = (\theta_{P,\ell} + \theta_{P,u})/2.$$

Let

$$Z_\delta \sim \mathcal{N}\left(0, \Omega_0\right), \quad Z_{\ell,b} = \frac{A_{\ell,b} Z_\delta}{\sigma_{0,\ell,b}}, \quad Z_{u,b} = -\frac{A_{u,b} Z_\delta}{\sigma_{0,u,b}}$$

denote the limiting distribution of

$$\sqrt{n}\left(\hat{\delta}_n - \delta_{P_n}\right),\ \frac{\sqrt{n}\left(\hat{\lambda}_{\ell,b} - \lambda_{P_n,\ell,b}\right)}{\hat{\sigma}_{\ell,b}},\ \frac{\sqrt{n}\left(\lambda_{P_n,u,b} - \hat{\lambda}_{u,b}\right)}{\hat{\sigma}_{u,b}}$$

with Ω_0 and P_n specified in Lemma 3 and $\sigma_{0,\ell,b} = \sqrt{A_{\ell,b}\Omega_0 A'_{\ell,b}}$, $\sigma_{0,u,b} = \sqrt{A_{u,b}\Omega_0 A'_{u,b}}$. For $k = \ell, m, u$, let

$$T_k = \max\left\{\min_{b\in\mathcal{B}} Z_{\ell,b} + \lambda_{k\ell,b},\ \min_{b\in\mathcal{B}} Z_{u,b} + \lambda_{ku,b}\right\} \tag{39}$$

be the asymptotic analog of $\hat{T}(\theta_{P_n,k})$, where $(\lambda_{k\ell}, \lambda_{ku})$ are specified in Lemma 3 (78).

Let $\mathcal{B}_\ell$ be a subset of $\mathcal{B}$ such that $A_{\ell,b_1} \neq A_{\ell,b_2}$ for all $b_1 \neq b_2$, $b_1, b_2 \in \mathcal{B}_\ell$. If there is $A_{\ell,b_1} = A_{\ell,b_2}$ for $b_1, b_2 \in \mathcal{B}$, keep only $\min\{b_1, b_2\}$ in $\mathcal{B}_\ell$. Construct $\mathcal{B}_u$ in the same way. Let $b_{k\ell}$, b_{ku} be the asymptotic analog of $\hat{b}_\ell(\theta_k)$ and $\hat{b}_u(\theta_k)$, with support $\mathcal{B}_{k\ell}$, $\mathcal{B}_{ku}$:

$$\begin{aligned} b_{k\ell} &= \min\left\{\arg\min_{b\in\mathcal{B}_\ell} Z_{\ell,b} + \lambda_{k\ell,b}\right\}, & b_{ku} &= \min\left\{\arg\min_{b\in\mathcal{B}_u} Z_{u,b} + \lambda_{ku,b}\right\}, \\ \mathcal{B}_{k\ell} &= \{b \in \mathcal{B}_\ell : \lambda_{k\ell,b} < \infty\}, & \mathcal{B}_{ku} &= \{b \in \mathcal{B}_u : \lambda_{ku,b} < \infty\}. \end{aligned} \tag{40}$$

Define the asymptotic analog of $(t_{\ell,1}, t_{\ell,2})$ in Lemma 1 evaluated at $\theta_{P_n,k}$ as

$$t_{k\ell,1}(b) = \begin{cases} \min_{\tilde{b}\in\mathcal{B}} \frac{Z_{u,\tilde{b}} + \rho_{\ell u}(b,\tilde{b}) Z_{\ell,b} + t^\dagger_{k\ell,1}(b,\tilde{b})}{1+\rho_{\ell u}(b,\tilde{b})}, & \text{if } \min_{\tilde{b}\in\mathcal{B}} \rho_{\ell u}(b,\tilde{b}) > -1 \\ -\infty & \text{elsewhere} \end{cases} \tag{41}$$

$$t^\dagger_{k\ell,1}(b,\tilde{b}) = \lambda_{ku,\tilde{b}} + \rho_{\ell u}(b,\tilde{b})\lambda_{k\ell,b}.$$

Similarly, we can define $t_{k\ell,2}, t_{u,1}, t_{u,2}$. If $\left|\lambda_{ku,\tilde{b}}\right| = \infty$ and $|\lambda_{k\ell,b}| = \infty$, $t^\dagger_{k\ell,1}(b,\tilde{b})$ may not be well defined. However, as we will see later, this case is irrelevant for the proof. Let

$$c^{\mathrm{c}}_k = \begin{cases} \Phi^{-1}\Big(\alpha^{\mathrm{c}}\Phi\big(t_{k\ell,1}(b_{k\ell})\big) + (1-\alpha^{\mathrm{c}})\Phi\big(t_{k\ell,2}(b_{k\ell})\big)\Big), & \text{if } Z_{\ell,b_{k\ell}} + \lambda_{k\ell,b_{k\ell}} \geq Z_{u,b_{ku}} + \lambda_{ku,b_{ku}} \\ \Phi^{-1}\Big(\alpha^{\mathrm{c}}\Phi\big(t_{ku,1}(b_{ku})\big) + (1-\alpha^{\mathrm{c}})\Phi\big(t_{ku,2}(b_{ku})\big)\Big), & \text{if } Z_{\ell,b_{k\ell}} + \lambda_{k\ell,b_{k\ell}} < Z_{u,b_{ku}} + \lambda_{ku,b_{ku}} \end{cases} \tag{42}$$

be the asymptotic analog of $\hat{c}^{\mathrm{c}}(\theta_k, \alpha^{\mathrm{c}})$. Let

$$\begin{aligned} p(c) = \max\{&P\left(T_\ell > c^{\mathrm{m}}_\ell(c) \text{ or } \{T_m > c^{\mathrm{m}}_m(c) \text{ and } T_u > c^{\mathrm{m}}_u(c)\}\right), \\ &P\left(T_u > c^{\mathrm{m}}_u(c) \text{ or } \{T_m > c^{\mathrm{m}}_m(c) \text{ and } T_\ell > c^{\mathrm{m}}_\ell(c)\}\right)\}, \end{aligned} \tag{43}$$

where $c_k^{\mathrm{m}}(c) = \max\{c_k^{\mathrm{c}}, c\}$ and T_k, c_k^{c} are defined in (39), (42). Lastly, let

$$c^t = \inf_c \{c \geq 0 : p(c) \leq \alpha - \eta\}, \tag{44}$$

be the asymptotic analog of $\hat{c}^{\mathrm{t}}$ defined (30).

I use Φ for the CDF of $\mathcal{N}(0,1)$ and $\Phi_2(x_1, x_2; \rho)$ for the CDF of $\mathcal{N}(0, [1, \rho; \rho, 1])$.

Proof of Lemma 1.

Proof. Let b_1 satisfy $\lambda_{\ell,b_1} \leq \theta$, and I show (22). Similar proof applies to the upper bound. The proof mainly uses Theorem 5.2 and Lemma A.1 in Lee, Sun, Sun, and Taylor (2016), and below I follow their notation. For $s \in \mathcal{B}$, let

$$A_s = \begin{pmatrix} \mathbf{1}_{|\mathcal{B}|\times 1} \\ -1 \end{pmatrix}, \; b_s = \begin{pmatrix} \mathcal{Z}_\ell \\ -\mathcal{Z}_{u,s} \end{pmatrix}.$$

It is easy to see that

$$\left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\right\} = \bigcup_{s\in\mathcal{B}} \{A_s \mathcal{Z}_{\ell,b_1} \leq b_s\}. \tag{45}$$

To simplify $A_s \mathcal{Z}_{\ell,b_1} \leq b_s$, note that for all $b \in \mathcal{B}$,

$$\mathcal{Z}_{\ell,b_1} \leq \mathcal{Z}_{\ell,b} \Leftrightarrow \begin{cases} \mathcal{Z}_{\ell,b_1} \leq (1-\rho_\ell(b_1,b))^{-1}(\mathcal{Z}_{\ell,b} - \rho_\ell(b_1,b)\mathcal{Z}_{\ell,b_1}) & \text{if } \rho_\ell(b_1,b) < 1 \\ 0 \leq \mathcal{Z}_{\ell,b} - \mathcal{Z}_{\ell,b_1} & \text{if } \rho_\ell(b_1,b) = 1 \end{cases}$$

$$\mathcal{Z}_{\ell,b_1} \geq \mathcal{Z}_{u,s} \Leftrightarrow \begin{cases} \mathcal{Z}_{\ell,b_1} \geq (1+\rho_{\ell u}(b_1,s))^{-1}(\mathcal{Z}_{u,s} + \rho_{\ell u}(b_1,s)\mathcal{Z}_{\ell,b_1}) & \text{if } \rho_{\ell u}(b_1,s) > -1 \\ 0 \geq \mathcal{Z}_{u,s} - \mathcal{Z}_{\ell,b_1} & \text{if } \rho_{\ell u}(b_1,s) = -1. \end{cases}$$

Therefore,

$$\{A_s \mathcal{Z}_{\ell,b_1} \leq b_s\} = \left\{\mathcal{V}_s^- \leq \mathcal{Z}_{\ell,b_1} \leq \mathcal{V}^+, \mathcal{V}^0 \geq 0\right\}, \text{ where} \tag{46}$$

$$\mathcal{V}_s^- = \begin{cases} (1+\rho_{\ell u}(b_1,s))^{-1}(\mathcal{Z}_{u,s} + \rho_{\ell u}(b_1,s)\mathcal{Z}_{\ell,b_1}) & \text{if } \rho_{\ell u}(b_1,s) > -1 \\ -\infty & \text{if } \rho_{\ell u}(b_1,s) = -1, \end{cases}$$

$$\mathcal{V}^+ = \begin{cases} \min\limits_{b\in\mathcal{B}:\rho_\ell(b_1,b)<1} (1-\rho_\ell(b_1,b))^{-1}(\mathcal{Z}_{\ell,b} - \rho_\ell(b_1,b)\mathcal{Z}_{\ell,b_1}) & \text{if } \{b\in\mathcal{B}, \rho_\ell(b_1,b)<1\} \neq \emptyset \\ +\infty & \text{if } \{b\in\mathcal{B}, \rho_\ell(b_1,b)<1\} = \emptyset, \end{cases}$$

$$\mathcal{V}_s^0 = \begin{cases} \min_{b\in\mathcal{B}:\rho_\ell(b_1,b)=1} \mathcal{Z}_{\ell,b} - \mathcal{Z}_{\ell,b_1} & \text{if } \rho_{\ell u}(b_1,s) > -1,\ \max_{b\in\mathcal{B}}\rho_\ell(b_1,b)=1 \\ \min\left\{\min_{b\in\mathcal{B}:\rho_\ell(b_1,b)=1} \mathcal{Z}_{\ell,b} - \mathcal{Z}_{\ell,b_1}, \mathcal{Z}_{\ell,b_1} - \mathcal{Z}_{u,s}\right\} & \text{if } \rho_{\ell u}(b_1,s) = -1,\ \max_{b\in\mathcal{B}}\rho_\ell(b_1,b)=1 \\ 1 & \text{elsewhere.} \end{cases}$$

$\mathcal{Z}_{\ell,b_1} \perp \left\{\mathcal{V}^+, \{\mathcal{V}_s^-, \mathcal{V}_s^0\}_{s\in\mathcal{B}}\right\}$ by construction. With $t_{\ell,1}(\theta,b_1)$ and $t_{\ell,2}(\theta,b_1)$ in Lemma 1,

$$[t_{\ell,1}(\theta,b_1), t_{\ell,2}(\theta,b_1)] = \bigcup_{s\in\mathcal{B}} \left[\mathcal{V}_s^-, \mathcal{V}^+\right].$$

Let $F_\mu(x;t_1,t_2)$ denote CDF of a $\mathcal{N}(\mu,1)$ random variable truncated to $[t_1,t_2]$, i.e.

$$F_\mu(x;t_1,t_2) = \frac{\Phi(x-\mu) - \Phi(t_1-\mu)}{\Phi(t_2-\mu) - \Phi(t_1-\mu)}. \tag{47}$$

where $\mu = E\left[\mathcal{Z}_{\ell,b_1}\right] = \frac{\lambda_{\ell,b_1}-\theta}{\sigma_{\ell,b_1}} \leq 0$. Then by Theorem 5.3 in Lee et al. (2016),

$$F_\mu\left(\mathcal{Z}_{\ell,b_\ell}; t_{\ell,1}(\theta,b_1), t_{\ell,2}(\theta,b_1)\right) \Bigg| \bigcup_{s\in\mathcal{B}} \{A_s \mathcal{Z}_{\ell,b_1} \leq b_s\} \sim \text{Unif}(0,1), \tag{48}$$

and by Lemma A.1 in Lee et al. (2016), for all $z \in \mathbb{R}$,

$$F_0\left(z; t_{\ell,1}(\theta,b_1), t_{\ell,2}(\theta,b_1)\right) \leq F_\mu\left(z; t_{\ell,1}(\theta,b_1), t_{\ell,2}(\theta,b_1)\right). \tag{49}$$

Therefore, we have

$$\begin{aligned} &\frac{\Phi\left(\hat{T}(\theta)\right) - \Phi\left(t_{\ell,1}(\theta,b_1)\right)}{\Phi\left(t_{\ell,2}(\theta,b_1)\right) - \Phi\left(t_{\ell,1}(\theta,b_1)\right)} \Bigg| \left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\right\} \\ \sim &F_0\left(\mathcal{Z}_{\ell,b_1}; t_{\ell,1}(\theta,b_1), t_{\ell,2}(\theta,b_1)\right) \Big| \left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\right\} \\ \overset{\text{FOSD}}{\preceq} &F_\mu\left(\mathcal{Z}_{\ell,b_1}; t_{\ell,1}(\theta,b_1), t_{\ell,2}(\theta,b_1)\right) \Big| \left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\right\} \sim \text{Unif}(0,1). \end{aligned}$$

□

Proof of Proposition 1.

Proof. To simplify notation, let

$$\mathcal{B}_{\ell 0} = \left\{ b \in \mathcal{B}_\ell : \lambda_{\ell,b} \leq \theta, P\left(b = \hat{b}_\ell\right) > 0 \right\},$$
$$\mathcal{B}_{u0} = \left\{ b \in \mathcal{B}_u : \lambda_{u,b} \geq \theta, P\left(b = \hat{b}_u\right) > 0 \right\}.$$

Let $b_1 \in \mathcal{B}_{\ell 0}$. F_μ is defined in (47). By Lemma 1, it holds that

$$\begin{aligned} & P\left(\hat{T}(\theta) > \hat{c}^{\mathrm{c}}(\theta, \alpha^{\mathrm{c}}) \middle| \hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\right) \\ =& P\left(F_0\left(\hat{T}(\theta); t_{\ell,1}(\theta,b_1), t_{\ell,2}(\theta,b_1)\right) > F_0\left(\hat{c}^{\mathrm{c}}(\theta,\alpha^{\mathrm{c}}); t_{\ell,1}(\theta,b_1), t_{\ell,2}(\theta,b_1)\right) \middle| \hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\right) \\ \leq& P\left(F_\mu\left(\hat{T}(\theta); t_{\ell,1}(\theta,b_1), t_{\ell,2}(\theta,b_1)\right) > 1-\alpha^{\mathrm{c}} \middle| \hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\right) \\ =& P\left(\mathrm{Unif}(0,1) > 1-\alpha^{\mathrm{c}}\right) = \alpha^{\mathrm{c}}, \end{aligned} \tag{50}$$

where the second line follows from $F_0(x; t_1, t_2)$ strictly increasing in x, the inequality follows from (49) and by construction

$$F_0\left(\hat{c}^{\mathrm{c}}(\theta,\alpha^{\mathrm{c}}); t_{\ell,1}(\theta,b_1), t_{\ell,2}(\theta,b_1)\right) = 1-\alpha^{\mathrm{c}},$$

and the last line follows from (48). Let $b_2 \in \mathcal{B}_{u0}$. Similar argument gives

$$P\left(\hat{T}(\theta) > \hat{c}^{\mathrm{c}}(\theta, \alpha^{\mathrm{c}}) \middle| \hat{T}(\theta) = \mathcal{Z}_{u,b_2}\right) \leq \alpha^{\mathrm{c}}. \tag{51}$$

Therefore, we have

$$\begin{aligned} & P\left(\hat{T}(\theta) > \hat{c}^{\mathrm{c}}(\theta, \alpha^{\mathrm{c}}) \middle| E_\ell \cup E_u\right) \\ =& \sum_{b_1 \in \mathcal{B}_{\ell 0}} P\left(\hat{T}(\theta) > \hat{c}^{\mathrm{c}}(\theta, \alpha^{\mathrm{c}}) \middle| \hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\right) P\left(\hat{T}(\theta) = \mathcal{Z}_{\ell,b_1} \middle| E_\ell \cup E_u\right) \\ &+ \sum_{b_2 \in \mathcal{B}_{u0}} P\left(\hat{T}(\theta) > \hat{c}^{\mathrm{c}}(\theta, \alpha^{\mathrm{c}}) \middle| \hat{T}(\theta) = \mathcal{Z}_{u,b_2}\right) P\left(\hat{T}(\theta) = \mathcal{Z}_{u,b_2} \middle| E_\ell \cup E_u\right) \\ \leq& \alpha^{\mathrm{c}} \sum_{b_1 \in \mathcal{B}_{\ell 0}} P\left(\hat{T}(\theta) = \mathcal{Z}_{\ell,b_1} \middle| E_\ell \cup E_u\right) + \alpha^{\mathrm{c}} \sum_{b_2 \in \mathcal{B}_{u0}} P\left(\hat{T}(\theta) = \mathcal{Z}_{u,b_2} \middle| E_\ell \cup E_u\right) = \alpha^{\mathrm{c}} \end{aligned}$$

where the first equality follows from $\left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\right\}_{b_1 \in \mathcal{B}_1}, \left\{\hat{T}(\theta) = \mathcal{Z}_{u,b_2}\right\}_{b_2 \in \mathcal{B}_2}$ is a partition of $E_\ell \cup E_u$ under (21), and the inequality follows from (50) and (51). □

Proof of Theorem 1.

Proof. By Lemma 3, we only need to show (75) for sequence P_n satisfying Lemma 3 conditions 1 and 2.

