Inference on Union Bounds with Applications to DiD, RDD, Bunching, and Structural Counterfactuals

Xinyue Bei*
Duke University
<xinyue.bei@duke.edu>

January 10, 2024

Latest Version

Abstract

A union bound is a union of multiple bounds. Union bounds occur in a wide variety of empirical settings, from relaxations of the difference-in-differences parallel trends assumption to counterfactual analysis with partially identified structural parameters. In this paper, I provide the first general and systematic study of inference on these kinds of bounds. When the union is taken over a finite set, I propose a confidence interval based on modified conditional inference. I show that it improves upon existing methods in a large set of data generating processes. When the union is taken over an infinite set, I consider the set defined by moment inequalities, as is common in practice. I then propose a calibrated projection based inference procedure that generalizes results from the moment inequality subvector inference literature and is computationally simple. Finally, the new procedures give statistically significant results while the pre-existing alternatives do not in two empirical applications, the sensitivity analysis in Dustmann, Lindner, Schönberg, Umkehrer, and Vom Berge (2022) and the counterfactual analysis in Dickstein and Morales (2018).

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

JEL classification: C01, C12, C15

*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 Duke microeconometrics breakfast, Triangle Econometrics Conference, Université de Montréal Econometrics Seminar for their useful comments and suggestions.

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:

1. 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) 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 (ATT) 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.

2. Counterfactual analysis in structural models. Dickstein and Morales (2018) study how the information set processed by exporters affects their decisions. One of the counterfactuals of interest is the change in the number of exporters with a change in their information set or fixed cost. The structural parameter satisfies a set of moment inequalities, and the counterfactual outcome may only be partially identified even if the structural parameter is known. Consequently, the identified set of the counterfactual outcome is a union bound, where the *set* contains all the structural parameters that satisfy the moment inequality restrictions, and each *bound* is the identified set of the counterfactual outcome given a potential value of the structural parameter.

I revisit these two applications in detail for empirical illustration. I also discuss applications to regression discontinuity designs, bunching strategies to identify the elasticity of taxable income, marginal treatment effects, and misspecification in instrumental variable models in Section 2.

In this paper, I provide the first general and systematic study of inference on union bounds. I consider two categories of union bound inference: (i) the *set* is finite as in the DiD example, and (ii) the *set* is infinite as in the counterfactual example. I study the inference procedures for the target object in both cases.

In the first case when the *set* is finite, 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). 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. Rambachan and Roth (2023) propose an inference procedure for their sensitivity analysis in DiD. However, the inference 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 CI at a rate of $\sqrt{n}$, and thus has material power improvement upon Ye et al. (2023). 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.

In Section 5, I conduct extensive simulations based on the DiD settings in Rambachan and Roth (2023) and compare the performance of my modified conditional CI to the simple CI, the adjusted bootstrap in Ye et al. (2023) and the hybrid CI in Rambachan and Roth (2023). 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.[1] 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 the second case when the *set* is infinite, I consider the case when the *set*, e.g. the identified set of the structural parameter, is formed by moment inequalities, as is often the case in practice. I assume that each *bound*, e.g. the bound of the counterfactual given a potential true structural parameter, is a function of data moments. Among the empirical applications in this setup, it is common to estimate point-identified nuisance parameters separately from the structural parameters to improve computational efficiency. Thus I allow the *set* and the *bounds* to include plug-in estimands. One of the current practices

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

for counterfactual inference is to first construct a valid confidence set for the structural parameter (the *set*), and then take a union of the estimated counterfactuals (the *bounds*) over this confidence set, treating the plug-in estimator and the counterfactual as known. This simple projection CI can be wide since it projects a confidence set of a higher dimensional structural parameter. Moreover, it may not have proper coverage because it does not adjust for sampling uncertainty in the plug-in estimator and the counterfactual.

In Section 4, I propose a calibrated projection based procedure. The calibrated CI is a union of the single calibrated CI of each *bound* over the calibrated confidence set for the *set* of structural parameters. The calibration means that the critical value used to construct the single CI and the confidence set are chosen so that the coverage rate of the target object, rather than the structural parameter, is above the nominal value. Calculation of the critical value is done by repeatedly solving a set of linear programs, which makes it computationally simple. This method uses insights from subvector inference in Kaido, Molinari, and Stoye (2019), where a subvector is a known function, usually a single element, of the structural parameter. Different from the subvector inference procedure, in this paper (i) the target object can be unknown as well as partially identified even if the structural parameter is known, and (ii) first step plug-in estimators are allowed. These two differences allow the new inference procedure to have broader application. Simulations in Section 5 confirm good size and power properties.