Step 1. I show that for all $c \in \mathbb{R}_+$,

$$\bar{p}\left(c, \lambda_{P_n}, \hat{\Sigma}_n/n\right) \xrightarrow{p} p(c) \tag{52}$$

where $\bar{p}\left(c, \lambda_{P_n}, \hat{\Sigma}_n/n\right)$ is defined in (29) and $p(c)$ is defined in (43). Note that by Assumption 4 and (77), $\hat{\Omega}_n \xrightarrow{p} \Omega_0$, thus there is $\tau_n = o(1)$ such that

$$\hat{\Omega}_n = \Omega_0 + o_p(\tau_n). \tag{53}$$

Let

$$\boldsymbol{\Sigma}_n = \left\{ \begin{bmatrix} A_\ell \\ A_u \end{bmatrix} \Omega \begin{bmatrix} A_\ell \\ A_u \end{bmatrix}' : \Omega \in \mathcal{S}, \|\Omega - \Omega_0\| \leq \tau_n \right\}.$$

To show (52), note that for all $\varepsilon > 0$,

$$\begin{aligned}
&P_n\left(\left|\bar{p}\left(c, \lambda_{P_n}, \hat{\Sigma}_n/n\right) - p(c)\right| > \varepsilon\right) \\
\leq& P_n\left(\left|\bar{p}\left(c, \lambda_{P_n}, \hat{\Sigma}_n/n\right) - p(c)\right| > \varepsilon, \hat{\Sigma}_n \in \boldsymbol{\Sigma}_n\right) + P_n\left(\hat{\Sigma}_n \notin \boldsymbol{\Sigma}_n\right) \\
\leq& P_n\left(\sup_{\Sigma \in \boldsymbol{\Sigma}_n} \left|\bar{p}\left(c, \lambda_{P_n}, \Sigma/n\right) - p(c)\right| > \varepsilon\right) + o(1) \\
=& \mathbf{1}\left[\sup_{\Sigma \in \boldsymbol{\Sigma}_n} \left|\bar{p}\left(c, \lambda_{P_n}, \Sigma/n\right) - p(c)\right| > \varepsilon\right] + o(1),
\end{aligned}$$

where the first inequality follows from $P(A) \leq P(A \cap B) + P(B^c)$, the second inequality follows from (53), and the last line is because $\bar{p}\left(c, \lambda_{P_n}, \Sigma/n\right)$ and $p(c)$ are non-random. Thus it suffices to show

$$\sup_{\Sigma \in \boldsymbol{\Sigma}_n} \left|\bar{p}\left(c, \lambda_{P_n}, \Sigma/n\right) - p(c)\right| \to 0.$$

To do so, there is a sequence $\Sigma_n \in \boldsymbol{\Sigma}_n$ such that

$$\limsup_n \sup_{\Sigma \in \boldsymbol{\Sigma}_n} \left|\bar{p}\left(c, \lambda_{P_n}, \Sigma/n\right) - p(c)\right| = \limsup_n \left|\bar{p}\left(c, \lambda_{P_n}, \Sigma_n/n\right) - p(c)\right|$$

and it suffices to show

$$\lim_n \bar{p}\left(c, \lambda_{P_n}, \Sigma_n/n\right) = p(c). \tag{54}$$

First consider the case when $\sqrt{n}\left(\lambda_{P_n,u,b_u} - \lambda_{P_n,\ell,b_\ell}\right) \in \mathbb{R}$ along P_n. Note that

$$g(T_\ell, T_m, T_u, c^{\mathrm{c}}_\ell, c^{\mathrm{c}}_m, c^{\mathrm{c}}_u) = \mathbf{1}\left[T_\ell > c^{\mathrm{m}}(c^{\mathrm{c}}_\ell, c) \text{ or } \{T_m > c^{\mathrm{m}}(c^{\mathrm{c}}_m, c) \text{ and } T_u > c^{\mathrm{m}}(c^{\mathrm{c}}_u, c)\}\right]$$

is bounded and continuous on D with

$$P(\overline{D}) = P\left(T_\ell = c^{\mathrm{m}}(c^{\mathrm{c}}_\ell, c) \text{ or } T_m = c^{\mathrm{m}}(c^{\mathrm{c}}_m, c) \text{ or } T_u = c^{\mathrm{m}}(c^{\mathrm{c}}_u, c)\right) = 0.$$

The second equality follows from (i) (T_ℓ, T_m, T_u) is continuously distributed and (ii) $T_\ell \perp c^{\mathrm{c}}_\ell$, $T_m \perp c^{\mathrm{c}}_m$, $T_u \perp c^{\mathrm{c}}_u$ by construction. Thus (54) follows from Portmanteau's Lemma.

Second, when $\sqrt{n}\left(\lambda_{P_n,u,b_u} - \lambda_{P_n,\ell,b_\ell}\right) \to \infty$ along P_n, let

$$\begin{aligned}\tilde{p}(c, \lambda_{P_n}, \Sigma_n/n) = \max\Big\{ & P\left(\hat{T}(\lambda_{P_n,\ell,b_\ell}) > \tilde{c}^{\mathrm{m}}(\lambda_{P_n,\ell,b_\ell}, c); \mathcal{N}(\lambda_{P_n}, \Sigma_n)\right), \\ & P\left(\hat{T}(\lambda_{P_n,u,b_u}) > \tilde{c}^{\mathrm{m}}(\lambda_{P_n,u,b_u}, c); \mathcal{N}(\lambda_{P_n}, \Sigma_n)\right)\Big\}\end{aligned} \tag{55}$$

and we have

$$0 \leq \bar{p}\left(c, \lambda_{P_n}, \Sigma_n/n\right) - \tilde{p}\left(c, \lambda_{P_n}, \Sigma_n/n\right) \leq P\left(\hat{T}(\theta_m) > \tilde{c}^{\mathrm{m}}(\theta_m, c); \mathcal{N}(\lambda_{P_n}, \Sigma_n)\right) = o(1)$$

where the last equality follows from Lemma 4. Then, (54) follows from

$$\begin{aligned}\tilde{p}(c, \lambda_{P_n}, \Sigma_n/n) =& \max\Big\{P\left(\hat{T}(\theta_{n,\ell}) > \tilde{c}^{\mathrm{m}}(\theta_{n,\ell}, c); \mathcal{N}(\lambda_{P_n}, \Sigma_n)\right), \\ & \qquad P\left(\hat{T}(\theta_{n,u}) > \tilde{c}^{\mathrm{m}}(\theta_{n,u}, c); \mathcal{N}(\lambda_{P_n}, \Sigma_n)\right)\Big\} \\ & \to \max\left\{P\left(T_\ell > c^{\mathrm{m}}_\ell(c)\right),\ P\left(T_u > c^{\mathrm{m}}_u(c)\right)\right\} = p(c)\end{aligned} \tag{56}$$

(56) follows from Portmanteau's Lemma and Lemma 4.

Step 2. I show that for all $\varepsilon > 0$,

$$\limsup_n P_n\left(\hat{c}^t_{P_n} \leq c^t - \varepsilon\right) = 0$$

where c^{t} is defined in (44) and $\hat{c}^t_{P_n}$ is defined in (76). Note that by definition

$$\bar{p}\left(\hat{c}^t_{P_n}, \lambda_{P_n}, \hat{\Sigma}_n/n\right) \leq \alpha - \eta$$

and $\bar{p}\left(c, \lambda_{P_n}, \hat{\Sigma}_n/n\right)$ is decreasing in c. Thus

$$\limsup_n P_n\left(\hat{c}^t_{P_n} \leq c^t - \varepsilon\right) \leq \limsup_n P_n\left(\bar{p}\left(c^t - \varepsilon, \lambda_{P_n}, \hat{\Sigma}_n/n\right) \leq \alpha - \eta\right) = 0$$

where the last equation is by

$$\bar{p}\left(c^t - \varepsilon, \lambda_{P_n}, \hat{\Sigma}_n/n\right) \xrightarrow{p} p\left(c^t - \varepsilon\right) > \alpha - \eta, \tag{57}$$

and (57) follows from Step 1 (52).

Step 3. Lastly, we show (75). Let

$$E(\theta, c) = \left\{\hat{T}(\theta) > \tilde{c}^{\mathrm{m}}(\theta, c)\right\}.$$

For all $\varepsilon > 0$, it holds that

$$\begin{aligned}
&\limsup_{n\to\infty} \max\left\{P_n\left(E(\theta_\ell, \hat{c}^t_{P_n}) \vee \left\{E(\theta_m, \hat{c}^t_{P_n}) \wedge E(\theta_u, \hat{c}^t_{P_n})\right\}\right),\right.\\
&\qquad\qquad\left. P_n\left(E(\theta_u, \hat{c}^t_{P_n}) \vee \left\{E(\theta_m, \hat{c}^t_{P_n}) \wedge E(\theta_\ell, \hat{c}^t_{P_n})\right\}\right)\right\}\\
\leq&\limsup_{n\to\infty} \max\left\{P_n\Big(E(\theta_\ell, c^t - \varepsilon) \vee \left\{E(\theta_m, c^t - \varepsilon) \wedge E(\theta_u, c^t - \varepsilon)\right\}\Big),\right.\\
&\qquad\qquad\left. P_n\Big(E(\theta_u, c^t - \varepsilon) \vee \left\{E(\theta_m, c^t - \varepsilon) \wedge E(\theta_\ell, c^t_{P_n} - \varepsilon)\right\}\Big)\right\}\\
&+\limsup_{n\to\infty} P_n\left(\hat{c}^t_{P_n} \leq c^t - \varepsilon\right)\\
=&p(c^t - \varepsilon)
\end{aligned} \tag{58}$$

(58) follows from Lemma 4 and Portmanteau's Lemma. Here I omit the subscript P_n in θ_ℓ, θ_m, θ_u and α in $\hat{c}^{\mathrm{m}}$, $\tilde{c}^{\mathrm{m}}$ for simplicity. Since (58) holds at all $\varepsilon > 0$, we can take a sequence of $\varepsilon \to 0$, then by Lemma 9,

$$\begin{aligned}
\limsup_{n\to\infty} \max&\left\{P_n\left(\hat{T}(\theta_\ell) > \hat{c}^{\mathrm{m}}(\theta_\ell; \alpha) \vee \left\{\hat{T}(\theta_m) > \hat{c}^{\mathrm{m}}(\theta_m; \alpha) \wedge \hat{T}(\theta_u) > \hat{c}^{\mathrm{m}}(\theta_u; \alpha)\right\}\right),\right.\\
&\left. P_n\left(\hat{T}(\theta_u) > \hat{c}^{\mathrm{m}}(\theta_u, \alpha) \vee \left\{\hat{T}(\theta_m) > \hat{c}^{\mathrm{m}}(\theta_m; \alpha) \wedge \hat{T}(\theta_\ell) > \hat{c}^{\mathrm{m}}(\theta_\ell; \alpha)\right\}\right)\right\}
\end{aligned}$$

$$\leq \lim_{\varepsilon \to 0} p(c^t - \varepsilon) = p(c^t) \leq \alpha - \eta.$$

The last inequality is by construction. This completes the proof. □

Proof of Theorem 2.

Proof. Part I. Symmetric Bounds. Since $\lambda_\ell = \lambda_u$ and $\hat{\lambda}_\ell = \hat{\lambda}_u$, let

$$\hat{T}(\theta) = \max \left\{ \min_{b \in \mathcal{B}} \left\{ \mathcal{Z}_b \right\}, \min_{b \in \mathcal{B}} \left\{ -\mathcal{Z}_b \right\} \right\},$$

where $\mathcal{Z}_{u,b} = \mathcal{Z}_b = \frac{\hat{\lambda}_b - \theta}{\hat{\sigma}_b / \sqrt{n}}$, $\mathcal{Z}_{\ell,b} = -\mathcal{Z}_b$.

Step 1. By Lemma 8, there is $\alpha_1' > \alpha$ such that

$$\liminf_n \inf_{P \in \mathcal{P}} P \left(\hat{c}^{\mathrm{t}} \leq \Phi^{-1}(1 - \frac{\alpha_1'}{2}) \right) = 1. \tag{59}$$

By Lemma 11, there is $\alpha_2' > \alpha$ such that

$$\liminf_n \inf_{P \in \mathcal{P}} P \left(\hat{T}(\theta) > \hat{c}^{\mathrm{c}}(\theta; \alpha^{\mathrm{c}}) \text{ for all } \theta \notin CI^{\mathrm{sim}}(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha_2') \right) = 1.$$

Let $\alpha' = \min\{\alpha_1', \alpha_2'\} > \alpha$, and then (36) follows from

$$\begin{aligned}
&\liminf_n \inf_{P \in \mathcal{P}} P \left(CI^{\mathrm{m}} \left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha \right) \subseteq CI^{\mathrm{sim}} \left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha' \right) \right) \\
= &\liminf_n \inf_{P \in \mathcal{P}} P \left(\hat{T}(\theta) > \hat{c}^{\mathrm{m}}(\theta; \alpha) \text{ for all } \theta \notin CI^{\mathrm{sim}}(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha') \right) \\
\geq &\liminf_n \inf_{P \in \mathcal{P}} P \left(\hat{T}(\theta) > \hat{c}^{\mathrm{c}}(\theta; \alpha^{\mathrm{c}}) \text{ for all } \theta \notin CI^{\mathrm{sim}}(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha'), \hat{c}^{\mathrm{t}} \leq \Phi^{-1}(1 - \frac{\alpha'}{2}) \right) \\
\geq &\liminf_n \inf_{P \in \mathcal{P}} P \left(\hat{T}(\theta) > \hat{c}^{\mathrm{c}}(\theta; \alpha^{\mathrm{c}}) \text{ for all } \theta \notin CI^{\mathrm{sim}}(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha') \right) \\
&\quad + \liminf_n \inf_{P \in \mathcal{P}} P \left(\hat{c}^{\mathrm{t}} \leq \Phi^{-1}(1 - \frac{\alpha'}{2}) \right) - 1 \\
\geq &\liminf_n \inf_{P \in \mathcal{P}} P \left(\hat{T}(\theta) > \hat{c}^{\mathrm{c}}(\theta; \alpha^{\mathrm{c}}) \text{ for all } \theta \notin CI^{\mathrm{sim}}(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha_2') \right) \\
&\quad + \liminf_n \inf_{P \in \mathcal{P}} P \left(\hat{c}^{\mathrm{t}} \leq \Phi^{-1}(1 - \frac{\alpha_1'}{2}) \right) - 1 = 1
\end{aligned}$$

Step 2. I show (37) with $\theta_n = \theta_\ell - \frac{\kappa}{\sqrt{n}}$. Note that by (36), there is $\alpha' > \alpha$ such that

$$
\begin{aligned}
&\liminf_n P_n\left(\theta_n \notin CI^{\mathrm{m}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) - P_n\left(\theta_n \notin CI^{\mathrm{sim}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) \\
\geq &\liminf_n P_n\left(\theta_n \notin CI^{\mathrm{sim}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha'\right)\right) - P_n\left(\theta_n \notin CI^{\mathrm{sim}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) \\
\geq &\liminf_n P_n\left(\hat{T}(\theta_n) \in \left(\Phi^{-1}(1-\frac{\alpha'}{2}), \Phi^{-1}(1-\frac{\alpha}{2})\right)\right) \qquad (60)
\end{aligned}
$$

Under P_n, we can show that there is a subsequence P_{a_n} such that (77), (78) hold and

$$
\hat{T}(\theta_n) \xrightarrow{d} T^* := \max\left\{\min_{b\in\mathcal{B}}\left\{Z_b + \bar{\lambda}_b + \frac{\kappa}{\sigma_b}\right\},\ \min_{b\in\mathcal{B}}\left\{-Z_b - \bar{\lambda}_b - \frac{\kappa}{\sigma_b}\right\}\right\},
$$

where

$$
\bar{\lambda}_b = \lim_{a_n} \frac{\sqrt{a_n}(\lambda_b - \theta_\ell)}{\sigma_b} \geq 0.
$$

Let $c_1 = \Phi^{-1}(1-\frac{\alpha'}{2})$, $c_2 = \Phi^{-1}(1-\frac{\alpha}{2})$. Then I show that there is $\kappa \in \mathbb{R}$ such that

$$
P\left(T^* \in (c_1, c_2)\right) > 0.
$$

To do so, let b^* be the element with largest variance, i.e. $\sigma_{b^*} \geq \max_{b\in\bar{\mathcal{B}}} \sigma_b$, where

$$
\bar{\mathcal{B}} = \left\{b \in \mathcal{B} : \bar{\lambda}_b \in \mathbb{R}\right\}.
$$

Note that we have $\bar{\lambda}_{b_\ell} = 0$, thus $\bar{\mathcal{B}} \neq \emptyset$. Then

$$
\begin{aligned}
&P\left(T^* \in (c_1, c_2)\right) \\
\geq &P\left(c_2 \geq Z_{b^*} + \bar{\lambda}_{b^*} + \frac{\kappa}{\sigma_{b^*}} \geq c_1, Z_b + \bar{\lambda}_b + \frac{\kappa}{\sigma_b} \geq c_1, b \in \bar{\mathcal{B}}\backslash\{b^*\}\right) \\
= &P\left(c_2 \geq Z_{b^*} + \bar{\lambda}_{b^*} + \frac{\kappa}{\sigma_{b^*}} \geq c_1, \mathbb{E}_b \geq c_1 - \rho_{b^*b}Z_{b^*} - \bar{\lambda}_b - \frac{\kappa}{\sigma_b}, b \in \bar{\mathcal{B}}\backslash\{b^*\}\right) \\
\geq &P\left(c_2 \geq Z_{b^*} + \bar{\lambda}_{b^*} + \frac{\kappa}{\sigma_{b^*}} \geq c_1, \mathbb{E}_b \geq c_1 - \bar{\lambda}_b - \frac{\kappa}{\sigma_b} - |\rho_{b^*b}|(c_1 - \bar{\lambda}_{b^*} - \frac{\kappa}{\sigma_{b^*}}), b \in \bar{\mathcal{B}}\backslash\{b^*\}\right) \\
= &P\left(c_2 \geq Z_{b^*} + \bar{\lambda}_{b^*} + \frac{\kappa}{\sigma_{b^*}} \geq c_1\right) \times \\
&\quad P\left(\mathbb{E}_b \geq c_1 - \bar{\lambda}_b - |\rho_{b^*b}|\left(c_1 - \bar{\lambda}_{b^*}\right) - \left(\frac{\sigma_{b^*}}{\sigma_b} - |\rho_{b^*b}|\right)\frac{\kappa}{\sigma_{b^*}}, b \in \bar{\mathcal{B}}\backslash\{b^*\}\right)
\end{aligned}
$$

where $\mathbb{E}_b = Z_b - \rho(b, b^*) Z_{b^*}$. There is $\kappa \in \mathbb{R}$ such that

$$P\left(\mathbb{E}_b \geq \bar{c} - \bar{\lambda}_b - |\rho_{b^*b}|\left(\bar{c} - \bar{\lambda}_{b^*}\right) - \left(\frac{\sigma_{b^*}}{\sigma_b} - |\rho_{b^*b}|\right)\frac{\kappa}{\sigma_{b^*}}, b \in \bar{\mathcal{B}} \backslash \{b^*\}\right) > 0$$

and therefore

$$P\left(T^* \in (c_1, c_2)\right) > 0. \tag{61}$$

(37) follows from (60) and (61).