In Section 6, I illustrate the proposed inference procedures in two empirical applications. First, I consider the application using Rambachan and Roth (2023)'s sensitivity analysis in Dustmann et al. (2022), which provides an example with a finite *set*. 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 1. The authors conduct the analysis using DiD and relax the parallel trends assumption following Rambachan and Roth (2023). Under all levels of relaxation, the modified conditional CI is shorter than the simple CI and the one provided by Rambachan and Roth (2023). Under the benchmark relaxation, my CI suggests that the elasticity is smaller than 1 with a 95% confidence level, while Rambachan and Roth (2023) and a simple CI do not give results significantly smaller than 1. My method gives a breakdown relaxation 33% to 66% larger than Rambachan and Roth (2023) and the simple CI. Next, I apply the calibrated projection CI to Dickstein and Morales (2018). In this case, the *set* of *bounds* over which the union is taken is infinite. The authors are interested in the percentage change in the number of exporters under a counterfactual information set. They report a simple projection CI, where the results for all three subsamples are significant. However, as previously discussed, the simple projection CI is invalid. For that reason, I first validate their method by properly adjusting for the estimation uncertainty in plug-in estimators and counterfactuals, and with this adjustment, two of the CIs cross zero. Then I report the calibrated CI, which is not only valid but also more efficient, and the calibrated projection CI restores statistical significance.

Related Literature

In the rest of this introductory section, I review the related literature.

When the *set* is finite

Although there are many empirical examples where the identified set is a union of finite bounds, 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. 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, Ye et al. (2023) study the relaxation of the parallel trends assumption in DiD based on a negative correlation bracketing strategy. The resulting identified set for the ATT 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, Rambachan and Roth (2023) 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. Rambachan and Roth (2023) first construct the CI for each element in the partition based on I. 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 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 inference procedure constructed in this paper also contributes to other related literature, such as intersection bounds, moment inequalities, directional differentiable functions, and conditional inference.

The union bound inference complements the large literature on intersection bounds and testing 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), D. W. K. Andrews and Guggenberger (2009), D. W. K. Andrews and Soares (2010), D. W. K. Andrews and Shi (2013), and Bugni, Canay, and Shi (2015), among others, for different inference 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 fundamentally different from intersection bounds and requires a different treatment.

My method also sheds light on inference on directionally differentiable functions. In many cases, 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.[2] 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 a growing literature on conditional inference, see, e.g. Moreira (2003), Kleibergen (2005), I. Andrews and Mikusheva (2016), I. Andrews, Kitagawa, and McCloskey (2019), I. Andrews, Kitagawa, and McCloskey (2021), I. Andrews et al. (2023), and Rambachan and Roth (2023), 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]Specifically, the space of λ_ℓ and λ_u under (9) is not convex.

When the *set* is infinite

Inference procedures with an infinite *set* are closely related to the literature on subvector inference in moment inequality models, where a subvector is a known function, usually a single component, of the structural parameter. Kaido et al. (2019) proposed a calibrated projection procedure for subvector inference that uses a local linearization approach to compute the critical value through linear programming. See I. Andrews et al. (2023), Bugni, Canay, and Shi (2017), Bei (2024), Chernozhukov, Newey, and Santos (2023), Chen, Christensen, and Tamer (2018), Cox and Shi (2023), among others, for different subvector inference procedures. However, the subvector usually does not contain sufficient information for decision making or policy suggestions. Hence, it is important to extend the previous work to construct the confidence intervals for counterfactual outcomes. In this paper, I follow the insights from Kaido et al. (2019) and propose a calibrated projection CI for unknown and potentially partially identified target objects, which has broad application to counterfactuals. In addition, previous papers assume that all parameters are estimated jointly by a set of moment inequality restrictions, which rules out plug-in estimators. Nevertheless, in practice, it is common to estimate the point identified parameters in a first step separately from the structural model. In this context, I propose a simple way to adjust for estimation uncertainty in plug-in estimators.

My paper is also related to the literature on counterfactual analysis and marginal treatment effect models, but my procedure applies to general moment inequality models. Kalouptsidi, Kitamura, Lima, and Souza-Rodrigues (2021) study the identification of counterfactuals for structural dynamic discrete choice models and propose an inference procedure that bypasses model estimation and directly obtains the confidence sets for the counterfactuals. However, their procedure requires a specific structure. Cho and Russell (2023) propose an inference procedure in a similar but more restrictive setting where the *bound* and *set* are linear in the structural parameters. Their procedure involves bootstrapping the value functions of randomly perturbed linear programming problems, which is computationally attractive but also produces a confidence set with a coverage probability of one. Unfortunately, their method does not apply to nonlinear moment inequalities and counterfactuals. Mogstad, Santos, and Torgovitsky (2016) propose a profile based inference procedure for estimated functionals of partially identified parameters that allows a nonparametric framework.[3] In their context, the equality restrictions are allowed to be random while the inequalities are deterministic, formed from the parameter space.

[3] Mogstad et al. (2016) is a working paper version of Mogstad, Santos, and Torgovitsky (2018).

2 Setup and Examples

2.1 Setup

A union bound is defined as a union of *bounds* $[\lambda_\ell(\beta),\ \lambda_u(\beta)]$ over *set* $\mathcal{B}$. 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

$$\theta \in \bigcup_{\beta \in \mathcal{B}} [\lambda_\ell(\beta),\ \lambda_u(\beta)]\,. \tag{1}$$

When the union bound is a connected interval, we can imply write it as

$$\theta \in \left[\inf_{\beta\in\mathcal{B}} \lambda_\ell(\beta),\ \sup_{\beta\in\mathcal{B}} \lambda_u(\beta)\right]. \tag{2}$$

In this paper, I assume that λ_ℓ and λ_u are unknown but consistently estimable with an asymptotically normal estimator. $\mathcal{B}$ is either known or consistently estimable, in the sense that the Hausdorff distance between $\hat{\mathcal{B}}$ and $\mathcal{B}$ converges to zero in probability. Below I illustrate this setting in different examples.

I consider separately finite $\mathcal{B}$ and infinite $\mathcal{B}$. The inference procedure for these two cases is distinct because (i) with finite $\mathcal{B}$, I consider the estimation uncertainty for all *bounds* jointly, while with infinite $\mathcal{B}$, I focus on a small subset of *bounds* each time, which is conservative by valid, to simplify the computation; (ii) when $\mathcal{B}$ is finite, asymptotically we can treat the *set* as known, while when $\mathcal{B}$ is infinite, we need to adjust the *set* for estimation uncertainty.[4]

2.2 Examples

When $\mathcal{B}$ is finite

Example 1. (Difference in Differences). Rambachan and Roth (2023) study a more credible approach to the parallel trends assumption in DiD. To illustrate, consider a simple 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

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

[4] For simplicity, I assume that the index set for the lower and upper bounds are the same, and this is the case in all the empirical examples listed below. Moreover, this assumption is without loss of generality, as we can add redundant bounds to achieve this. For instance, if $\theta \in [\min\{\lambda_{\ell,1}, \lambda_{\ell,2}\}, \lambda_{u,1}]$, we can add $\lambda_{u,2} = \lambda_{u,1}$ then the identified set of θ has form (1) with $\mathcal{B} = \{1, 2\}$.

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

$$\left|\xi^{post} - 0\right| \leq M \max_{t=-1,...,-\underline{T}} \left|\xi^{pre}_{t+1} - \xi^{pre}_t\right|, \tag{3}$$

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 (3), the identified set of θ is a union bound in (2) with $\mathcal{B} = \{1, ..., 2\underline{T}\}$, [5]

$$\lambda_\ell(\beta) = \lambda_u(\beta) = \begin{cases} \gamma^{post} + M\left(\gamma^{pre}_{-\beta+1} - \gamma^{pre}_{-\beta}\right) & \text{if } \beta = 1, ..., \underline{T}, \\ \gamma^{post} - M\left(\gamma^{pre}_{\underline{T}-\beta+1} - \gamma^{pre}_{\underline{T}-\beta}\right) & \text{if } \beta = \underline{T}+1, ..., 2\underline{T}. \end{cases} \tag{4}$$

Hasegawa et al. (2019), Ye et al. (2023), 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 Figure 1a. Assume that the preference is specified as in Saez (2010) by the isoelastic utility function:

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

where y is the before-tax income, $c = y - T(y)$ is the after-tax income (or consumption), 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 1b. 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

$$\sigma_\ell \min\left\{g(\xi_1), g(\xi_2)\right\} \leq g(\xi) \leq \sigma_u \max\left\{g(\xi_1), g(\xi_2)\right\} \text{ for } \xi \in [\xi_1, \xi_2] \tag{5}$$

[5] In Appendix A.5, I give the union bound form of Rambachan and Roth (2023) "relative magnitudes" relaxation and "second differences relative magnitudes" relaxation with multiple post policy periods.

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

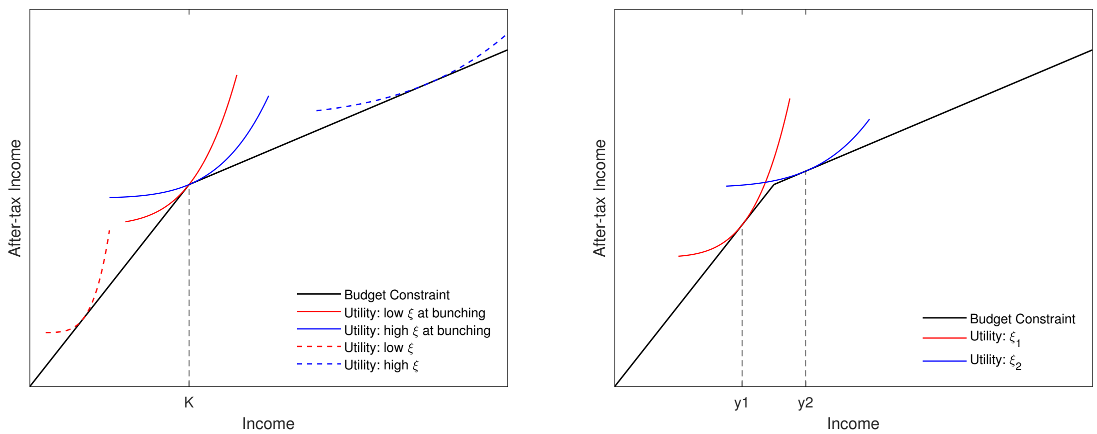
$$\theta \in \left[\min_{\beta \in \mathcal{B}} \lambda_\ell(\beta),\ \max_{\beta \in \mathcal{B}} \lambda_u(\beta)\right] \cap \mathbb{R}_+$$