Part II. Large Bounds. I first show (36). By Lemma 12, there is $\alpha_1' > \alpha$ such that

$$\liminf_n \inf_n P\left(\hat{c}^{\mathrm{t}} \leq \Phi^{-1}(1 - \frac{\alpha_1'}{2})\right) = 1.$$

Let $\alpha' = \min\{\alpha_1', 2\alpha^{\mathrm{c}}\} > \alpha$ and $c_1 = \Phi^{-1}(1 - \frac{\alpha'}{2})$. I show that

$$\liminf_n \inf_n P\left(\theta \notin CI^{\mathrm{m}}(\hat{\lambda}, \hat{\Sigma}_n/n, \alpha) \text{ for all } \theta > \max_{b \in \mathcal{B}} \hat{\lambda}_{u,b} + \frac{\hat{\sigma}_{u,b}}{\sqrt{n}} c_1\right) = 1.$$

The proof for the lower bound is symmetric.

Let $\kappa_n' \to \infty$ and $\kappa_n' \ll \kappa_n$. Lemma 2 suggests that

$$\liminf_n \inf_n P\left(\theta \notin CI^{\mathrm{m}}(\hat{\lambda}, \hat{\Sigma}_n/n, \alpha) \text{ for all } \theta > \max_{b \in \mathcal{B}} \hat{\lambda}_{u,b} + \frac{\hat{\sigma}_{u,b}}{\sqrt{n}} \kappa_n'\right) = 1.$$

Then I simplify $\hat{c}^{\mathrm{c}}(\theta, \alpha^{\mathrm{c}})$ for

$$\theta \in \left(\max_{b \in \mathcal{B}} \hat{\lambda}_{u,b} + \frac{\hat{\sigma}_{u,b}}{\sqrt{n}} c_1,\ \max_{b \in \mathcal{B}} \hat{\lambda}_{u,b} + \frac{\hat{\sigma}_{u,b}}{\sqrt{n}} \kappa_n'\right].$$

In this case, under (35),

$$\mathcal{Z}_{\ell,\hat{b}_\ell} \leq \mathcal{Z}_{\ell,b_\ell} = \frac{\hat{\lambda}_{\ell,b_\ell} - \theta}{\hat{\sigma}_{\ell,b_\ell}/\sqrt{n}} \leq \frac{\hat{\lambda}_{\ell,b_\ell} - \hat{\lambda}_{u,b_u} - \hat{\sigma}_{u,b_u}/\sqrt{n} c_1}{\hat{\sigma}_{\ell,b_\ell}/\sqrt{n}} \to -\infty \tag{62}$$

$$\mathcal{Z}_{u,\hat{b}_u} = \frac{\theta - \hat{\lambda}_{u,\hat{b}_u}}{\hat{\sigma}_{u,\hat{b}_u}/\sqrt{n}} \in (c_1,\ \kappa_n']. \tag{63}$$

Thus, with probability approaching one,

$$\hat{T}(\theta) = \mathcal{Z}_{u,\hat{b}_u} > c_1, \tag{64}$$
$$\hat{c}^{\mathrm{c}}(\theta, \alpha^{\mathrm{c}}) = \Phi^{-1}\left(\alpha^{\mathrm{c}} \Phi\left(t_{u,1}(\theta, \hat{b}_u)\right) + (1 - \alpha^{\mathrm{c}}) \Phi\left(t_{u,2}(\theta, \hat{b}_u)\right)\right).$$

Moreover, (62) and (63) implies that

$$\Phi\left(t_{u,1}(\theta,\hat{b}_u)\right) \leq \Phi\left(\left(1+\rho_{\ell u}(b_\ell,\hat{b}_u)\right)^{-1}\left(\mathcal{Z}_{\ell,b_\ell}+\rho_{\ell u}(b_\ell,\hat{b}_u)\mathcal{Z}_{u,\hat{b}_u}\right)\right)$$
$$=\Phi\left(\frac{\mathcal{Z}_{\ell,b_\ell}-\mathcal{Z}_{u,\hat{b}_u}}{1+\rho_{\ell u}(b_\ell,\hat{b}_u)}+\mathcal{Z}_{u,\hat{b}_u}\right) \xrightarrow{p} 0.$$

Therefore,

$$\hat{c}^{\mathrm{c}}(\theta,\alpha^{\mathrm{c}}) \leq \Phi^{-1}\left((1-\alpha^{\mathrm{c}})\Phi\left(t_{u,2}(\theta,\hat{b}_u)\right)\right)+o(1) \leq c_1 \text{ w.p.a. } 1,$$

where the last inequality follows from $\alpha' \leq 2\alpha^{\mathrm{c}}$. Thus by construction,

$$\hat{c}^{\mathrm{m}}(\theta,\alpha) \leq c_1 < \hat{T}(\theta),$$

and θ is rejected.

The proof for (37) is similar to Part I Step 2. □

Proof of Theorem 3.

Proof. I show the results with $\theta_n = \min_{b\in\mathcal{B}} \lambda_{\ell,b} - \frac{\kappa_n'}{\sqrt{n}}a$. By Lemma 2,

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n} P\left(\theta_{\ell,n} \notin CI^{\mathrm{m}}\left(\hat{\lambda}_n,\hat{\Sigma}_n/n;\alpha\right)\right)=1. \tag{65}$$

YKHS23 confidence interval has form

$$\left[\hat{\lambda}_{m,\min}-\sqrt{\frac{n}{m}}Q^*\left(\hat{\lambda}^*_{n,\min}-\hat{\lambda}_{n,\min},\hat{p}\right),\hat{\lambda}_{m,\max}-\sqrt{\frac{n}{m}}Q^*\left(\hat{\lambda}^*_{n,\max}-\hat{\lambda}_{n,\max},1-\hat{p}\right)\right],$$

with $\hat{\lambda}^*_{n,\ell,b}$ calculated by empirical bootstrap, $\hat{\lambda}_{m,\ell,b}$ calculate by a subsample of size m,

$$\hat{\lambda}_{m,\min}=\min_{b\in\mathcal{B}}\hat{\lambda}_{m,\ell,b},\ \hat{\lambda}_{n,\min}=\min_{b\in\mathcal{B}}\hat{\lambda}_{n,\ell,b},\ \hat{\lambda}^*_{n,\min}=\min_{b\in\mathcal{B}}\hat{\lambda}^*_{n,\ell,b},$$

$\hat{p}\xrightarrow{p} p^*\in\left[\frac{\alpha}{2},\alpha\right]$. The upper bound is defined symmetrically. Note that

$$P\left(\theta_n \notin CI^{\mathrm{YKHS}}\left(\hat{\lambda}_n,\hat{\Sigma}_n/n;\alpha\right)\right)$$
$$\leq P\left(\theta_n < \hat{\lambda}_{m,\min}-\sqrt{\frac{n}{m}}Q^*\left(\hat{\lambda}^*_{n,\min}-\hat{\lambda}_{n,\min},\hat{p}\right)\right)$$

$$+P\left(\theta_n > \hat{\lambda}_{m,\max} - \sqrt{\frac{n}{m}}Q^*\left(\hat{\lambda}^*_{n,\max} - \hat{\lambda}_{n,\max}, 1-\hat{p}\right)\right)$$

$$=P\left(Q^*\left(\sqrt{n}\left(\hat{\lambda}^*_{n,\min} - \hat{\lambda}_{n,\min}\right), \hat{p}\right) < \sqrt{m}(\hat{\lambda}_{m,\min} - \theta_n)\right) \tag{66}$$

$$+P\left(Q^*\left(\sqrt{n}(\hat{\lambda}^*_{n,\max} - \hat{\lambda}_{n,\max}), 1-\hat{p}\right) > \sqrt{m}(\hat{\lambda}_{m,\max} - \theta_n)\right). \tag{67}$$

As for (66), note that

$$\sqrt{m}(\hat{\lambda}_{m,\min} - \theta_n) = \sqrt{m}(\hat{\lambda}_{m,\min} - \lambda_{\ell,b_\ell}) + \sqrt{m}(\lambda_{\ell,b_\ell} - \theta_n)$$
$$= \sqrt{m}(\hat{\lambda}_{m,\min} - \lambda_{\ell,b_\ell}) + o(1) \xrightarrow{d} \min_{b\in\mathcal{B}} Z_{\ell,b} + \tau_{\ell,b}$$

where $\tau_{\ell,b} = \lim_m \sqrt{m}\,(\lambda_{m,b} - \lambda_{\ell,b_\ell})$, and the limit distribution is continuous. Thus

$$(66) = P\left(Q^*\left(\sqrt{n}\left(\hat{\lambda}^*_{n,\min} - \hat{\lambda}_{n,\min}\right), \hat{p}\right) < \sqrt{m}(\hat{\lambda}_{m,\min} - \lambda_{\ell,b_\ell})\right) + o(1).$$

Similarly, if $\sqrt{m}(\lambda_{u,b_u} - \theta_n) \in \mathbb{R}$, we have

$$(67) = P\left(Q^*\left(\sqrt{n}(\hat{\lambda}^*_{n,\max} - \hat{\lambda}_{n,\max}), 1-\hat{p}\right) > \sqrt{m}(\hat{\lambda}_{m,\max} - \lambda_{\ell,b_\ell})\right) + o(1) \tag{68}$$

with similar argument. If $\sqrt{m}(\lambda_{u,b_u} - \theta_n) \to \infty$, (68) still holds since both side of the equation is $o(1)$. In sum, we have

$$P\left(\theta_n \notin CI^{\text{YKHS}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) = P\left(\lambda_{\ell,b_\ell} \notin CI^{\text{YKHS}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) + o(1),$$

thus by Theorem 2(d) in YKHS23, it holds that

$$\limsup_n \sup_P P\left(\theta_n \notin CI^{\text{YKHS}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) \leq \alpha. \tag{69}$$

(65) and (69) complete the proof. □

References

Andrews, Donald W K and Patrik Guggenberger (2009): "Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities," *Econometric Theory*, 25 (3), 669–709.

Andrews, Donald W K and Soonwoo Kwon (2023): "Misspecified Moment

Inequality Models: Inference and Diagnostics," *The Review of Economic Studies.*

Andrews, Donald W K and Xiaoxia Shi (2013): "Inference based on conditional moment inequalities," *Econometrica*, 81 (2), 609–666.

Andrews, Donald W K and Gustavo Soares (2010): "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78 (1), 119–157.

Andrews, Isaiah, Toru Kitagawa, and Adam McCloskey (2021): "Inference after estimation of breaks," *Journal of Econometrics*, 224 (1), 39–59.

——— (2024): "Inference on winners," *The Quarterly Journal of Economics*, 139 (1), 305–358.

Andrews, Isaiah and Anna Mikusheva (2016): "Conditional inference with a functional nuisance parameter," *Econometrica*, 84 (4), 1571–1612.

Andrews, Isaiah, Jonathan Roth, and Ariel Pakes (2023): "Inference for Linear Conditional Moment Inequalities," *The Review of Economic Studies.*

Apfel, Nicolas and Frank Windmeijer (2022): "The Falsification Adaptive Set in Linear Models with Instrumental Variables that Violate the Exogeneity or Exclusion Restriction," *arXiv preprint arXiv:2212.04814.*

Ban, Kyunghoon and Desire Kedagni (2022): "Generalized Difference-in-differences Models: Robust Bounds," *arXiv preprint arXiv:2211.06710.*

Benzarti, Youssef and Dorian Carloni (2019): "Who really benefits from consumption tax cuts? Evidence from a large VAT reform in France," *American Economic Journal: Economic Policy*, 11 (1), 38–63.

Berger, Roger L (1989): "Uniformly more powerful tests for hypotheses concerning linear inequalities and normal means," *Journal of the American Statistical Association*, 84 (405), 192–199.

Blomquist, Sören, Whitney K Newey, Anil Kumar, and Che-Yuan Liang (2021): "On bunching and identification of the taxable income elasticity," *Journal of Political Economy*, 129 (8), 2320–2343.

Bugni, Federico A, Ivan A Canay, and Xiaoxia Shi (2015): "Specification tests for partially identified models defined by moment inequalities," *Journal of Econometrics*, 185 (1), 259–282.

Casella, George and Roger L Berger (2021): *Statistical inference*, Cengage Learning.

Chernozhukov, Victor, Han Hong, and Elie Tamer (2007): "Estimation and confidence regions for parameter sets in econometric models 1," *Econometrica*, 75 (5), 1243–1284.

Chernozhukov, Victor, Sokbae Lee, and Adam M Rosen (2013): "Intersection bounds: estimation and inference," *Econometrica*, 81 (2), 667–737.

Christensen, Peter, David A Keiser, and Gabriel E Lade (2023): "Economic effects of environmental crises: Evidence from Flint, Michigan," *American Economic Journal: Economic Policy*, 15 (1), 196–232.

Cohen, Arthur, Constantine Gatsonis, and John Marden (1983): "Hypothesis tests and optimality properties in discrete multivariate analysis," in *Studies in Econometrics, Time Series, and Multivariate Statistics*, 379–405.

Conley, Timothy G, Christian B Hansen, and Peter E Rossi (2012): "Plausibly exogenous," *Review of Economics and Statistics*, 94 (1), 260–272.

Dustmann, Christian, Attila Lindner, Uta Schönberg, Matthias Umkehrer, and Philipp Vom Berge (2022): "Reallocation effects of the minimum wage," *The Quarterly Journal of Economics*, 137 (1), 267–328.

Fang, Zheng (2018): "Optimal plug-in estimators of directionally differentiable functionals," *Unpublished Manuscript*.

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

Hasegawa, Raiden B, Daniel W Webster, and Dylan S Small (2019): "Evaluating Missouri's handgun purchaser law: a bracketing method for addressing concerns about history interacting with group," *Epidemiology*, 30 (3), 371–379.

Hirano, Keisuke and Jack R Porter (2012): "Impossibility results for nondifferentiable functionals," *Econometrica*, 80 (4), 1769–1790.

Horowitz, Joel L (2019): "Bootstrap methods in econometrics," *Annual Review of Economics*, 11, 193–224.

Imbens, Guido W and Charles F Manski (2004): "Confidence intervals for partially identified parameters," *Econometrica*, 72 (6), 1845–1857.

Kim, Deborah (2024): "Testing Sign Agreement," *Unpublished Manuscript*.

Kleibergen, Frank (2005): "Testing parameters in GMM without assuming that they are identified," *Econometrica*, 73 (4), 1103–1123.

Kolesár, Michal and Christoph Rothe (2018): "Inference in regression discontinuity designs with a discrete running variable," *American Economic*

Review, 108 (8), 2277–2304.

Lee, Jason D, Dennis L Sun, Yuekai Sun, and Jonathan E Taylor (2016): "Exact post-selection inference, with application to the lasso," *The Annals of Statistics*, 907–927.

Lovenheim, Michael F and Alexander Willén (2019): "The long-run effects of teacher collective bargaining," *American Economic Journal: Economic Policy*, 11 (3), 292–324.

Manski, Charles F and John V Pepper (2018): "How do right-to-carry laws affect crime rates? Coping with ambiguity using bounded-variation assumptions," *Review of Economics and Statistics*, 100 (2), 232–244.

Masten, Matthew A and Alexandre Poirier (2021): "Salvaging falsified instrumental variable models," *Econometrica*, 89 (3), 1449–1469.

Miller, Douglas L, Francesca Molinari, and Jörg Stoye (2024): "Testing Sign Congruence," *arXiv preprint arXiv:2405.11759*.

Moreira, Marcelo J (2003): "A conditional likelihood ratio test for structural models," *Econometrica*, 71 (4), 1027–1048.

Perlman, Michael D and Lang Wu (1999): "The emperor's new tests," *Statistical Science*, 14 (4), 355–369.

Ponomarev, Kirill (2022): "Efficient Estimation of Directionally Differentiable Functionals," .

Rambachan, Ashesh and Jonathan Roth (2023): "A More Credible Approach to Parallel Trends," *The Review of Economic Studies*, 90 (5), 2555–2591.

Romano, Joseph P and Azeem M Shaikh (2008): "Inference for identifiable parameters in partially identified econometric models," *Journal of Statistical Planning and Inference*, 138 (9), 2786–2807.

Rosen, Adam M (2008): "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," *Journal of Econometrics*, 146 (1), 107–117.

Saez, Emmanuel (2010): "Do taxpayers bunch at kink points?" *American Economic Journal: Economic Policy*, 2 (3), 180–212.

Stoye, Jörg (2009): "More on confidence intervals for partially identified parameters," *Econometrica*, 77 (4), 1299–1315.

——— (2020): "Simple misspecification adaptive inference for interval identified parameters," Tech. rep., cemmap working paper.

Ye, Ting, Luke Keele, Raiden Hasegawa, and Dylan S Small (2023): "A negative correlation strategy for bracketing in difference-in-differences," *Journal of the American Statistical Association*, 1–13.

Online Appendices

B Auxiliary Lemmas

Lemma 2. *($\sqrt{n}$ Convergence Rate) Suppose Assumptions 1, 2, 3, 4, and 5 hold. For all $\varepsilon > 0$, there is $\bar{\kappa} \in \mathbb{R}_+$ such that*

$$\liminf_n \inf_{P \in \mathcal{P}} P\left(CI^m \subseteq \left[\theta_\ell - \frac{\bar{\kappa}}{\sqrt{n}}, \theta_u + \frac{\bar{\kappa}}{\sqrt{n}}\right]\right) > 1 - \varepsilon.$$

Proof. It suffices to show that there is $\bar{\kappa} \in \mathbb{R}_+$ such that

$$\liminf_n \inf_{P \in \mathcal{P}} P\left(\hat{T}(\theta) > \hat{c}^{\mathrm{m}}(\theta;\alpha) \text{ for all } \theta \notin \left[\theta_\ell - \frac{\bar{\kappa}}{\sqrt{n}}, \theta_u + \frac{\bar{\kappa}}{\sqrt{n}}\right]\right) > 1 - \varepsilon.$$

Following similar argument in Lemma 3, there is subsequence $P_{a_n} \in \mathcal{P}$ such that

$$\begin{aligned}
&\liminf_n \inf_{P \in \mathcal{P}} P\left(\hat{T}(\theta) > \hat{c}^{\mathrm{m}}(\theta;\alpha) \text{ for all } \theta \notin \left[\theta_\ell - \frac{\bar{\kappa}}{\sqrt{n}}, \theta_u + \frac{\bar{\kappa}}{\sqrt{n}}\right]\right) \\
=&\lim_{a_n} P_{a_n}\left(\hat{T}(\theta) > \hat{c}^{\mathrm{m}}(\theta;\alpha) \text{ for all } \theta \notin \left[\theta_\ell - \frac{\bar{\kappa}}{\sqrt{a_n}}, \theta_u + \frac{\bar{\kappa}}{\sqrt{a_n}}\right]\right)
\end{aligned}$$

and $\Sigma(P_{a_n}) \to \Sigma_0$. In addition, note that

$$\begin{aligned}
&P_{a_n}\left(\hat{T}(\theta) > \hat{c}^{\mathrm{m}}(\theta;\alpha) \text{ for all } \theta \notin \left[\theta_\ell - \frac{\bar{\kappa}}{\sqrt{a_n}}, \theta_u + \frac{\bar{\kappa}}{\sqrt{a_n}}\right]\right) \\
\geq &P_{a_n}\left(\hat{T}(\theta) > \hat{c}^{\mathrm{m}}(\theta;\alpha) \text{ for all } \theta < \theta_\ell - \frac{\bar{\kappa}}{\sqrt{a_n}}\right) \\
&\quad + P_{a_n}\left(\hat{T}(\theta) > \hat{c}^{\mathrm{m}}(\theta;\alpha) \text{ for all } \theta > \theta_u + \frac{\bar{\kappa}}{\sqrt{a_n}}\right) - 1.
\end{aligned}$$

Therefore, it suffices to show that for all $\varepsilon > 0$, there is $\bar{\kappa} \in \mathbb{R}_+$ such that

$$P_{a_n}\left(\hat{T}(\theta) > \hat{c}^{\mathrm{m}}(\theta;\alpha) \text{ for all } \theta < \theta_\ell - \frac{\bar{\kappa}}{\sqrt{a_n}}\right) \geq 1 - \frac{\varepsilon}{2}, \tag{70}$$

The proof for the upper bound is symmetric. In the following proof, I use n for subsequence a_n to simplify notation.

First, I show that for all $\varepsilon > 0$, there is $\bar{\kappa}_1$ such that

$$\liminf_n P_n\left(A_n(\bar{\kappa}_1)\right) \geq 1 - \frac{\varepsilon}{6}, \text{ where} \tag{71}$$
$$A_n(\bar{\kappa}_1) = \left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,\hat{b}_\ell} > \Phi^{-1}(1 - \frac{\alpha-\eta}{2}), \text{ for all } \theta \leq \theta_\ell - \frac{\bar{\kappa}_1}{\sqrt{n}}\right\}.$$

To see this,

$$\begin{aligned}
&P_n\left(A_n(\bar{\kappa}_1)\right) \\
=&P_n\left(\mathcal{Z}_{\ell,\hat{b}_\ell} \geq \min_{b\in\mathcal{B}} \mathcal{Z}_{u,b}, \mathcal{Z}_{\ell,\hat{b}_\ell} > \Phi^{-1}(1-\frac{\alpha-\eta}{2}), \text{ for all } \theta \leq \theta_\ell - \frac{\bar{\kappa}_1}{\sqrt{n}}\right) \\
\geq&P_n\left(\mathcal{Z}_{\ell,\hat{b}_\ell} \geq \mathcal{Z}_{u,b_u}, \mathcal{Z}_{\ell,\hat{b}_\ell} > \Phi^{-1}(1-\frac{\alpha-\eta}{2}), \text{ for all } \theta \leq \theta_\ell - \frac{\bar{\kappa}_1}{\sqrt{n}}\right) \\
\geq&P_n\left(\frac{\hat{\lambda}_{\ell,\hat{b}_\ell} - \lambda_{\ell,\hat{b}_\ell}}{\hat{\sigma}_{\ell,\hat{b}_\ell}/\sqrt{n}} + \frac{\bar{\kappa}_1}{\hat{\sigma}_{\ell,\hat{b}_\ell}} \geq \frac{\theta_u - \hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}} - \frac{\bar{\kappa}_1}{\hat{\sigma}_{u,b_u}}, \frac{\hat{\lambda}_{\ell,\hat{b}_\ell} - \lambda_{\ell,\hat{b}_\ell}}{\hat{\sigma}_{\ell,\hat{b}_\ell}/\sqrt{n}} + \frac{\bar{\kappa}_1}{\hat{\sigma}_{\ell,\hat{b}_\ell}} > \Phi^{-1}(1-\frac{\alpha-\eta}{2})\right) \\
=&P_n\Bigg(\bar{\kappa}_1 \geq \left(\frac{1}{\hat{\sigma}_{\ell,\hat{b}_\ell}} + \frac{1}{\hat{\sigma}_{u,b_u}}\right)^{-1}\left(\frac{\theta_u - \hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}} - \frac{\hat{\lambda}_{\ell,\hat{b}_\ell} - \lambda_{\ell,\hat{b}_\ell}}{\hat{\sigma}_{\ell,\hat{b}_\ell}/\sqrt{n}}\right), \\
&\qquad \bar{\kappa}_1 > \hat{\sigma}_{\ell,\hat{b}_\ell}\Phi^{-1}(1-\frac{\alpha-\eta}{2}) - \sqrt{n}(\hat{\lambda}_{\ell,\hat{b}_\ell} - \lambda_{\ell,\hat{b}_\ell})\Bigg).
\end{aligned}$$

The existence of $\bar{\kappa}_1$ follows from

$$\sqrt{n}\left(\theta_u - \hat{\lambda}_{u,b_u}\right) = O_P(1), \ \sqrt{n}\left(\hat{\lambda}_{\ell,\hat{b}_\ell} - \lambda_{\ell,\hat{b}_\ell}\right) = O_P(1).$$

Second, if $\min_{b\in\mathcal{B}} \rho_{\ell u}(\tilde{b}, b) > -1$, there is $\xi \in (0,1)$ such that $\hat{\rho}_{\ell u}(\tilde{b}, b_u) > \xi - 1$ with probability approaching one by Assumption 1, 2, 3, and 4. Then, for all $\varepsilon > 0$, there is $\bar{M} \in \mathbb{R}$ such that

$$\liminf_n P_n\left(B_n\right) \geq 1 - \frac{\varepsilon}{6}, \text{ where } B_n = B_{1n} \cup B_{2n}, \tag{72}$$
$$B_{1n} = \left\{\min_{b\in\mathcal{B}} \rho_{\ell u}(\hat{b}_\ell, b) = -1\right\},$$
$$B_{2n} = \left\{\min_{b\in\mathcal{B}} \rho_{\ell u}(\hat{b}_\ell, b) > -1, \left(1 + \hat{\rho}_{\ell u}(\hat{b}_\ell, b_u)\right)^{-1} \frac{\lambda_{u,b_u} - \hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}} \leq \bar{M}\right\}$$

because

$$
\begin{aligned}
&\liminf_n P_n\left(B_{1n} \cup B_{2n}\right) \\
=&1-\liminf_n P_n\left(\min_{b \in \mathcal{B}} \rho_{\ell u}(\hat{b}_\ell, b) > -1, \left(1+\hat{\rho}_{\ell u}(\hat{b}_\ell, b_u)\right)^{-1} \frac{\lambda_{u,b_u}-\hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}} > \bar{M}\right) \\
\geq&1-\liminf_n P_n\left(\min_{b \in \mathcal{B}} \rho_{\ell u}(\hat{b}_\ell, b) > -1, \ \frac{1}{\xi}\left|\frac{\lambda_{u,b_u}-\hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}}\right| > \bar{M}\right) \\
\geq&1-\liminf_n P_n\left(\frac{1}{\xi}\left|\frac{\lambda_{u,b_u}-\hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b}/\sqrt{n}}\right| > \bar{M}\right)
\end{aligned}
$$

and the existence of $\bar{M}$ follows from $\left|\frac{\lambda_{u,b_u}-\hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}}\right| = O_P(1)$.

By similar argument in (71), there is $\bar{\kappa}_2$ such that

$$\liminf_n P_n\left(C_n(\bar{\kappa}_2)\right) \geq 1-\frac{\varepsilon}{6}, \text{ where} \tag{73}$$

$$C_n(\bar{\kappa}_2) = \left\{\hat{T}(\theta) > \bar{z} \text{ for all } \theta \leq \theta_\ell - \frac{\bar{\kappa}_2}{\sqrt{n}}\right\}, \tag{74}$$

$\bar{z}$ is defined in Lemma 6 with $\bar{M}$ given above (72).

In sum, let $\bar{\kappa} = \max\{\bar{\kappa}_1, \bar{\kappa}_2, 0\}$,

$$D_n = \left\{\hat{T}(\theta_n) > \hat{c}^{\mathrm{m}}(\theta_n, \alpha^{\mathrm{c}}) \text{ for all } \theta \leq \theta_\ell - \bar{\kappa}/\sqrt{n}\right\},$$

we have

$$
\begin{aligned}
&\liminf_n P_n\left(D_n\right) \\
\geq&\liminf_n P_n\left(A_n(\bar{\kappa}) \cap B_n \cap C_n(\bar{\kappa}) \cap D_n\right) \\
=&\liminf_n P_n\left(A_n(\bar{\kappa}) \cap B_n \cap C_n(\bar{\kappa})\right) \\
\geq&\liminf_n P_n\left(A_n(\bar{\kappa})\right) + P_n\left(B_n\right) + P_n\left(C_n(\bar{\kappa})\right) - 2 \geq 1-\frac{\varepsilon}{2}.
\end{aligned}
$$

where the equality follows from Lemma 5: the three assumptions in Lemma 5 hold because (i) $\bar{\kappa} \geq 0$, (ii) $A_n(\bar{\kappa})$, (iii) $B_n \cap C_n(\bar{\kappa})$. The last inequality follows from (71), (72) and (73), and thus (70) holds. □

Lemma 3. *Under Assumptions 1, 2, 3, 4, 5, to prove that*

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \sup_{\theta\in\left[\theta_{P,\ell},\theta_{P,u}\right]} P\left(\theta \notin CI^m\left(\hat{\lambda}_n, \hat{\Sigma}_n/n;\alpha\right)\right) \leq \alpha,$$

it suffices to show that we have

$$\limsup_{n\to\infty} \max\left\{P_n\left(E_\ell \cup \{E_m \cap E_u\}\right), P_n\left(E_u \cup \{E_m \cap E_\ell\}\right)\right\} \leq \alpha - \eta, \tag{75}$$

for all sequence $\{P_n\} \in \mathcal{P}^\infty = \times_{n=1}^\infty \mathcal{P}_n$ *with*

$$E_k = \left\{\hat{T}(\theta_k) > \tilde{c}^m(\theta_k, \hat{c}^t_{P_n})\right\},\ \hat{c}^t_{P_n} = \inf_c \left\{c \geq 0 : \bar{p}\left(c, \lambda_{P_n}, \hat{\Sigma}_n/n\right) + \eta \leq \alpha\right\}, \tag{76}$$

and P_n *satisfying*

1. *The convergence of* Ω,
$$\Omega(P_n) \to \Omega_0 \in \mathcal{S}; \tag{77}$$

2. *The convergence of*
$$(\lambda_{n,k\ell}, \lambda_{n,ku}) = \left(\left(\frac{\lambda_{P_n,\ell,b} - \theta_k}{\sigma_{P_n,\ell,b}/\sqrt{n}}\right)_{b\in\mathcal{B}}, \left(\frac{\theta_k - \lambda_{P_n,u,b}}{\sigma_{P_n,u,b}/\sqrt{n}}\right)_{b\in\mathcal{B}}\right) \to (\lambda_{k\ell}, \lambda_{ku}), \tag{78}$$

   with $\lambda_{\ell\ell} \in \Lambda_0$, $\lambda_{uu} \in \Lambda_0$, $\lambda_{\ell u}, \lambda_{mu}, \lambda_{m\ell}, \lambda_{u\ell} \in \Lambda_-$,

$$\Lambda_0 = \left\{\lambda \in [0, +\infty]^{|\mathcal{B}|} : \min_{b\in\mathcal{B}} \lambda_b = 0\right\}$$
$$\Lambda_- = \left\{\lambda \in [-\infty, +\infty]^{|\mathcal{B}|} : \min_{b\in\mathcal{B}} \lambda_b \leq 0\right\}.$$

Recall that $\tilde{c}^m(\theta, c; \alpha)$ *is defined in (26) and* $\bar{p}\left(c, \lambda_{P_n}, \hat{\Sigma}_n/n\right)$ *is defined in (29). Here I omit* P_n *in* $\theta_{P_n,k}$ *and* α *in* $\tilde{c}^m(\theta_\ell, \hat{c}^t_{P_n}; \alpha)$ *to simplify notation.*

Proof. There is always a subsequence $\{n_a\}$, $\{P_{n_a}, \theta_{n_a}\}$ such that

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \sup_{\theta\in[\theta_{P,\ell},\theta_{P,u}]} P\Big(\theta \notin CI^{\mathrm{m}}(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha)\Big) = \lim_{n_a} P_{n_a}\Big(\theta_{n_a} \notin CI^{\mathrm{m}}(\hat{\lambda}_{n_a}, \hat{\Sigma}_{n_a}/n_a; \alpha)\Big).$$

Since $\mathcal{S}$ defined in Assumption 3 is compact (e.g. in the Frobenius norm), and Assumption 3 implies that $\Omega(P_{n_a}) \in \mathcal{S}$ for all n_a, there exists a further subsequence

$\{n_r\} \subseteq \{n_a\}$ such that

$$\lim_{r\to\infty} \Omega\left(P_{n_r}\right) \to \Omega_0 \in \mathcal{S}.$$

Also, note that the set $[-\infty, +\infty]^{|\mathcal{B}|}$ is compact under metric $d(\lambda, \tilde{\lambda}) = \left\|\Phi(\lambda) - \Phi(\tilde{\lambda})\right\|$ for $\Phi(\cdot)$ the standard normal CDF applied elementwise, and $\|\cdot\|$ the Euclidean norm. Therefore, there is a further subsequence $\{n_s\} \subseteq \{n_r\}$ along which (78) holds. We have found a subsequence n_s such that (77) and (78) hold. Therefore,

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \sup_{\theta\in[\theta_\ell,\theta_u]} P\left(\theta \notin CI^{\mathrm{m}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) = \lim_{n_s} P_{n_s}\left(\theta_{n_s} \notin CI^{\mathrm{m}}\left(\hat{\lambda}_{n_s}, \hat{\Sigma}_{n_s}/n_s; \alpha\right)\right).$$

With slight abuse of notation, in the following equations I use n for subsequence n_s:

$$\begin{aligned}
&P_n\left(\theta_n \notin CI^{\mathrm{m}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) \\
\leq &P_n\left(\theta_n \notin CI^{\mathrm{m}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right), \lambda_{P_n} \in \hat{\Lambda}_n\right) + P_n\left(\lambda_{P_n} \notin \hat{\Lambda}_n\right) \\
\leq &\max\Big\{P_n\left([\theta_\ell, \theta_m] \not\subseteq CI^{\mathrm{m}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right), \lambda_{P_n} \in \hat{\Lambda}_n\right), \\
&\quad P_n\left([\theta_m, \theta_u] \not\subseteq CI^{\mathrm{m}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right), \lambda_{P_n} \in \hat{\Lambda}_n\right)\Big\} + P_{n_s}\left(\lambda_{P_n} \notin \hat{\Lambda}_n\right) \\
\leq &\max\Big\{P_n\left(\hat{T}(\theta_\ell) > \hat{c}^{\mathrm{m}}(\theta_\ell; \alpha) \text{ or } \left\{\hat{T}(\theta_m) > \hat{c}^{\mathrm{m}}(\theta_m; \alpha) \text{ and } \hat{T}(\theta_u) > \hat{c}^{\mathrm{m}}(\theta_u; \alpha)\right\}, \lambda_{P_n} \in \hat{\Lambda}_n\right), \\
&\quad P_n\left(\hat{T}(\theta_u) > \hat{c}^{\mathrm{m}}(\theta_u; \alpha) \text{ or } \left\{\hat{T}(\theta_m) > \hat{c}^{\mathrm{m}}(\theta_m; \alpha) \text{ and } \hat{T}(\theta_\ell) > \hat{c}^{\mathrm{m}}(\theta_\ell; \alpha)\right\}, \lambda_{P_n} \in \hat{\Lambda}_n\right)\Big\} \\
&\quad + \eta + o(1) \\
\leq &\max\left\{P_n\left(E_\ell \cup \{E_m \cap E_u\}\right), P_n\left(E_u \cup \{E_m \cap E_\ell\}\right)\right\} + \eta + o(1).
\end{aligned}$$

Recall that $\hat{c}^{\mathrm{t}}$ is defined in (30) and $\hat{c}^{\mathrm{t}}_{P_n}$ is defined in (76), thus the last inequality follows from the fact that $\hat{c}^{\mathrm{t}}_{P_n} \leq \hat{c}^{\mathrm{t}}$ if $\lambda_{P_n} \in \hat{\Lambda}_n$. Therefore it suffices to show (75). □

Lemma 4. *Under Assumptions 1, 2, 3, 4, 5, sequences (77) and (78), if*

$$\min_{b\in\mathcal{B}} \lambda_{\ell u,b} \in \mathbb{R}, \tag{79}$$

it holds that

$$\left(\hat{T}(\theta_{P_n,k}), \hat{c}^c(\theta_{P_n,k}, \alpha^c)\right)_{k=\ell,m,u} \xrightarrow{d} (T_k, c^c_k)_{k=\ell,m,u}. \tag{80}$$

If

$$\min_{b\in\mathcal{B}} \lambda_{\ell u,b} = -\infty, \tag{81}$$

it holds that

$$\left(\hat{T}(\theta_{P_n,k}), \hat{c}^c(\theta_{P_n,k}, \alpha^c)\right)_{k=\ell,u} \xrightarrow{d} (T_k, c_k^c)_{k=\ell,u}, \tag{82}$$

and for all $c \in \mathbb{R}$

$$P_n\left(\hat{T}(\theta_{P_n,m}) \geq c\right) \to 0. \tag{83}$$

Proof. Note that

$$\lim_n \sqrt{n}\left(\lambda_{P_n,\ell,b_\ell} - \lambda_{P_n,u,b_u}\right) = \lim_n \min_{b\in\mathcal{B}} \sigma_{P_n,u,b} \frac{\theta_{P_n,\ell} - \lambda_{P_n,u,b}}{\sigma_{P_n,u,b}/\sqrt{n}} = \min_{b\in\mathcal{B}} \sigma_{0,u,b}\lambda_{\ell u,b}. \tag{84}$$

Thus the two cases in (79) and (81) correspond to whether the length of the identified set of θ is large asymptotically. I will show (80) under (79) in Step 1 and 2, then show (82) and (83) under (81) in Step 3.