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

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

$$f^-(y_1) = \lim_{y \uparrow y_1} f(y), \quad f^+ = \lim_{y \downarrow y_2} f(y),$$

and $f(y)$ is the density of y. Note that the identified set of θ is restricted by $\mathbb{R}_+$, but it is easy to see that if we have a valid CI for $\tilde{\theta} \in \bigcup_{\beta \in \mathcal{B}} [\lambda_\ell(\beta),\ \lambda_u(\beta)]$, then 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. □



(a) Budget Constraint and Utility

(b) Bunching Interval

Figure 1: Example 2 Bunching and Taxable Income Elasticity

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 with running variable X. 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 at the threshold is

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

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

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

Let γ_h be the regression coefficient and $\theta_h = e_1'\gamma_h$, where $e_1 = (1, 0, ..., 0)'$. 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 specification errors 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 (2) with

$$\mathcal{B} = \left\{(s_\ell, s_u, x_\ell, x_u) : s_\ell, s_u \in \{-1, 1\}, x_\ell \in S_X^-, x_u \in S_X^+\right\},$$

$$\lambda_\ell(s_\ell, s_u, x_\ell, x_u) = \lambda_u(s_\ell, s_u, x_\ell, x_u) = \theta_h + s_\ell\xi(x_\ell) + s_u\xi(x_u).$$

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

Example 4. (Falsification Adaptive Set). 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 $\text{cov}(Z, U) = 0$, (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 $\xi \in \mathbb{R}$, where $\xi \geq 0$ measures the level of relaxation. The corresponding identified set of θ accounting for the relaxation by ξ is

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

where 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, they show that FAS is characterized by (2), where

$$\lambda_\ell(\beta) = \lambda_u(\beta) = \frac{\psi_\beta}{\pi_\beta},$$

ψ_β and π_β are the β-th element of $\psi = \text{var}(Z)^{-1}\text{cov}(Z, Y)$, $\pi = \text{var}(Z)^{-1}\text{cov}(Z, X)$, and

$$\mathcal{B} = \{\beta = 1, ..., L : \pi_\beta \neq 0\}.$$

In their empirical application, the authors implicitly assume that either $\pi_\beta = 0$ or $|\pi_\beta| \geq \varepsilon > 0$ for all β, so that $\mathcal{B}$ is consistently estimable, in which case my procedure applies. If we allow $\pi_\beta \to 0$ as the sample size increases, we may not be able to consistently estimate $\mathcal{B}$, and inference is more complicated. I leave the second case for future research. 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 misspecification inference for interval identified parameters. The identified set for θ is $[\theta_L, \theta_u]$, and this set is empty under misspecification where $\theta_L > \theta_U$. Stoye (2020) suggests reporting the misspecification robust identified set

$$[\theta_L, \theta_U] \cup \left\{ \frac{\sigma_U\theta_L + \sigma_L\theta_U}{\sigma_L + \sigma_U} \right\}, \tag{6}$$

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 (2) with $\mathcal{B} = \{1, 2\}$,

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

Stoye (2020) proposes a CI for (6), but it does not apply to general union bounds. □

When $\mathcal{B}$ is infinite

Example 5. (Counterfactual Analysis). Dickstein and Morales (2018) study how the information potential exporters possess influences their decisions. In the structural model, all firms located in the home country are indexed by $i = 1, \ldots, N$ and choose whether to sell in each export market $j = 1, \ldots, J$. The export profit that i would obtain in market j is

$$\pi_{ij} = d_{ij}\left(r_{ij} - \beta_1 - \beta_2 dist_j - \beta_3 v_{ij}\right)$$

where $d_{ij} \in \{0,1\}$ is firm i's export decision, r_{ij} is the revenue in market j, $dist_j$ denotes the distance from the home country to j, $v_{ij} \sim \mathcal{N}(0,1)$ represents the determinant of π_{ij} observed by the firm i but not by the researcher, and $(\beta_1, \beta_2, \beta_3)$ are structural parameters. Let $\mathcal{J}_{ij}$ be the information that firm i possess. A risk-neutral firm i will decide to export to j if and only if

$$E\left[r_{ij} \left| \mathcal{J}_{ij}\right.\right] - \beta_1 - \beta_2 dist_j \geqslant \beta_3 v_{ij}$$

which implies that

$$d_{ij}\left(\beta_3^{-1} E\left[r_{ij} \left| \mathcal{J}_{ij}\right.\right] - \beta_3^{-1}\beta_1 - \beta_3^{-1}\beta_2 dist_j - v_{ij}\right) \geqslant 0. \tag{7}$$

Based on (7), the authors construct a set of moment inequalities to get the identified set of β, which is, in union bound notation, $\mathcal{B}$. The counterfactual outcome of interest is the proportion change in exporter numbers $\theta = \frac{E\left[d_{ij}; \mathcal{J}_{ij}^c, g(\beta)\right]}{E[d_{ij}; \mathcal{J}_{ij}, \beta]}$ under a different information set or a different fixed cost, where $\mathcal{J}_{ij}^c$ is the counterfactual information set and $g(\beta)$ is the counterfactual structural parameter. Given β, the authors show that $\theta \in [\lambda_\ell(\beta),\ \lambda_u(\beta)]$ with $\lambda_\ell(\beta)$ and $\lambda_u(\beta)$ point identified. Consequently, the identified set of θ is given by (1). Further details of the moment conditions and counterfactuals are given in Appendix B.3.