Step 1. Show that under (79),

$$\begin{aligned}
&\left(\hat{b}_\ell(\theta_{P_n,k}), \hat{b}_u(\theta_{P_n,k}), \hat{T}(\theta_{P_n,k}), \Phi\left(\hat{t}_\ell(\theta_{P_n,k}, \mathcal{B}_{k\ell})\right), \Phi\left(\hat{t}_u(\theta_{P_n,k}, \mathcal{B}_{ku})\right)\right)_{k=\ell,m,u} \\
&\quad \xrightarrow{d} \left(b_{k\ell}, b_{ku}, T_k, \Phi\left(t_{k\ell}(\mathcal{B}_{k\ell})\right), \Phi\left(t_{ku}(\mathcal{B}_{ku})\right)\right)_{k=\ell,m,u}
\end{aligned}$$

where $\hat{t}_\ell$, $\hat{t}_u$ is t_ℓ, t_u in Lemma 1 with Σ_n replaced with $\hat{\Sigma}_n/n$,

$$\begin{aligned}
t_{kk'}(\mathcal{B}_{kk'}) &= \left(t_{kk',1}(b),\ t_{kk',2}(b)\right)_{b\in\mathcal{B}_{kk'}} \\
\hat{t}_{k'}(\theta_{P_n,k}, \mathcal{B}_{kk'}) &= \left(\hat{t}_{k',1}(\theta_{P_n,k}, b),\ \hat{t}_{k',2}(\theta_{P_n,k}, b)\right)_{b\in\mathcal{B}_{kk'}},\ k \in \{\ell, u, m\}, k' \in \{\ell, u\}.
\end{aligned}$$

$t_{kk'}$ is defined in (41) and $\mathcal{B}_{k\ell}$, $\mathcal{B}_{kl}$ are defined in (40).

Step 1.1. Note that

$$\begin{aligned}
\hat{T}(\theta_{P_n,k}) &= \max\left\{\min_{b\in\mathcal{B}} \frac{\hat{\lambda}_{\ell,b} - \theta_{P_n,k}}{\hat{\sigma}_{\ell,b}/\sqrt{n}},\ \min_{b\in\mathcal{B}} \frac{\theta_{P_n,k} - \hat{\lambda}_{u,b}}{\hat{\sigma}_{u,b}/\sqrt{n}}\right\} \\
&= \max\left\{\min_{b\in\mathcal{B}_\ell} \frac{\hat{\lambda}_{\ell,b} - \lambda_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}/\sqrt{n}} + \frac{\sigma_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}}\lambda_{n,k\ell,b},\ \min_{b\in\mathcal{B}_u} \frac{\lambda_{P_n,u,b} - \hat{\lambda}_{u,b}}{\hat{\sigma}_{u,b}/\sqrt{n}} + \frac{\sigma_{P_n,u,b}}{\hat{\sigma}_{u,b}}\lambda_{n,ku,b}\right\} \\
&\overset{\text{w.p.a. 1}}{=} \max\left\{\min_{b\in\mathcal{B}_{k\ell}} \frac{\hat{\lambda}_{\ell,b} - \lambda_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}/\sqrt{n}} + \frac{\sigma_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}}\lambda_{n,k\ell,b},\ \min_{b\in\mathcal{B}_{ku}} \frac{\lambda_{P_n,u,b} - \hat{\lambda}_{u,b}}{\hat{\sigma}_{u,b}/\sqrt{n}} + \frac{\sigma_{P_n,u,b}}{\hat{\sigma}_{u,b}}\lambda_{n,ku,b}\right\}
\end{aligned}$$

$$\xrightarrow{d} \max\left\{\min_{b\in\mathcal{B}_{k\ell}} Z_{\ell,b}+\lambda_{k\ell,b},\ \min_{b\in\mathcal{B}_{ku}} Z_{u,b}+\lambda_{ku,b}\right\}. \tag{85}$$

The first line is by definition, the second line simply rearranges terms with $\lambda_{n,k\ell,b}$, $\lambda_{n,ku,b}$ defined in (78). To see the third line, note that by Assumption 1, 2, 3, 4, we have

$$\left(\left(\frac{\hat{\lambda}_{\ell,b}-\lambda_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}/\sqrt{n}}\right)_{b\in\mathcal{B}},\left(\frac{\lambda_{P_n,u,b}-\hat{\lambda}_{u,b}}{\hat{\sigma}_{u,b}/\sqrt{n}}\right)_{b\in\mathcal{B}},\left(\frac{\sigma_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}}\right)_{b\in\mathcal{B}},\left(\frac{\sigma_{P_n,u,b}}{\hat{\sigma}_{u,b}}\right)_{b\in\mathcal{B}}\right) \tag{86}$$
$$\xrightarrow{d}\left(Z_\ell, Z_u, \mathbf{1}_{2|\mathcal{B}|}\right).$$

In addition, for $b\in\mathcal{B}_\ell\backslash\mathcal{B}_{k\ell}$, $\lambda_{k\ell,b}=\infty$, thus with probability going to one,

$$\min_{\tilde b\in\mathcal{B}}\frac{\hat{\lambda}_{\ell,\tilde b}-\theta_{P_n,k}}{\hat{\sigma}_{\ell,\tilde b}/\sqrt{n}}\le\frac{\hat{\lambda}_{\ell,b_\ell}-\theta_{P_n,k}}{\hat{\sigma}_{\ell,b_\ell}/\sqrt{n}}\le\frac{\hat{\lambda}_{\ell,b_\ell}-\lambda_{P_n,\ell,b_\ell}}{\hat{\sigma}_{\ell,b_\ell}/\sqrt{n}}<\frac{\hat{\lambda}_{\ell,b}-\lambda_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}/\sqrt{n}}+\frac{\sigma_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}}\lambda_{n,k\ell,b},$$

$\lambda_{n,k\ell,b}$ is defined in (78). Thus asymptotically, we can ignore $\mathcal{B}_\ell\backslash\mathcal{B}_{k\ell}$. With the same argument, we can replace $\mathcal{B}_u$ with $\mathcal{B}_{ku}$ in the second part. The fourth line follows from (i) (86), (ii) Slustsky's Lemma and (iii) the limit distribution is well defined because

$$\lambda_{k\ell,b}=\lim_n\frac{\lambda_{P_n,\ell,b}-\theta_{P_n,k}}{\sigma_{P_n,\ell,b}/\sqrt{n}}\ge\lim_n\frac{\lambda_{P_n,\ell,b_\ell}-\lambda_{P_n,u,b_u}}{\sigma_{P_n,\ell,b}/\sqrt{n}}=\frac{\min_{b\in\mathcal{B}}\sigma_{0,u,b}\lambda_{\ell u,b}}{\sigma_{0,\ell,b}}\in\mathbb{R}, \tag{87}$$
$$\lambda_{ku,b}=\lim_n\frac{\theta_{P_n,k}-\lambda_{P_n,u,b}}{\sigma_{P_n,u,b}/\sqrt{n}}\ge\lim_n\frac{\lambda_{P_n,\ell,b_\ell}-\lambda_{P_n,u,b_u}}{\sigma_{P_n,u,b}/\sqrt{n}}=\frac{\min_{b\in\mathcal{B}}\sigma_{0,u,b}\lambda_{\ell u,b}}{\sigma_{0,u,b}}\in\mathbb{R}.$$

Step 1.2. As for $\Phi\left(\hat{t}_{\ell,1}(\theta_k,\mathcal{B}_{k\ell})\right)$, let $b\in\mathcal{B}_{k\ell}$. If $\min_{\tilde b\in\mathcal{B}}\rho_{\ell u}(b,\tilde b)=-1$, then $\Phi\left(t_{k\ell,1}(b)\right)=0$ by construction in (41). Note that $\min_{b\in\mathcal{B}}\rho_{\ell u}(b,\tilde b)=-1$ implies $A_{\ell,b}=-aA_{u,\tilde b}$ for some $a>0$, thus $\min_{b\in\mathcal{B}}\hat\rho_{\ell u}(b,\tilde b)=-1$ for all samples, thus with probability one, $\Phi\left(\hat{t}_{\ell,1}(\theta_k,b\right)=0$, and the convergence is trivial. Then consider $\min_{\tilde b\in\mathcal{B}}\rho_{\ell u}(b,\tilde b)>-1$. By (87) and the definition of $\mathcal{B}_{k\ell}$, we have $\lambda_{k\ell,b}\in\mathbb{R}$, and thus

$$\begin{aligned}
&\Phi\left(t_{\ell,1}(\theta_k,b)\right)\\
=&\min_{\tilde b\in\mathcal{B}}\Phi\left(\left(1+\hat\rho_{\ell u}(b,\tilde b)\right)^{-1}\left(\hat Z_{u,\tilde b}+\frac{\sigma_{u,\tilde b}}{\hat\sigma_{u,\tilde b}}\lambda_{P_n,ku,\tilde b}+\hat\rho_{\ell u}(b,\tilde b)\left(\hat Z_{\ell,b}+\frac{\sigma_{\ell,b}}{\hat\sigma_{\ell,b}}\lambda_{P_n,k\ell,b}\right)\right)\right)\\
\xrightarrow{d}&\min_{\tilde b\in\mathcal{B}}\Phi\left(\left(1+\rho_{\ell u}(b,\tilde b)\right)^{-1}\left(Z_{u,\tilde b}+\lambda_{ku,\tilde b}+\rho_{\ell u}(b,\tilde b)\left(Z_{\ell,b}+\lambda_{k\ell,b}\right)\right)\right)
\end{aligned}$$

$$= \Phi\left(t_{k\ell,1}(b)\right).$$

Thus

$$\Phi\left(t_{\ell,1}(\theta_k, \mathcal{B}_{k\ell})\right) \xrightarrow{d} \Phi\left(t_{k\ell,1}(\mathcal{B}_{k\ell})\right). \tag{88}$$

This argument also applies to $\Phi\left(t_{\ell,2}(\theta_k, \mathcal{B}_{k\ell})\right)$, $\Phi\left(t_{u,1}(\theta_k, \mathcal{B}_{ku})\right)$, $\Phi\left(t_{u,2}(\theta_k, \mathcal{B}_{ku})\right)$.

Step 1.3. Let

$$g\left(X, Y\right) = \mathbf{1}\left[X \leq Y\right].$$

For $b_1, b_2 \in \mathcal{B}_{k\ell}$, $b_1 \neq b_2$,

$$P\left(Z_{\ell,b_1} + \lambda_{k\ell,b_1} = Z_{\ell,b_2} + \lambda_{k\ell,b_2}\right) = P\left((A_{\ell,b_1} - A_{\ell,b_2})Z_\delta = \lambda_{k\ell,b_2} - \lambda_{k\ell,b_1}\right) = 0,$$

following from $A_{\ell,b_1} \neq A_{\ell,b_2}$, $Z_\delta \sim \mathcal{N}(0, \Omega_0)$, Ω_0 non-singular and $\lambda_{k\ell,b_2}, \lambda_{k\ell,b_1} \in \mathbb{R}$. Thus

$$g\left(Z_{\ell,b_1} + \lambda_{k\ell,b_1}, Z_{\ell,b_2} + \lambda_{k\ell,b_2}\right)$$

is almost sure continuous. Thus by continuous mapping theorem, it holds that

$$g\left(\mathcal{Z}_{\ell,b_1}, \mathcal{Z}_{\ell,b_2}\right) \xrightarrow{d} g\left(Z_{\ell,b_1} + \lambda_{k\ell,b_1}, Z_{\ell,b_2} + \lambda_{k\ell,b_2}\right). \tag{89}$$

Similarly, we have

$$g\left(\mathcal{Z}_{u,b_2}, \mathcal{Z}_{\ell,b_1}\right) \xrightarrow{d} g\left(Z_{u,b_2} + \lambda_{ku,b_2}, Z_{\ell,b_1} + \lambda_{k\ell,b_1}\right). \tag{90}$$

Then consider $b_1 \in \mathcal{B}_{k\ell}$ and $b_2 \in \mathcal{B}_{ku}$. (i) If $A_{\ell,b_1} \neq A_{u,b_2}$, similar argument holds,

$$g\left(\mathcal{Z}_{\ell,b_1}, \mathcal{Z}_{u,b_2}\right) \xrightarrow{d} g\left(Z_{\ell,b_1} + \lambda_{k\ell,b_1}, Z_{u,b_2} + \lambda_{ku,b_2}\right); \tag{91}$$

(ii) if $A_{\ell,b_1} = A_{u,b_2}$, then

$$g\left(\mathcal{Z}_{\ell,b_1}, \mathcal{Z}_{u,b_2}\right) = g\left(Z_{\ell,b_1} + \lambda_{k\ell,b_1}, Z_{u,b_2} + \lambda_{ku,b_2}\right) = 1$$

for all samples, thus the convergence holds trivially. The convergence in (85), (88), (89), (90), (91) holds jointly.

Step 2. Then I show the convergence of $\Phi\left(\hat{c}^c(\theta_k, \alpha)\right)$.

$$\Phi\left(\hat{c}^c(\theta_k, \alpha)\right)$$

$$
\begin{aligned}
=&\left((1-\alpha)\Phi\left(t_{\ell,2}(\theta_k,\hat{b}_\ell)\right)+\alpha\Phi\left(t_{\ell,1}(\theta_k,\hat{b}_\ell)\right)\right)\mathbf{1}\left[\mathcal{Z}_{\ell,\hat{b}_\ell}\geq\mathcal{Z}_{u,\hat{b}_u}\right]\\
&+\left((1-\alpha)\Phi\left(t_{u,2}(\theta_k,\hat{b}_u)\right)+\alpha\Phi\left(t_{u,1}(\theta_k,\hat{b}_u)\right)\right)\mathbf{1}\left[\mathcal{Z}_{\ell,\hat{b}_\ell}<\mathcal{Z}_{u,\hat{b}_u}\right]\\
=&\sum_{b_1\in\mathcal{B}_{k\ell}}\sum_{b_2\in\mathcal{B}_{ku}}\mathbf{1}\left[\mathcal{Z}_{\ell,b_1}\geq\mathcal{Z}_{u,b_2},\mathcal{Z}_{\ell,b_1}\leq\mathcal{Z}_{\ell,\mathcal{B}_{k\ell}\backslash b_1},\mathcal{Z}_{u,b_2}\leq\mathcal{Z}_{u,\mathcal{B}_{ku}\backslash b_2}\right]\times\\
&\qquad\qquad((1-\alpha)\Phi\left(t_{\ell,2}(\theta_k,b_1)\right)+\alpha\Phi\left(t_{\ell,1}(\theta_k,b_1)\right))\\
&+\sum_{b_1\in\mathcal{B}_{k\ell}}\sum_{b_2\in\mathcal{B}_{ku}}\mathbf{1}\left[\mathcal{Z}_{\ell,b_1}<\mathcal{Z}_{u,b_2},\mathcal{Z}_{\ell,b_1}\leq\mathcal{Z}_{\ell,\mathcal{B}_{k\ell}\backslash b_1},\mathcal{Z}_{u,b_2}\leq\mathcal{Z}_{u,\mathcal{B}_{ku}\backslash b_2}\right]\times\\
&\qquad\qquad((1-\alpha)\Phi\left(t_{u,2}(\theta_k,b_2)\right)+\alpha\Phi\left(t_{u,1}(\theta_k,b_2)\right))\ \text{ w.p.a. } 1
\end{aligned}
$$

In addition,

$$
\begin{aligned}
&\mathbf{1}\left[\mathcal{Z}_{\ell,b_1}\geq\mathcal{Z}_{u,b_2},\mathcal{Z}_{\ell,b_1}\leq\mathcal{Z}_{\ell,\mathcal{B}_{k\ell}\backslash b_1},\mathcal{Z}_{u,b_2}\leq\mathcal{Z}_{u,\mathcal{B}_{ku}\backslash b_2}\right]\\
=&g\left(\mathcal{Z}_{u,b_2},\mathcal{Z}_{\ell,b_1}\right)\prod_{\tilde{b}_1\in\mathcal{B}_{k\ell}\backslash b_1}g\left(\mathcal{Z}_{\ell,b_1},\mathcal{Z}_{\ell,\tilde{b}_1}\right)\prod_{\tilde{b}_2\in\mathcal{B}_{ku}\backslash b_2}g\left(\mathcal{Z}_{u,b_2},\mathcal{Z}_{u,\tilde{b}_2}\right)\\
&\mathbf{1}\left[\mathcal{Z}_{\ell,b_1}<\mathcal{Z}_{u,b_2},\mathcal{Z}_{\ell,b_1}\leq\mathcal{Z}_{\ell,\mathcal{B}_{k\ell}\backslash b_1},\mathcal{Z}_{u,b_2}\leq\mathcal{Z}_{u,\mathcal{B}_{ku}\backslash b_2}\right]\\
=&\left[1-g\left(\mathcal{Z}_{u,b_2},\mathcal{Z}_{\ell,b_1}\right)\right]\prod_{\tilde{b}_1\in\mathcal{B}_{k\ell}\backslash b_1}g\left(\mathcal{Z}_{\ell,b_1},\mathcal{Z}_{\ell,\tilde{b}_1}\right)\prod_{\tilde{b}_2\in\mathcal{B}_{ku}\backslash b_2}g\left(\mathcal{Z}_{u,b_2},\mathcal{Z}_{u,\tilde{b}_2}\right)
\end{aligned}
$$

Since all function are almost sure continuous as discussed before, we have

$$
\Phi\left(\hat{c}^c(\theta_k,\alpha)\right)\xrightarrow{d}\Phi\left(c_k^c(\alpha)\right)
$$

following from (88), (89), (90), (91).