Structural counterfactual analysis with union bound identified set is very common in applied microeconomics, such as industry organization, trade, political economy, etc. To list a few, see examples Berry, Eizenberg, and Waldfogel (2016), Bombardini, Li, and Trebbi (2023), Crawford and Yurukoglu (2012), Ciliberto, Murry, and Tamer (2021), Eizenberg (2014), Jia (2008), Kalouptsidi et al. (2021), Kireyev (2020), Wollmann (2018), Yang (2020), among many others. □

Example 6. (Marginal Treatment Effects). Mogstad et al. (2018) propose a method to partially identify the policy relevant treatment parameters, exploiting the insight that the IV estimand and many treatment parameters can be expressed as weighted averages of the same underlying marginal treatment effects. Assume that the treatment is determined by

$$D = \mathbf{1}\left[U \leq p(Z)\right]$$

where U is an unobservable with uniform $[0,1]$ distribution, $p(Z)$ is the propensity score, and Z are exogenous instruments. Assume that the marginal treatment response functions have parametric form $m_0(x,u;\beta)$ and $m_1(x,u;\beta)$, where m_0 and m_1 are known functions, x is other covariates and β are parameters. Let $E\left[s(D,Z)Y\right]$ be an IV-like estimand using instrument Z, where $s(D,Z)$ is a known function. Then β is partially identified by

$$\mathcal{B} = \left\{\beta \in \overline{\mathcal{B}} : E\left[\int_0^1 m_0(u,X;\beta)s(0,Z)\mathbf{1}\left[u > p(Z)\right]du\right]\right.$$

$$+E\left[\int_0^1 m_1(u,X;\beta)s(1,Z)\mathbf{1}\left[u\leq p(Z)\right]du\right]=E\left[s(D,Z)Y\right]\Big\}$$

where $\overline{\mathcal{B}}$ is the feasible set of structural parameter β. Assume that the target object θ is the average treatment effect. Then it is point identified by

$$\lambda_\ell(\beta)=\lambda_u(\beta)=E\left[\int_0^1 m_1(u,X;\beta)du\right]-E\left[\int_0^1 m_0(u,X;\beta)du\right]$$

for a given β, and the identified set of θ is given in (1). □

Example 7. (Plausibly Exogenous IVs) Conley et al. (2012) consider an instrumental variable model when the instruments are only plausibly exogenous:

$$Y=X\theta+Z\beta+U$$

where Y is the outcome, X is a $L_1\times 1$ endogenous variable and Z is a $L_2\times 1$ vector of potentially invalid instruments. To simplify the illustration, let $L_1=1$. Similar to Masten and Poirier (2021) in Example 4, under exogeneity $E\left[U|Z\right]=0$, proper rank conditions, and given a plausible exogenous value β, the average treatment effect θ is point identified by

$$\lambda_\ell(\beta)=\lambda_u(\beta)=\frac{E\left[XZ'\right]E\left[ZZ'\right]E\left[Z(Y-Z\beta)\right]}{E\left[XZ'\right]E\left[ZZ'\right]E\left[ZX\right]}.$$

In their empirical application, Conley et al. (2012) suggest using a continuous relaxation where

$$\mathcal{B}=\left\{\beta:\beta_j\in\left[-\bar{\beta}_j,\bar{\beta}_j\right],j=1,...,\dim(\beta)\right\}$$

and $\bar{\beta}$ is a user chosen tuning parameter. In this case, $\mathcal{B}$ is known, which is a degenerate case of consistently estimable $\mathcal{B}$, and the inference procedure proposed in Section 4 applies. □

3 Inference with Finite $\mathcal{B}$

In this section, I study inference on θ with finite $\mathcal{B}$, and I focus on connected union bounds where

$$\theta\in\left[\min_{\beta\in\mathcal{B}}\lambda_\ell(\beta),\ \max_{\beta\in\mathcal{B}}\lambda_u(\beta)\right]$$

since this is the case in all examples with finite $\mathcal{B}$ in Section 2.2. A similar inference procedure applies to general, potentially non-connected, union bounds with the form (1), which I discuss in Appendix A.3.

To simplify the presentation, I first assume that $\mathcal{B}$ is known. In this case, $\lambda_\ell(\beta)$

and $\lambda_u(\beta)$ are finite dimensional vectors indexed by β, so I write each of them as $|\mathcal{B}|$ dimensional vectors λ_ℓ and λ_u, with the b-th element $\lambda_{\ell,b}$ and $\lambda_{u,b}$. I illustrate with a normally distributed estimator $\hat{\lambda}_n = \left(\hat{\lambda}_\ell, \hat{\lambda}_u\right)$ such that

$$\begin{pmatrix} \hat{\lambda}_\ell \\ \hat{\lambda}_u \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \lambda_\ell \\ \lambda_u \end{pmatrix}, \Sigma_n\right), \ \Sigma_n = \begin{bmatrix} \Sigma_{\ell,n} & \Sigma'_{\ell u,n} \\ \Sigma_{\ell u,n} & \Sigma_{u,n} \end{bmatrix} \tag{8}$$

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 later present theorems under general DGPs where this condition holds in Section 3.4.

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

$$H_0 : \min_{b\in\mathcal{B}} \lambda_{\ell,b} \leq \theta \leq \max_{b\in\mathcal{B}} \lambda_{u,b}. \tag{9}$$

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 $1-\alpha$ confidence interval is

$$CI^{\mathrm{m}}(\hat{\lambda}_n, \Sigma_n; \alpha) = \left[\inf_{\phi(\theta,\hat{\lambda}_n,\Sigma_n)=0} \theta, \ \sup_{\phi(\theta,\hat{\lambda}_n,\Sigma_n)=0} \theta\right]. \tag{10}$$

3.1 The Test Statistic

The test statistic has a max-min form

$$\hat{T}(\theta) = \max\left\{\min_{b\in\mathcal{B}} \mathcal{Z}_{\ell,b}, \ \min_{b\in\mathcal{B}} \mathcal{Z}_{u,b}\right\} \tag{11}$$

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

$$\mathcal{Z}_{\ell,b} = \frac{\hat{\lambda}_{\ell,b} - \theta}{\sigma_{\ell,b}}, \text{ and } \mathcal{Z}_{u,b} = \frac{\theta - \hat{\lambda}_{u,b}}{\sigma_{u,b}}. \tag{12}$$

Observing that H_0 in (9) is equivalent to

$$H_0 : \max\left\{\min_{b\in\mathcal{B}}(\lambda_{\ell,b} - \theta), \ \min_{b\in\mathcal{B}}(\theta - \lambda_{u,b})\right\} \leq 0, \tag{13}$$

and the test statistic is constructed by replacing λ_ℓ and λ_u in (13) 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

$$CI^{\text{sim}} = \left[\min_{b\in\mathcal{B}} \hat{\lambda}_{\ell,b} - \sigma_{\ell,b}\Phi^{-1}(1-\frac{\alpha}{2}),\ \max_{b\in\mathcal{B}} \hat{\lambda}_{u,b} + \sigma_{u,b}\Phi^{-1}(1-\frac{\alpha}{2})\right], \tag{14}$$

which is often used in current practice, see e.g. Kolesár and Rothe (2018), Hasegawa et al. (2019), Ban and Kedagni (2022). The simple confidence interval 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

$$b_\ell = \arg\min_{b\in\mathcal{B}} \lambda_{\ell,b}, \quad b_u = \arg\max_{b\in\mathcal{B}} \lambda_{u,b}. \tag{15}$$

and observe that

$$\begin{aligned}
P\left(\theta \notin CI^{\text{sim}}\right) &= P\left(\max\left\{\min_{b\in\mathcal{B}} \mathcal{Z}_{\ell,b},\ \min_{b\in\mathcal{B}} \mathcal{Z}_{u,b}\right\} > \Phi^{-1}(1-\frac{\alpha}{2})\right) \\
&\le P\left(\max\left\{\mathcal{Z}_{\ell,b_\ell},\ \mathcal{Z}_{u,b_u}\right\} > \Phi^{-1}(1-\frac{\alpha}{2})\right) \\
&\le P\left(\mathcal{Z}_{\ell,b_\ell} > \Phi^{-1}(1-\frac{\alpha}{2})\right) + P\left(\mathcal{Z}_{u,b_u} > \Phi^{-1}(1-\frac{\alpha}{2})\right) \\
&= P\left(\frac{\hat{\lambda}_{\ell,b_\ell} - \lambda_{\ell,b_\ell}}{\sigma_{\ell,b}} + \frac{\lambda_{\ell,b_\ell} - \theta}{\sigma_{\ell,b}} > \Phi^{-1}(1-\frac{\alpha}{2})\right) \qquad (16) \\
&\quad + P\left(\frac{\lambda_{u,b_u} - \hat{\lambda}_{u,b_u}}{\sigma_{u,b}} + \frac{\theta - \lambda_{u,b_u}}{\sigma_{u,b}} > \Phi^{-1}(1-\frac{\alpha}{2})\right) \qquad (17) \\
&\le \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.
\end{aligned}$$

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 in the sample. The second inequality follows from $P(A \cup B) \le P(A) + P(B)$. The final inequality holds under the null hypothesis (9).