Step 3. Now assume (81) holds. We can show that

$$
\begin{aligned}
&\left(\hat{b}_\ell(\theta_{P_n,k}),\hat{b}_u(\theta_{P_n,k}),\hat{T}(\theta_{P_n,k}),\Phi\left(\hat{t}_\ell(\theta_{P_n,k},\mathcal{B}_{k\ell})\right),\Phi\left(\hat{t}_u(\theta_{P_n,k},\mathcal{B}_{ku})\right)\right)_{k=\ell,u}\\
&\quad\xrightarrow{d}\left(b_{k\ell},b_{ku},T_k,\Phi\left(t_{k\ell}(\mathcal{B}_{k\ell})\right),\Phi\left(t_{ku}(\mathcal{B}_{ku})\right)\right)_{k=\ell,u}
\end{aligned}
$$

with similar argument as Step 1 and 2. Regarding (83), note that

$$
\begin{aligned}
\hat{T}(\theta_m)&=\max\left\{\min_{b\in\mathcal{B}}\frac{\hat{\lambda}_{\ell,b}-\theta_m}{\hat{\sigma}_{\ell,b}/\sqrt{n}},\ \min_{b\in\mathcal{B}}\frac{\theta_m-\hat{\lambda}_{u,b}}{\hat{\sigma}_{u,b}/\sqrt{n}}\right\}\\
&\leq\max\left\{\frac{\hat{\lambda}_{\ell,b_\ell}-\theta_m}{\hat{\sigma}_{\ell,b_\ell}/\sqrt{n}},\ \frac{\theta_m-\hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}}\right\}
\end{aligned}
$$

$$= \max\left\{ \frac{\hat{\lambda}_{\ell,b_\ell} - \lambda_{\ell,b_\ell}}{\hat{\sigma}_{\ell,b_\ell}/\sqrt{n}} + \frac{\sigma_{\ell,b_\ell}}{\hat{\sigma}_{\ell,b_\ell}} \frac{\lambda_{\ell,b_\ell} - \theta_m}{\sigma_{\ell,b_\ell}/\sqrt{n}},\ \frac{\lambda_{u,b_u} - \hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b}/\sqrt{n}} + \frac{\sigma_{u,b_u}}{\hat{\sigma}_{u,b_u}} \frac{\theta_m - \lambda_{u,b}}{\sigma_{u,b_u}/\sqrt{n}} \right\}.$$

By (81), (84) and $\theta_m = (\theta_\ell + \theta_u)/2$,

$$\lim_n \frac{\lambda_{\ell,b_\ell} - \theta_m}{\sigma_{\ell,b_\ell}/\sqrt{n}} = -\infty,\ \lim_n \frac{\theta_m - \lambda_{u,b}}{\sigma_{u,b_u}/\sqrt{n}} = -\infty.$$

Thus it is easy to see (83) holds. □

Lemma 5. *Assume that (i) $\theta \le \theta_\ell$; (ii) $\eta \in [0, \alpha/4)$,*

$$\hat{T}(\theta) = \mathcal{Z}_{\ell,\hat{b}_\ell} > \Phi^{-1}(1 - \frac{\alpha - \eta}{2}); \tag{92}$$

(iii) either

$$\min_{\tilde{b}\in\mathcal{B}} \hat{\rho}_{\ell u}(\hat{b}_\ell, \tilde{b}) = -1,$$

or

$$\left(1 + \hat{\rho}_{\ell u}(\hat{b}_\ell, b_u)\right)^{-1} \frac{\lambda_{u,b_u} - \hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}} \le \bar{M}, \tag{93}$$

$$\hat{T}(\theta) > \bar{z}, \tag{94}$$

where $\bar{M} \in \mathbb{R}$, $\bar{z}$ is defined in Lemma 6 with $\bar{M}$ given in (93). Then

$$\hat{T}(\theta) > \hat{c}^m(\theta, \alpha^c). \tag{95}$$

Proof. Note that $\hat{c}^{\text{t}} \le \Phi^{-1}(1 - \frac{\alpha-\eta}{2})$ by construction, thus under (92), $\hat{T}(\theta) > \hat{c}^{\text{t}}$ and (95) is equivalent to

$$\hat{T}(\theta) > \hat{c}^{\text{c}}(\theta, \alpha^c). \tag{96}$$

If $\min_{\tilde{b}\in\mathcal{B}} \hat{\rho}_{\ell u}(\hat{b}_\ell, \tilde{b}) = -1$, then

$$\hat{c}^{\text{c}}(\theta, \alpha^c) = \Phi^{-1}\left((1 - \alpha^c)\Phi\left(t_{\ell,2}(\theta, \hat{b}_\ell)\right)\right) \le \Phi^{-1}(1 - \alpha^c) < \Phi^{-1}(1 - \frac{\alpha}{2}).$$

In this case, (96) holds trivially. If $\min_{\tilde{b}\in\mathcal{B}} \rho_{\ell u}(\hat{b}_\ell, \tilde{b}) > -1$, we have

$$t_{\ell,1}(\theta, \hat{b}_\ell) = \min_{\tilde{b}\in\mathcal{B}} \left(1 + \hat{\rho}_{\ell u}(\hat{b}_\ell, \tilde{b})\right)^{-1} \left(\mathcal{Z}_{u,\tilde{b}} + \hat{\rho}_{\ell u}(\hat{b}_\ell, \tilde{b})\mathcal{Z}_{\ell,\hat{b}_\ell}\right)$$

$$
\begin{aligned}
&\leq \left(1+\hat{\rho}_{\ell u}(\hat{b}_\ell, b_u)\right)^{-1}\left(\frac{\theta-\hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}}+\hat{\rho}_{\ell u}(\hat{b}_\ell, b_u)\mathcal{Z}_{\ell,\hat{b}_\ell}\right)\\
&\leq \left(1+\hat{\rho}_{\ell u}(\hat{b}_\ell, b_u)\right)^{-1}\left(\frac{\lambda_{u,b_u}-\hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}}+\hat{\rho}_{\ell u}(\hat{b}_\ell, b_u)\mathcal{Z}_{\ell,\hat{b}_\ell}\right)\\
&\leq \bar{M}+\frac{1}{2}\mathcal{Z}_{\ell,\hat{b}_\ell},
\end{aligned}
$$

where the second inequality uses $\theta \leq \theta_\ell \leq \lambda_{u,b_u}$ by (i). Then

$$
\begin{aligned}
&\Phi\left(\hat{T}(\theta)\right)-\Phi\left(\hat{c}^{\mathrm{c}}(\theta,\alpha^{\mathrm{c}})\right)\\
=&\Phi\left(\hat{T}(\theta)\right)-\alpha^{\mathrm{c}}\Phi\left(t_{\ell,1}(\theta,\hat{b}_\ell)\right)-(1-\alpha^{\mathrm{c}})\Phi\left(t_{\ell,2}(\theta,\hat{b}_\ell)\right)\\
\geq&\Phi\left(\mathcal{Z}_{\ell,\hat{b}_\ell}\right)-\alpha^{\mathrm{c}}\Phi\left(\bar{M}+\frac{1}{2}\mathcal{Z}_{\ell,\hat{b}_\ell}\right)-(1-\alpha^{\mathrm{c}})\\
=&H\left(\mathcal{Z}_{\ell,\hat{b}_\ell},\bar{M}\right)>0,
\end{aligned}
$$

with $H(z,\bar{M})$ defined in Lemma 6. $H(\mathcal{Z}_{\ell,\hat{b}_\ell},\bar{M})>0$ follows from (94) and Lemma 6. ☐

Lemma 6. *Let*

$$
H(z,\bar{M})=\Phi(z)-\alpha^c\Phi\left(\bar{M}+\frac{1}{2}z\right)-(1-\alpha^c).
$$

For all $\bar{M}\in\mathbb{R}$, there is some $\bar{z}_M\in\mathbb{R}$ such that $H(z,\bar{M})>0$ for all $z\geq\bar{z}_M$.

Proof. Note that

$$
\frac{\mathrm{d}H(z,\bar{M})}{\mathrm{d}z}=\phi(z)\left(1-\frac{\alpha}{2}\exp\left(\frac{3}{8}z^2-\frac{\bar{M}}{2}z-\frac{\bar{M}^2}{2}\right)\right),
$$

and thus there is $\bar{z}_M\in\mathbb{R}$ such that $\frac{\mathrm{d}H(z,\bar{M})}{\mathrm{d}z}<0$ for all $z\geq\bar{z}_M$. Also note that

$$
\lim_{z\to\infty}H(z,\bar{M})=0.
$$

Therefore, for all $z\geq\bar{z}_M$, we have $H(z,\bar{M})>0$. ☐

Lemma 7. *Let $\alpha\in(0,\frac{1}{2})$, $\alpha^c\in(\frac{\alpha}{2},\alpha)$, $\eta\in\left[0,\frac{\alpha}{4}\right)$. Recall that $c^{sim}=\Phi^{-1}(1-\frac{\alpha}{2})$.*

Let

$$H\left(c,\Delta,\rho\right)=\Phi_2\left(-c,\Delta-c;\rho\right)+\Phi\left(-\frac{\Delta}{2}-c\right), \tag{97}$$

$$\rho_2^*(\alpha,\eta)=\sup_{\rho\in(-1,1)}\left\{\rho:\sup_{\Delta\geq0}H\left(c^{sim},\Delta,\rho\right)\leq\alpha-\eta\right\}. \tag{98}$$

For all $\xi>0$*, there is* $\bar{c}<c^{sim}$ *such that*

$$\sup_{\rho\leq\rho_2^*(\alpha,\eta)-\xi}\sup_{\Delta\geq0}H\left(\bar{c},\Delta,\rho\right)<\alpha-\eta. \tag{99}$$

Proof. First, we can check numerically that for $\alpha\in(0,\frac{1}{2})$,

$$\sup_{\Delta\geq0}H\left(c^{\mathrm{sim}},\Delta,0\right)=\sup_{\Delta\geq0}\frac{\alpha}{2}\Phi\left(\Delta-c^{\mathrm{sim}}\right)+\Phi\left(-\frac{\Delta}{2}-c^{\mathrm{sim}}\right)<\frac{3}{4}\alpha<\alpha-\eta$$

and thus $\rho_2^*(\alpha,\eta)$ is well defined.

Second, I show that for all $c\in(0,c^{\mathrm{sim}}]$, it holds that for all $|\rho|<1$,

$$\sup_{\Delta\geq0}H\left(c,\Delta,\rho\right)=\sup_{\Delta\in[0,\bar{\Delta}]}H\left(c,\Delta,\rho\right) \tag{100}$$

where $\bar{\Delta}=2c^{\mathrm{sim}}+\sqrt{4(c^{\mathrm{sim}})^2+8/3\log(2)}$. The first order derivative gives that for all $\Delta>\bar{\Delta}$,

$$\begin{aligned}\frac{\mathrm{d}H\left(c,\Delta,\rho\right)}{\mathrm{d}\Delta}&=\phi(\Delta-c)\left[\Phi\left(\frac{(\rho-1)c-\rho\Delta}{\sqrt{1-\rho^2}}\right)-\frac{1}{2}\exp\left(\frac{3}{8}\Delta(\Delta-4c)\right)\right]\\&\leq\phi(\Delta-c)\left[1-\frac{1}{2}\exp\left(\frac{3}{8}\Delta(\Delta-4c^{\mathrm{sim}})\right)\right]\leq0.\end{aligned}$$

Therefore, (100) holds for all $c\in(0,c^{\mathrm{sim}}]$.

Third, let $\bar{\rho}=\rho_2^*(\alpha,\eta)-\xi$, and by construction,

$$\begin{aligned}\alpha-\eta&\geq\sup_{\Delta\in[0,\bar{\Delta}]}H\left(c^{\mathrm{sim}},\Delta,\rho_2^*(\alpha,\eta)\right)\\&=\sup_{\Delta\in[0,\bar{\Delta}]}H\left(c^{\mathrm{sim}},\Delta,\bar{\rho}\right)+\frac{\mathrm{d}H\left(c^{\mathrm{sim}},\Delta,\tilde{\rho}(\Delta)\right)}{\mathrm{d}\rho}\xi\\&\geq\sup_{\Delta\in[0,\bar{\Delta}]}H\left(c^{\mathrm{sim}},\Delta,\bar{\rho}\right)+a\xi\end{aligned} \tag{101}$$

where

$$a = \inf_{\substack{\Delta \in [0,\bar{\Delta}] \\ \tilde{\rho} \in [\bar{\rho}, \rho_2^*(\alpha,\eta)]}} \frac{\mathrm{d}H\left(c^{\mathrm{sim}}, \Delta, \tilde{\rho}\right)}{\mathrm{d}\rho} = \inf_{\substack{\Delta \in [0,\bar{\Delta}] \\ \tilde{\rho} \in [\bar{\rho}, \rho_2^*(\alpha,\eta)]}} \phi\left(-c^{\mathrm{sim}}, \Delta - c^{\mathrm{sim}}; \tilde{\rho}\right) > 0.$$

Rewrite (101) we get

$$\sup_{\Delta\in[0,\bar{\Delta}]} H\left(c^{\mathrm{sim}}, \Delta, \bar{\rho}\right) \leq \alpha - \eta - a\xi.$$

Lastly,

$$\frac{\mathrm{d}H\left(c, \Delta, \rho\right)}{\mathrm{d}c} = -\phi(\Delta-c)\Phi\left(\frac{(\rho-1)c - \Delta\rho}{\sqrt{1-\rho^2}}\right) - \phi(-c)\Phi\left(\frac{c\rho - c + \Delta}{\sqrt{1-\rho^2}}\right) - \phi(-\frac{\Delta}{2} - c).$$

Let

$$b = -\inf_{\rho\in\left[0,\rho_2^*(\alpha,\eta)\right], c\in[0,c^{\mathrm{sim}}], \Delta\in[0,\bar{\Delta}]} \frac{\mathrm{d}H\left(c, \Delta, \rho\right)}{\mathrm{d}c} > 0.$$

Choose $\bar{c} = c^{\mathrm{sim}} - \frac{a\xi}{2b}$, and then for all $\rho \leq \rho_2^*(\alpha,\eta) - \xi$,

$$\begin{aligned} \sup_{\Delta\in[0,\bar{\Delta}]} H\left(c^{\mathrm{sim}}, \Delta, \rho\right) &= \sup_{\Delta\in[0,\bar{\Delta}]} H\left(\bar{c}, \Delta, \rho\right) + \frac{\mathrm{d}H\left(\tilde{c}(\Delta), \Delta, \rho\right)}{\mathrm{d}c}(c^{\mathrm{sim}} - \bar{c}) \\ &\geq \sup_{\Delta\in[0,\bar{\Delta}]} H\left(\bar{c}, \Delta, \rho\right) - b(c^{\mathrm{sim}} - \bar{c}). \end{aligned} \tag{102}$$

In sum, for all $\rho \leq \rho_2^*(\alpha,\eta) - \xi$, by (100), (101), (102),

$$\sup_{\Delta\geq 0} H\left(\bar{c}, \Delta, \rho\right) = \sup_{\Delta\in[0,\bar{\Delta}]} H\left(\bar{c}, \Delta, \rho\right) \leq \alpha - \eta - a\xi + b\frac{a\xi}{2b} < \alpha - \eta.$$

□

Lemma 8. *Suppose Assumptions 1, 2, 3, 4, and 5 hold. Let* $\alpha \in (0, \frac{1}{2})$, $\alpha^c \in (\frac{\alpha}{2}, \alpha)$, $\eta \in \left[0, \frac{\alpha-\alpha^c}{2}\right)$. *Assume that* $A_\ell = A_u$, *and* $\mathcal{P}$ *satisfies that*

$$\sup_{P\in\mathcal{P}} \rho_\ell(b_\ell, b_u) < \rho_2^*(\alpha,\eta), \tag{103}$$

where $\rho_2^*(\alpha,\eta)$ *is defined in Lemma 7 equation (98). Then there is* $\alpha' > \alpha$ *such* *that*

$$\liminf_n \inf_{P\in\mathcal{P}} P\left(\hat{c}^t \leq \Phi^{-1}(1-\frac{\alpha'}{2})\right) = 1. \tag{104}$$

Proof. Let

$$\xi = \frac{1}{2}\left(\rho_2^*(\alpha,\eta) - \sup_{P\in\mathcal{P}} \rho_\ell(b_\ell, b_u)\right) > 0,$$

and it is easy to see that $\eta < \frac{\alpha-\alpha^{c}}{2} < \frac{\alpha}{4}$. Therefore, by Lemma 7, there is $\bar{c} < \Phi^{-1}(1-\frac{\alpha}{2})$ such that (99) holds. To show (104), note that

$$\begin{aligned}\liminf_n \inf_{P\in\mathcal{P}} P\left(\hat{c}^{\mathrm{t}} \leq \bar{c}\right) &\geq \liminf_n \inf_{P\in\mathcal{P}} P\left(\sup_{\lambda\in\hat{\Lambda}} \bar{p}(\bar{c},\lambda,\hat{\Sigma}_n/n) \leq \alpha-\eta\right)\\ &\geq \liminf_n \inf_{P\in\mathcal{P}} P\left(\sup_{\lambda\in\Lambda} \bar{p}(\bar{c},\lambda,\hat{\Sigma}_n/n) \leq \alpha-\eta\right)\end{aligned}$$