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.

$$\frac{\lambda_{u,b_u} - \lambda_{\ell,b_\ell}}{\max\left\{\sigma_{u,b_u}, \sigma_{\ell,b_\ell}\right\}} \gg 0.$$

In such cases, either (16) or (17) 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 the second inequality 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.

$$\min_{b\in\mathcal{B}\backslash b_\ell} \frac{\lambda_{\ell,b}-\lambda_{\ell,b_\ell}}{\sigma_{\ell,b}} \gg 0, \quad \min_{b\in\mathcal{B}\backslash b_u} \frac{\lambda_{u,b_u}-\lambda_{u,b}}{\sigma_{u,b}} \gg 0, \quad \frac{\lambda_{u,b_u}-\lambda_{\ell,b_\ell}}{\min\left\{\sigma_{\ell,b_\ell},\sigma_{u,b_u}\right\}} \approx 0. \tag{18}$$

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 (18) 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 (18),

$$P\left(E_\ell \cup E_u\right) \approx 1 \tag{19}$$

where[6]

$$\begin{aligned} E_\ell &= \left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,\hat{b}_\ell}\right\} \cap \left\{\lambda_{\ell,\hat{b}_\ell} \le \theta\right\}, \\ E_u &= \left\{\hat{T}(\theta) = \mathcal{Z}_{u,\hat{b}_u}\right\} \cap \left\{\lambda_{u,\hat{b}_u} \ge \theta\right\}, \\ \hat{b}_\ell &= \arg\min_{b\in\mathcal{B}_\ell} \mathcal{Z}_{\ell,b},\ \hat{b}_u = \arg\min_{b\in\mathcal{B}_u} \mathcal{Z}_{u,b}. \end{aligned} \tag{20}$$

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

$$P\left(\hat{T}(\theta) > \hat{c}(\theta)\,|E_\ell \cup E_u\right) \le \alpha^{\text{c}} < \alpha, \tag{21}$$

[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 $\hat{b}_u$.

the unconditional rejection rate is bounded above by α following from (19). Therefore, I construct a conditional critical value based on the conditional distribution

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

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

Lemma 1. *Under H_0 and (8). Let b_1, b_2 satisfy $\lambda_{\ell,b_1}\le\theta\le\lambda_{u,b_2}$, then*

$$\frac{\Phi\left(\hat{T}(\theta)\right)-\Phi\left(t_{\ell,1}(\theta,b_1)\right)}{\Phi\left(t_{\ell,2}(\theta,b_1)\right)-\Phi\left(t_{\ell,1}(\theta,b_1)\right)}\Big|\left\{\hat{T}(\theta)=\mathcal{Z}_{\ell,b_1}\right\} \overset{\text{FOSD}}{\preceq} \text{Unif}(0,1)$$

$$\frac{\Phi\left(\hat{T}(\theta)\right)-\Phi\left(t_{u,1}(\theta,b_1)\right)}{\Phi\left(t_{u,2}(\theta,b_1)\right)-\Phi\left(t_{u,1}(\theta,b_1)\right)}\Big|\left\{\hat{T}(\theta)=\mathcal{Z}_{u,b_2}\right\} \overset{\text{FOSD}}{\preceq} \text{Unif}(0,1)$$

where

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

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

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

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

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

and $\mathcal{Z}_\ell$, $\mathcal{Z}_u$ are defined in (12).

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 $\mathcal{TN}\left(0,1,[t_{\ell,1}(\theta,b_1),t_{\ell,2}(\theta,b_1)]\right)$, where $\mathcal{TN}\left(\mu,\sigma^2,[t_1,t_2]\right)$ is a normal distribution $\mathcal{N}\left(\mu,\sigma^2\right)$ truncated at $[t_1,t_2]$. Hence, we can guarantee conditional coverage by using the $1-\alpha^{\mathrm{c}}$ quantile of $\mathcal{TN}\left(0,1,[t_{\ell,1}(\theta,b),t_{\ell,2}(\theta,b)]\right)$ with $\alpha^{\mathrm{c}}<\alpha$.

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

$$\hat{c}^{\mathrm{c}}(\theta,\alpha^{\mathrm{c}})=\begin{cases}\Phi^{-1}\left(\alpha^{\mathrm{c}}\Phi\left(t_{\ell,1}(\theta,\hat{b}_\ell)\right)+(1-\alpha^{\mathrm{c}})\Phi\left(t_{\ell,2}(\theta,\hat{b}_\ell)\right)\right) & \text{if } \mathcal{Z}_{\ell,\hat{b}_\ell}\ge\mathcal{Z}_{u,\hat{b}_u}\\ \Phi^{-1}\left(\alpha^{\mathrm{c}}\Phi\left(t_{u,1}(\theta,\hat{b}_u)\right)+(1-\alpha^{\mathrm{c}})\Phi\left(t_{u,2}(\theta,\hat{b}_u)\right)\right) & \text{if } \mathcal{Z}_{\ell,\hat{b}_\ell}<\mathcal{Z}_{u,\hat{b}_u}\end{cases}\tag{22}$$

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 rejection rate under more and less favorable DGPs.