Recall that

$$\begin{aligned}&\bar{p}(\bar{c},\lambda,\hat{\Sigma}_n/n)\\ =&\max\Big\{P\left(\hat{T}(\theta_\ell) > \tilde{c}^{\mathrm{m}}(\theta_\ell,\bar{c}) \vee \left\{\hat{T}(\theta_m) > \tilde{c}^{\mathrm{m}}(\theta_m,\bar{c}) \wedge \hat{T}(\theta_u) > \tilde{c}^{\mathrm{m}}(\theta_u,\bar{c})\right\};\ \mathcal{N}(\lambda,\hat{\Sigma}_n/n)\right)\\ &\qquad P\left(\hat{T}(\theta_u) > \tilde{c}^{\mathrm{m}}(\theta_u,\bar{c}) \vee \left\{\hat{T}(\theta_m) > \tilde{c}^{\mathrm{m}}(\theta_m,\bar{c}) \wedge \hat{T}(\theta_\ell) > \tilde{c}^{\mathrm{m}}(\theta_\ell,\bar{c})\right\};\ \mathcal{N}(\lambda,\hat{\Sigma}_n/n)\right)\Big\}\\ \leq&\max\Big\{P\left(\hat{T}(\theta_\ell) > \bar{c} \text{ or } \hat{T}(\theta_m) > \bar{c};\ \mathcal{N}(\lambda,\hat{\Sigma}_n/n)\right),\\ &\qquad P\left(\hat{T}(\theta_m) > \bar{c} \text{ or } \hat{T}(\theta_u) > \bar{c};\ \mathcal{N}(\lambda,\hat{\Sigma}_n/n)\right)\Big\}.\end{aligned}$$

Thus it suffices to show that

$$\sup_{\lambda\in\Lambda} P\left(\hat{T}(\theta_\ell) > \bar{c} \text{ or } \hat{T}(\theta_m) > \bar{c};\ \mathcal{N}(\lambda,\hat{\Sigma}_n/n)\right) \leq \alpha-\eta \text{ w.p.a. } 1, \tag{105}$$

The proof for the upper bound is similar. To see (105),

$$\begin{aligned}&P\left(\hat{T}(\theta_\ell) > \bar{c} \text{ or } \hat{T}(\theta_m) > \bar{c};\ \mathcal{N}(\lambda,\hat{\Sigma}_n/n)\right)\\ \leq&P\left(\max\left\{\min_{b\in\mathcal{B}} \frac{\hat{\lambda}_b-\theta_\ell}{\hat{\sigma}_b/\sqrt{n}}, \min_{b\in\mathcal{B}} \frac{\theta_m-\hat{\lambda}_b}{\hat{\sigma}_b/\sqrt{n}}\right\} > \bar{c};\ \mathcal{N}(\lambda,\hat{\Sigma}_n/n)\right)\\ \leq&P\left(\min\left\{\frac{\hat{\lambda}_{b_\ell}-\theta_\ell}{\hat{\sigma}_{b_\ell}/\sqrt{n}}, \frac{\hat{\lambda}_{b_u}-\theta_\ell}{\hat{\sigma}_{b_u}/\sqrt{n}}\right\} > \bar{c}, \text{ or } \frac{\theta_m-\hat{\lambda}_{b_u}}{\hat{\sigma}_{b_u}/\sqrt{n}} > \bar{c};\ \mathcal{N}(\lambda,\hat{\Sigma}_n/n)\right)\\ =&P\left(\min\left\{\mathbb{Z}_{b_\ell}, \mathbb{Z}_{b_u} + \frac{\theta_u-\theta_\ell}{\hat{\sigma}_{b_u}/\sqrt{n}}\right\} > \bar{c}, \text{ or } \frac{\theta_m-\theta_u}{\hat{\sigma}_{b_u}/\sqrt{n}} - \mathbb{Z}_{b_u} > \bar{c}\right)\end{aligned}$$

$$
\begin{aligned}
&\leq P\left(\mathbb{Z}_{b_\ell} > \bar{c},\ \mathbb{Z}_{b_u} + \frac{\theta_u - \theta_\ell}{\hat{\sigma}_{b_u}/\sqrt{n}} > \bar{c};\right) + P\left(\frac{\theta_m - \theta_u}{\hat{\sigma}_{b_u}/\sqrt{n}} - \mathbb{Z}_{b_u} > \bar{c}\right)\\
&=\Phi\left(-\bar{c}, \frac{\theta_u - \theta_\ell}{\hat{\sigma}_{b_u}/\sqrt{n}} - \bar{c}; \hat{\rho}_{\ell u}(b_\ell, b_u)\right) + \Phi\left(\frac{\theta_m - \theta_u}{\hat{\sigma}_{b_u}/\sqrt{n}} - \bar{c}\right)\\
&\leq \Phi\left(-\bar{c}, \Delta - \bar{c}; \hat{\rho}_{\ell u}(b_\ell, b_u)\right) + \Phi\left(-\frac{\Delta}{2} - \bar{c}\right)\\
&= H\left(\bar{c}, \Delta, \hat{\rho}_\ell(b_\ell, b_u)\right) \qquad (106)
\end{aligned}
$$

where $(\mathbb{Z}_{b_u}, \mathbb{Z}_{b_\ell}) \sim \mathcal{N}(0, [1, \hat{\rho}_\ell(b_\ell, b_u)); \hat{\rho}_\ell(b_\ell, b_u)), 1]$, $\Delta = \frac{\theta_u - \theta_\ell}{\hat{\sigma}_{b_u}/\sqrt{n}} \geq 0$, and $H(c, \Delta, \rho)$ is in (97).

Under (103) and Assumptions 1, 2, 3, 4, and 5, it holds that

$$\hat{\rho}_\ell(b_\ell, b_u) \leq \rho_2^*(\alpha, \eta) - \xi \text{ w.p.a. } 1.$$

Thus (106) gives that w.p.a. 1,

$$
\begin{aligned}
P\left(\hat{T}(\theta_\ell) > \bar{c} \text{ or } \hat{T}(\theta_m) > \bar{c}; \mathcal{N}(\lambda, \hat{\Sigma})\right) &\leq H\left(\bar{c}, \Delta, \hat{\rho}_\ell(b_\ell, b_u)\right)\\
&\leq \sup_{\rho \leq \rho_2^*(\alpha,\eta) - \xi} \sup_{\Delta \geq 0} H\left(\bar{c}, \Delta, \rho\right) < \alpha - \eta
\end{aligned}
$$

where the last inequality follows from the construction of $\bar{c}$. □

Lemma 9. *$p(c)$ in (43) is continuous at $c \geq 0$.*

Proof. For $\varepsilon > 0$, let

$$p_k(c, \varepsilon) = P\left(c_k^{\mathrm{m}}(c) \geq T_k > c_k^{\mathrm{m}}(c - \varepsilon)\right) \leq P\left(c - \varepsilon < T_k \leq c\right).$$

Then

$$\lim_{\varepsilon \to 0} p_k(c, \varepsilon) = 0$$

for all $c \geq 0$ since (i) under (79) and $k = \ell, m, u$, or under (81) and $k = \ell, u$, T_k is continuously distributed, (ii) under (81) and $k = m$,

$$P\left(c - \varepsilon < T_k \leq c\right) \leq P\left(c - \varepsilon < T_k\right) = 0.$$

But then with $E_k(c) = \{P(T_k > c_k^{\mathrm{m}}(c))\}$,

$p(c - \varepsilon) - p(c)$

$$\le \max\left\{P\left(E_\ell(c-\varepsilon) \text{ or } \{E_m(c-\varepsilon) \text{ and } E_u(c-\varepsilon)\}\right) - P\left(E_\ell(c) \text{ or } \{E_m(c) \text{ and } E_u(c)\}\right),\right.$$
$$\left. P\left(E_u(c-\varepsilon) \text{ or } \{E_m(c-\varepsilon) \text{ and } E_\ell(c-\varepsilon)\}\right) - P\left(E_u(c) \text{ or } \{E_m(c) \text{ and } E_\ell(c)\}\right)\right\}$$
$$\le p_\ell(c,\varepsilon) + p_m(c,\varepsilon) + p_u(c,\varepsilon) \xrightarrow{\varepsilon \to 0} 0.$$

Thus $p(c)$ is continuous at $c \ge 0$. □

Lemma 10. *Let*

$$H(\rho, \alpha^c) = (1-\alpha^c)\Phi\left((1+\rho)\Xi\right) + \alpha^c\Phi\left((\rho-1)\Xi\right), \tag{107}$$
$$\Xi = \sqrt{\frac{1}{2\rho}\log\left(\frac{(1-\alpha^c)(1+\rho)}{\alpha^c(1-\rho)}\right)}. \tag{108}$$

It holds that

1. *for all* $\alpha \in (0, \frac{1}{2})$, $\alpha^c \in (\frac{\alpha}{2}, \alpha)$, *there is a unique solution* $\rho_1^*(\alpha, \alpha^c) \in (0,1)$ *such that*
$$H\left(\rho_1^*(\alpha,\alpha^c), \alpha^c\right) = 1 - \frac{\alpha}{2}. \tag{109}$$

2. *Let* $\xi > 0$. *There is* $\varepsilon > 0$ *such that for all* $\rho \in [0, \rho_1^*(\alpha,\alpha^c) - \xi]$,
$$H\left(\rho, \alpha^c\right) \le 1 - \frac{\alpha}{2} - \varepsilon. \tag{110}$$

Proof. Straightforward calculation gives that for all $\rho \in (0,1)$,

$$\frac{\mathrm{d}H(\rho,\alpha^{\mathrm{c}})}{\mathrm{d}\rho} = \frac{\alpha^{\mathrm{c}}}{\sqrt{\pi\rho}(\rho+1)}\left(\frac{(1-\alpha^{\mathrm{c}})(\rho+1)}{\alpha^{\mathrm{c}}(1-\rho)}\right)^{-\frac{(1-\rho)^2}{4\rho}}\sqrt{\log\left(\frac{(1-\alpha^{\mathrm{c}})(\rho+1)}{\alpha^{\mathrm{c}}(1-\rho)}\right)} > 0.$$

In addition, note that

$$\lim_{\rho\to 1}\Xi = \lim_{\rho\to 1}\sqrt{\frac{1}{2\rho}\log\left(\frac{(1-\alpha^c)(1+\rho)}{\alpha^c(1-\rho)}\right)} = +\infty,$$
$$\lim_{\rho\to 1}(\rho-1)\Xi = \lim_{\rho\to 1}(\rho-1)\sqrt{\frac{1}{2\rho}\log\left(\frac{(1-\alpha^c)(1+\rho)}{\alpha^c(1-\rho)}\right)} = \lim_{\rho\to 1} -\sqrt{\frac{(1-\rho)^2}{2\rho}\log\left(\frac{1}{1-\rho}\right)} = 0,$$
$$\lim_{\rho\to 0}\Xi = \lim_{\rho\to 0}\sqrt{\frac{1}{2\rho}\log\left(\frac{(1-\alpha^c)(1+\rho)}{\alpha^c(1-\rho)}\right)} = +\infty,$$

thus

$$\lim_{\rho\to 1} H(\rho,\alpha^{\mathrm{c}}) = (1-\alpha^{\mathrm{c}}) + \frac{1}{2}\alpha^{\mathrm{c}} = 1-\frac{\alpha^{\mathrm{c}}}{2} > 1-\frac{\alpha}{2},$$
$$\lim_{\rho\to 0} H(\rho,\alpha^{\mathrm{c}}) = (1-\alpha^{\mathrm{c}}) = 1-\alpha^{\mathrm{c}} < 1-\frac{\alpha}{2},$$

where the inequality follows from $\alpha^{\mathrm{c}} \in (\frac{\alpha}{2},\alpha)$. Since $H(\rho,\alpha^{\mathrm{c}})$ is strictly increasing in $\rho \in (0,1)$, there is a unique solution $\rho^* \in (0,1)$ that $H(\rho^*,\alpha^c) = 1-\frac{\alpha}{2}$.

(110) holds trivially with $\varepsilon = 1-\frac{\alpha}{2} - H(\rho_1^*(\alpha,\alpha^c)-\xi,\alpha^c)$. □

Lemma 11. *Suppose Assumptions 1, 2, 3, 4, and 5 hold. If $A_l = A_u$ and*

$$\sup_{P\in\mathcal{P}} \max_{b_1\in\mathcal{B}} \min_{b_2\in\mathcal{B}} \rho_\ell(b_1,b_2) < \rho_1^*(\alpha,\alpha^c), \tag{111}$$

where $\rho_1^*(\alpha,\alpha^c)$ is defined in Lemma 10, then there is $\alpha' > \alpha$ such that

$$\liminf_n \inf_{P\in\mathcal{P}} P\left(\hat{T}(\theta) > \hat{c}^c(\theta;\alpha^c) \text{ for all } \theta \notin CI^{sim}(\hat{\lambda}_n,\hat{\Sigma}_n/n,\alpha')\right) = 1. \tag{112}$$

Proof. Let

$$\xi = \frac{1}{2}\min\left\{\rho_1^*(\alpha,\alpha^c) - \sup_{P\in\mathcal{P}} \max_{b_1\in\mathcal{B}} \min_{b_2\in\mathcal{B}} \rho_\ell(b_1,b_2),\ \rho_1^*(\alpha,\alpha^c)\right\} > 0. \tag{113}$$

By Lemma 10, there is $\varepsilon > 0$ such that $H(\rho,\alpha^{\mathrm{c}}) \le 1-\frac{\alpha}{2}-\varepsilon$, with $H(\rho,\alpha^{\mathrm{c}})$ defined in (107), for all $\rho \in [0,\rho_1^*(\alpha,\alpha^{\mathrm{c}})-\xi]$. Next, I show (112) for $\alpha' = \min\{\alpha^{\mathrm{c}}+\frac{\alpha}{2},\alpha+\varepsilon\}$.

Consider $\theta \notin CI^{\mathrm{sim}}(\hat{\lambda}_n,\hat{\Sigma}_n/n,\alpha')$. Denote

$$\mathcal{Z}_b = \frac{\hat{\lambda}_{\ell,b}-\theta}{\hat{\sigma}_{\ell,b}/\sqrt{n}}, \quad \mathcal{Z}_{\ell,b} = -\mathcal{Z}_b, \quad \mathcal{Z}_{u,b} = \mathcal{Z}_b.$$

Without loss of generality, assume that

$$\hat{T}(\theta) = \mathcal{Z}_1 \text{ and } \hat{\rho}_{12} = \hat{\rho}_\ell(1,2) \le \rho_1^*(\alpha,\alpha^c) - \xi.$$

The inequality happens with probability approaching one. In this case

$$t_{u,1} = \min_{\tilde{b}\in\mathcal{B}} \left(1+\hat{\rho}_\ell(1,\tilde{b})\right)^{-1}\left(\mathcal{Z}_{\ell,\tilde{b}} + \hat{\rho}_\ell(1,\tilde{b})\mathcal{Z}_{u,1}\right) \le \frac{\hat{\rho}_{12}\mathcal{Z}_1 - \mathcal{Z}_2}{1+\hat{\rho}_{12}}$$
$$t_{u,2} = \min_{\tilde{b}\in\mathcal{B}:\hat{\rho}_u(1,\tilde{b})<1} \left(1-\hat{\rho}_\ell(1,\tilde{b})\right)^{-1}\left(\mathcal{Z}_{u,\tilde{b}} - \hat{\rho}_\ell(1,\tilde{b})\mathcal{Z}_{u,1}\right) \le \frac{\mathcal{Z}_2 - \hat{\rho}_{12}\mathcal{Z}_1}{1-\hat{\rho}_{12}}.$$

θ is rejected if

$$\Phi(\mathcal{Z}_1) > \Phi(\hat{c}^{\mathrm{c}}) = (1-\alpha^{\mathrm{c}})\Phi\left(t_{u,2}\right) + \alpha^{\mathrm{c}}\Phi\left(t_{u,1}\right).$$

By construction, $\mathcal{Z}_2 \geq \mathcal{Z}_1$, thus it suffices to show that

$$\Phi(\mathcal{Z}_1) > \sup_{z_2 \geq \mathcal{Z}_1} G(z_2) \text{ where} \tag{114}$$

$$G(z_2) = (1-\alpha^{\mathrm{c}})\Phi\left(\frac{z_2 - \hat{\rho}_{12}\mathcal{Z}_1}{1-\hat{\rho}_{12}}\right) + \alpha^{\mathrm{c}}\Phi\left(\frac{\hat{\rho}_{12}\mathcal{Z}_1 - z_2}{1+\hat{\rho}_{12}}\right). \tag{115}$$

The first order derivative of $G(z_2)$ with respect to z_2 is

$$g(z_2) = \frac{1-\alpha^{\mathrm{c}}}{1-\hat{\rho}_{12}}\phi\left(\frac{z_2 - \hat{\rho}_{12}\mathcal{Z}_1}{1-\hat{\rho}_{12}}\right) - \frac{\alpha^{\mathrm{c}}}{1+\hat{\rho}_{12}}\phi\left(\frac{\hat{\rho}_{12}\mathcal{Z}_1 - z_2}{1+\hat{\rho}_{12}}\right).$$

$g(z_2) \geq 0$ is equivalent to

$$\log\left(\frac{1-\alpha^{\mathrm{c}}}{\alpha^{\mathrm{c}}}\frac{1+\hat{\rho}_{12}}{1-\hat{\rho}_{12}}\right) \geq \frac{2\hat{\rho}_{12}\left(z_2 - \mathcal{Z}_1\hat{\rho}_{12}\right)^2}{\left(1-\hat{\rho}_{12}^2\right)^2}. \tag{116}$$

(i) If $2\alpha^c - 1 < \hat{\rho}_{12} \leq 0$, then (116) holds trivially, and thus

$$\sup_{z_2 \geq \mathcal{Z}_1} G(z_2) = \lim_{z_2 \to \infty} G(z_2) = 1 - \alpha^{\mathrm{c}} < \Phi\left(\mathcal{Z}_1\right),$$

and (114) holds. The inequality follows from $1 - \frac{\alpha'}{2} > 1 - \alpha^{\mathrm{c}}$.

(ii) If $-1 \leq \hat{\rho}_{12} < 2\alpha^c - 1$, straightforward calculation shows that $G(z_2)$ decreases in $[\max\left\{\mathcal{Z}_1, z_2^*\right\}, z_2^*]$ and increases in $[z_2^*, +\infty)$, where

$$z_2^* = \hat{\rho}_{12}\mathcal{Z}_1 + (1-\hat{\rho}_{12}^2)\sqrt{\Xi}.$$

and Ξ is defined in (108). Thus

$$\sup_{z_2 \geq \mathcal{Z}_1} G(z_2) \leq \max\left\{G(\mathcal{Z}_1), \lim_{z_2 \to \infty} G(z_2)\right\} < \Phi\left(\mathcal{Z}_1\right).$$

(iii) If $\hat{\rho}_{12} \in \left(0, \rho^*(\alpha, \alpha^c) - \frac{\xi}{2}\right)$, straightforward calculation shows that $G(z_2)$

increases in $[\max\{\mathcal{Z}_1, z_2^*\}, z_2^*]$ and decreases in $[z_2^*, +\infty)$, thus

$$\sup_{z_2 \geq \mathcal{Z}_1} G(z_2) \leq G(z_2^*) = (1-\alpha^c)\Phi((1+\rho)\Xi) + \alpha^c \Phi((\rho-1)\Xi) < \Phi(\mathcal{Z}_1)$$

The last inequality is by Lemma 10. $\square$

Lemma 12. *Suppose Assumptions 1, 2, 3, 4 hold, and* $\mathcal{P}_n$ *satisfies (35).* $\hat{\Lambda}_\eta$ *is defined as in (32). Let* $\alpha \in (0, \frac{1}{2})$, $\alpha^c \in (\frac{\alpha}{2}, \alpha)$, $\eta \in [0, \frac{\alpha-\alpha^c}{2})$. *Then there exists* $\alpha' > \alpha$ *such that*

$$\liminf_n \inf_{P \in \mathcal{P}_n} P\left(\hat{c}^t \leq \Phi^{-1}(1 - \frac{\alpha'}{2})\right) = 1.$$

Proof. Consider $P_n \in \mathcal{P}_n$ satisfying (78) and (77). I show that there is $\xi > 0$ such that

$$\sup_{\lambda \in \hat{\Lambda}_\eta} \bar{p}(c^{\text{sim}}, \lambda, \hat{\Sigma}_n/n) \leq \alpha - \eta - \xi \text{ w.p.a. } 1. \tag{117}$$

Then since $\sup_{\lambda \in \hat{\Lambda}_\eta} \bar{p}(c, \lambda, \hat{\Sigma}_n/n)$ is continuous in c, there is $c' < c^{\text{sim}}$ such that

$$\sup_{\lambda \in \hat{\Lambda}_\eta} \bar{p}(c', \lambda, \hat{\Sigma}_n/n) \leq \alpha$$

and the conclusion holds with $\alpha' = 2(1 - \Phi(c'))$. To show (117), I consider two cases.

Case 1: $\eta = 0$ and thus $\hat{\Lambda}_\eta = \Lambda$. By definition,

$$\begin{aligned}\bar{p}(c, \lambda, \Sigma) \leq \max\Big\{ & P\left(\hat{T}(\theta_\ell) > \tilde{c}^{\text{m}}(\theta_\ell, c) \text{ or } \hat{T}(\theta_m) > \tilde{c}^{\text{m}}(\theta_m, c); \mathcal{N}(\lambda, \Sigma)\right), \\ & P\left(\hat{T}(\theta_u) > \tilde{c}^{\text{m}}(\theta_u, c) \text{ or } \hat{T}(\theta_m) > \tilde{c}^{\text{m}}(\theta_m, c); \mathcal{N}(\lambda, \Sigma)\right)\Big\}. \end{aligned} \tag{118}$$

Therefore, it suffices to show that there is $\xi > 0$ such that

$$\sup_{\lambda \in \Lambda} P\left(\hat{T}(\theta_\ell) > \tilde{c}^{\text{m}}(\theta_\ell, c^{\text{sim}}) \text{ or } \hat{T}(\theta_m) > \tilde{c}^{\text{m}}(\theta_m, c^{\text{sim}}); \mathcal{N}(\lambda, \hat{\Sigma}_n/n)\right) \leq \alpha - \xi$$

and same argument applies to (118). To do so, let

$$\bar{\bar{p}}(\Delta; c^{\text{sim}}) = \sup_{(\lambda,\Sigma) \in \mathcal{D}(\Delta)} P\left(\hat{T}(\theta_\ell) > \tilde{c}^{\text{m}}(\theta_\ell, c^{\text{sim}}) \text{ or } \hat{T}(\theta_m) > \tilde{c}^{\text{m}}(\theta_m, c^{\text{sim}}); \mathcal{N}(\lambda, \Sigma)\right)$$

$$\mathcal{D}(\Delta) = \left\{(\lambda, \Sigma) : \frac{\theta_u - \theta_\ell}{\sigma_{u,b_u}} = \Delta_1, \min_{b: A_{\ell,b} \neq A_{\ell,b_\ell}} \frac{\lambda_{\ell,b} - \theta_\ell}{\sigma_{\ell,b}} = \Delta_2, \min_{b: A_{u,b} \neq A_{u,b_u}} \frac{\theta_u - \lambda_{u,b}}{\sigma_{u,b}} = \Delta_3\right\}.$$

If $\{b : A_{\ell,b} \neq A_{\ell,b_\ell}\} = \emptyset$, then $\Delta_2 = \infty$. Same for Δ_3.

For $\Delta_1 = 0$, $\theta_\ell = \theta_u = \theta$. If $0 \leq \Delta_2, \Delta_3 < \infty$,

$$
\begin{aligned}
&\bar{\bar{p}}(0, \Delta_2, \Delta_3; c^{\text{sim}}) \\
= &\sup_{(\lambda,\Sigma)\in\mathcal{D}(\Delta)} P\left(\hat{T}(\theta) > \tilde{c}^{\text{m}}(\theta, c^{\text{sim}}); \mathcal{N}(\lambda, \Sigma)\right) \\
\leq &\sup_{(\lambda,\Sigma)\in\mathcal{D}(\Delta)} P\left(\min_{b\in\mathcal{B}} \frac{\hat{\lambda}_{\ell,b} - \theta}{\sigma_{\ell,b}} > c^{\text{sim}}; \mathcal{N}(\lambda, \Sigma)\right) + P\left(\min_{b\in\mathcal{B}} \frac{\theta - \hat{\lambda}_{u,b}}{\sigma_{u,b}} > c^{\text{sim}}; \mathcal{N}(\lambda, \Sigma)\right) \\
\leq &\sup_{(\lambda,\Sigma)\in\mathcal{D}(\Delta)} P\left(\min\left\{\frac{\hat{\lambda}_{\ell,b_\ell} - \theta}{\sigma_{\ell,b_\ell}}, \frac{\hat{\lambda}_{\ell,b'} - \lambda_{\ell,b'}}{\sigma_{\ell,b'}} + \Delta_2\right\} > c^{\text{sim}}; \mathcal{N}(\lambda, \Sigma)\right) \\
&\quad + P\left(\min\left\{\frac{\theta - \hat{\lambda}_{u,b_u}}{\sigma_{u,b_u}}, \frac{\lambda_{u,\tilde{b}} - \hat{\lambda}_{u,\tilde{b}}}{\sigma_{u,\tilde{b}}} + \Delta_3\right\} > c^{\text{sim}}; \mathcal{N}(\lambda, \Sigma)\right) < \alpha.
\end{aligned}
$$

In addition,

$$
\begin{aligned}
&\lim_{\Delta_{2,3}\to\infty} \bar{\bar{p}}(0, \Delta_2, \Delta_3; \alpha^{\text{sim}}) \\
= &\lim_{\Delta_{2,3}\to\infty} \sup_{(\lambda,\Sigma)\in\mathcal{D}(\Delta)} P\left(\hat{T}(\theta) > \tilde{c}^{\text{m}}(\theta, c^{\text{sim}}) \middle| \hat{T}(\theta) = \frac{\hat{\lambda}_{\ell,b_\ell} - \theta}{\sigma_\ell} \text{ or } \hat{T}(\theta) = \frac{\theta - \hat{\lambda}_{u,b_u}}{\sigma_{u,b_u}}; \mathcal{N}(\lambda, \Sigma)\right) \\
\leq &\alpha^{\text{c}} < \alpha
\end{aligned}
$$

where the second line follows from

$$
\lim_{(\Delta_2,\Delta_3)\to\infty^2} P\left(\hat{T}(\theta) = \frac{\hat{\lambda}_{\ell,b_\ell} - \theta}{\sigma_{\ell,b_\ell}} \text{ or } \hat{T}(\theta) = \frac{\theta - \hat{\lambda}_{u,b_u}}{\sigma_{u,b_u}}\right) = 1.
$$

By the continuity of $\bar{\bar{p}}(\Delta_1, \Delta_2, \Delta_3; \alpha^{\text{sim}})$ in (Δ_2, Δ_3), we get

$$
\sup_{\Delta_2\Delta_3\geq 0} \bar{\bar{p}}(0, \Delta_2, \Delta_3; \alpha^{\text{sim}}) < \alpha.
$$

For all $\Delta_1 > 0$,

$$
\sup_{\Delta_2,\Delta_3\in[0,\infty)} \bar{\bar{p}}(\Delta_1, \Delta_2, \Delta_3; \alpha^{\text{sim}})
$$

$$
\begin{aligned}
\leq & \sup_{(\lambda,\Sigma):\frac{\theta_u-\theta_\ell}{\sigma_{u,b_u}}=\Delta_1} \sup_{(\lambda,\Sigma)\in\mathcal{D}(\Delta)} P\left(\hat{T}(\theta_\ell) > c^{\text{sim}} \text{ or } \hat{T}(\theta_m) > c^{\text{sim}}; \mathcal{N}(\lambda,\Sigma)\right) \\
\leq & \sup_{(\lambda,\Sigma):\frac{\theta_u-\theta_\ell}{\sigma_{u,b_u}}=\Delta_1} P\left(\max\left\{\min_{b\in\mathcal{B}} \frac{\hat{\lambda}_{\ell,b}-\theta_\ell}{\sigma_{\ell,b}}, \min_{b\in\mathcal{B}} \frac{\theta_m-\theta_u}{\sigma_{u,b}} + \frac{\theta_u-\hat{\lambda}_{u,b}}{\sigma_{u,b}}\right\} > c^{\text{sim}}; \mathcal{N}(\lambda,\Sigma)\right) \\
\leq & \sup_{(\lambda,\Sigma):\frac{\theta_u-\theta_\ell}{\sigma_{u,b_u}}=\Delta_1} P\left(\frac{\hat{\lambda}_{\ell,b_\ell}-\theta_\ell}{\sigma_{\ell,b_\ell}} > c^{\text{sim}} \text{ or } -\frac{\Delta_1}{2} + \frac{\theta_{u,b_u}-\hat{\lambda}_u}{\sigma_{u,b_u}} > c^{\text{sim}}; \mathcal{N}(\lambda,\Sigma)\right) \\
\leq & P\left(\mathcal{N}(0,1) > c^{\text{sim}}\right) + P\left(-\frac{\Delta_1}{2} + \mathcal{N}(0,1) > c^{\text{sim}}\right) < \alpha.
\end{aligned}
$$

In addition,

$$
\lim_{\Delta_1\to\infty} \sup_{\Delta_2,\Delta_3\in[0,\infty)} \bar{\bar{p}}(\Delta_1,\Delta_2,\Delta_3;\alpha^{\text{sim}}) = \frac{\alpha}{2} < \alpha.
$$

In sum, by the continuity of $\bar{\bar{p}}(\Delta_1,\Delta_2,\Delta_3;\alpha^{\text{sim}})$ in $(\Delta_1,\Delta_2,\Delta_3)$,

$$
\sup_{\Delta\in[0,\infty)^3} \bar{\bar{p}}(\Delta_1,\Delta_2,\Delta_3;\alpha^{\text{sim}}) < \alpha.
$$

It follows that there is $\xi > 0$ such that

$$
\begin{aligned}
& \sup_{\lambda\in\Lambda} P\left(\hat{T}(\theta_\ell) > \tilde{c}^{\text{m}}(\theta_\ell, c^{\text{sim}}) \text{ or } \hat{T}(\theta_m) > \tilde{c}^{\text{m}}(\theta_m, c^{\text{sim}}); \mathcal{N}(\lambda, \hat{\Sigma}_n/n)\right) \\
\leq & \sup_{\Delta\in[0,+\infty)^3} \bar{\bar{p}}(\Delta_1,\Delta_2,\Delta_3;c^{\text{sim}}) \leq \alpha - \xi
\end{aligned}
$$

and (117) holds.

Case 2. $\eta > 0$. There exists $\kappa_n' \to \infty$ and $\kappa_n' \ll \kappa_n$ such that

$$
\hat{\Lambda}_n \subseteq \Lambda_n = \left\{\lambda : \lambda_{u,b_u} - \lambda_{\ell,b_\ell} \geq \frac{\kappa_n'}{\sqrt{n}}\right\} \text{ w.p.a. 1.}
$$

Thus

$$
\begin{aligned}
& \sup_{\lambda\in\hat{\Lambda}_n} \bar{p}(c^{\text{sim}}, \lambda, \hat{\Sigma}_n/n) \\
\leq & \sup_{\lambda\in\Lambda_n} \max\left\{P\left(\hat{T}(\theta_\ell) > c^{\text{sim}}; \mathcal{N}(\lambda,\hat{\Sigma}_n/n)\right),\ P\left(\hat{T}(\theta_u) > c^{\text{sim}}; \mathcal{N}(\lambda,\hat{\Sigma}_n/n)\right)\right\} \text{ w.p.a. 1} \\
\leq & \sup_{\lambda\in\Lambda_n} \max\left\{P\left(\max\left\{\frac{\hat{\lambda}_{\ell,b_\ell}-\theta_\ell}{\hat{\sigma}_{\ell,b_\ell}/\sqrt{n}}, \frac{\theta_\ell-\theta_u}{\hat{\sigma}_{u,b_u}/\sqrt{n}} + \frac{\theta_u-\hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}}\right\} > c^{\text{sim}}; \mathcal{N}(\lambda,\hat{\Sigma}_n/n)\right),\right.
\end{aligned}
$$

$$P\left(\max\left\{\frac{\hat{\lambda}_{\ell,b_\ell}-\theta_\ell}{\hat{\sigma}_{\ell,b_\ell}/\sqrt{n}}+\frac{\theta_\ell-\theta_u}{\hat{\sigma}_{\ell,b_\ell}/\sqrt{n}},\frac{\theta_u-\hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}}\right\}>c^{\mathrm{sim}};\mathcal{N}(\lambda,\hat{\Sigma}_n/n)\right)\Bigg\}$$
$$=P\left(\mathcal{N}(0,1)>c^{\mathrm{sim}}\right)+o_p(1)=\frac{\alpha}{2}+o_p(1)<\alpha-\eta \text{ w.p.a. } 1.$$

Thus (117) holds with $\xi \in (0, \frac{\alpha}{2}-\eta)$. □

C Union Bounds in Rambachan and Roth (2023)

Consider a simple panel data model with $t=-\underline{T},...,\overline{T}$. Let $\gamma\in\mathbb{R}^{\underline{T}+\overline{T}}$ be a vector of "event study" coefficients, which can be decomposed as

$$\gamma=\begin{pmatrix}\gamma^{pre}\\ \gamma^{post}\end{pmatrix}=\begin{pmatrix}\xi^{pre}\\ \tau+\xi^{post}\end{pmatrix}.$$

$\xi^{pre}=\left(\xi^{pre}_{-\underline{T}},...,\xi^{pre}_{-1}\right)$, $\xi^{post}=\left(\xi^{post}_{1},...,\xi^{post}_{\overline{T}}\right)$, and $\gamma_0=\xi^{pre}_0$ is normalized to zero. The target object $\theta=\iota'\tau$ is the weighted average of the average treatment effect on the treated for post-policy years, and ξ represents the bias from differences in trends.

Under the relative magnitudes relaxation,

$$\left|\xi^{post}_t-\xi^{post}_{t-1}\right|\le M\max_{s=-1,...,-\underline{T}}\left|\xi^{pre}_{s+1}-\xi^{pre}_s\right|.$$

The identified set of θ is (1) with $\lambda_\ell=\lambda_u=A\delta$,

$$A=\begin{bmatrix} M\left|\iota' L\right|\mathbf{1}_{\overline{T}\times 1}\mathcal{I}_{\underline{T}}, & \mathbf{1}_{\underline{T}\times 1}\\ -M\left|\iota' L\right|\mathbf{1}_{\overline{T}\times 1}\mathcal{I}_{\underline{T}}, & \mathbf{1}_{\underline{T}\times 1}\end{bmatrix},\ \underbrace{L}_{\overline{T}\times\overline{T}}=\begin{bmatrix}1 & & & \\ 1 & 1 & & \\ & & \ddots & \\ 1 & 1 & \cdots & 1\end{bmatrix},\quad \underbrace{\delta}_{\underline{T}+1}=\begin{pmatrix}\xi^{pre}_{1-\underline{T}}-\xi^{pre}_{-\underline{T}}\\ \vdots\\ \xi^{pre}_0-\xi^{pre}_{-1}\\ \iota'\gamma^{post}\end{pmatrix}.$$

Under the second difference relative magnitudes relaxation,

$$\left|(\xi_t-\xi_{t-1})-(\xi_{t-1}-\xi_{t-2})\right|\le M\max_{s=-1,...,-\underline{T}}\left|(\xi_{s+1}-\xi_s)-(\xi_s-\xi_{s-1})\right|.$$

The identified set of θ is (1) with $\lambda_\ell = \lambda_u = A\delta$,

$$A = \begin{bmatrix} \iota' LM\mathcal{I}_{\underline{T}-1}, & \mathbf{1}_{(\underline{T}-1)\times 1} \\ \iota' LM\mathcal{I}_{\underline{T}-1}, & \mathbf{1}_{(\underline{T}-1)\times 1} \end{bmatrix}, \delta = \left(\Delta_{-\underline{T}+2}, \cdots, \Delta_0, \iota'\gamma^{post} + \iota' H\xi_{-1}\right)',$$
$$H = \left(1, ..., \bar{T}\right)',\ L = \left(1, 3, ..., \frac{t(t+1)}{2}, ..., \frac{\overline{T}(\overline{T}+1)}{2}\right)',$$
$$\Delta_t = \left(\xi_t - \xi_{t-1}\right) - \left(\xi_{t-1} - \xi_{t-2}\right).$$