Proposition 1. *Assume that*

$$P\left(\mathcal{Z}_{\ell,\hat{b}_\ell} = \mathcal{Z}_{u,\hat{b}_u}\right) = 0. \tag{23}$$

Under H_0 and (8), 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 (23), the set $E_\ell \cup E_u$ can be partitioned into $\left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\right\}$ and $\left\{\hat{T}(\theta) = \mathcal{Z}_{u,b_2}\right\}$ for b_1, b_2 satisfying $\lambda_{b_1} \leq \theta \leq \lambda_{b_2}$. Hence (24) follows directly from Lemma 1. Condition (23) holds in most examples previously discussed and is assumed in Proposition 1 for simplicity. Under more favorable DGPs diverging from (18), the conditional quantile 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 $\frac{\hat{\lambda}_{\ell,\hat{b}_\ell} - \hat{\lambda}_{u,b_u}}{\sigma_{u,b_u}} \approx -\infty$ for a large identified set. In this case,

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

Moreover, if the minimum $\lambda_{\ell,-b_\ell}$ is not well separated from λ_{ℓ,b_ℓ}, then the upper bound

$$t_{\ell,2}(\theta, \hat{b}_\ell) = \min_{\tilde{b} \in \mathcal{B}: \rho_\ell(\hat{b}_\ell, \tilde{b}) < 1} \left(1 - \rho_\ell(\hat{b}_\ell, \tilde{b})\right)^{-1} \left(\frac{\hat{\lambda}_{\ell,\tilde{b}} - \lambda_{\ell,\tilde{b}}}{\sigma_{\ell,\tilde{b}}} + \frac{\lambda_{\ell,\tilde{b}} - \lambda_{\ell,b_\ell}}{\sigma_{\ell,\tilde{b}}} - \mathcal{Z}_{\ell,\hat{b}_\ell}\right) + \mathcal{Z}_{\ell,\hat{b}_\ell}$$

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 8. (Simple Union Bounds) Consider a simple union bound

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

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

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

Without loss of generality, assume that $\hat{T}(\theta) = \hat{\lambda}_1 - \theta$. In this case, the conditional critical value is

$$\begin{aligned} \hat{c}^{\mathrm{c}}(\theta; \alpha^{\mathrm{c}}) &= \Phi^{-1}\left((1-\alpha^{\mathrm{c}})\Phi\left(\hat{\lambda}_2 - \theta\right) + \alpha^{\mathrm{c}}\Phi\left(\theta - \hat{\lambda}_2\right)\right) \\ &< \Phi^{-1}\left(1-\alpha^{\mathrm{c}}\right) < \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \end{aligned} \tag{26}$$

where the first line is by construction, and the first inequality follows from

$$\Phi\left(\hat{\lambda}_2 - \theta\right) + \Phi\left(\theta - \hat{\lambda}_2\right) = 1.$$

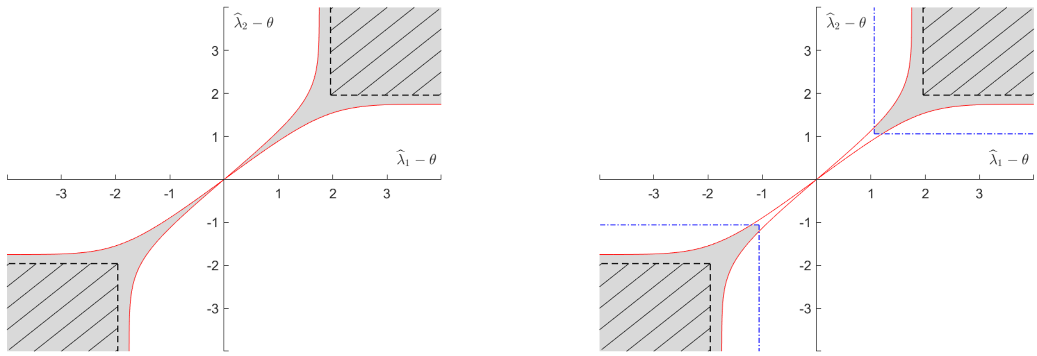
If the minimizer and maximizer are well separated, e.g. $\lambda_1 = \theta$ and $\lambda_2 \to \infty$, the efficient critical value is $\Phi^{-1}(1-\alpha)$, as discussed in Imbens and Manski (2004). In this case, $\hat{\lambda}_2$ will be large and $\hat{c}^{\mathrm{c}}(\theta; \alpha^{\mathrm{c}}) \to \Phi^{-1}(1-\alpha^{\mathrm{c}})$, which is slightly conservative. 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 2a, I plot the rejection region for the simple and conditional critical values with $\alpha = 0.05$ and $\alpha^{\mathrm{c}} = 0.04$. 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, resulting in larger power. □

It is important to note that $\hat{c}^{\mathrm{c}}(\theta, 1-\alpha^{\mathrm{c}})$ 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 (18). 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

To guarantee proper coverage, I introduce a novel modification to the conditional critical value:

$$\hat{c}^{\mathrm{m}}(\theta; \alpha) = \tilde{c}^{\mathrm{m}}(\theta, \hat{c}^{\mathrm{t}}; \alpha) = \max\left\{\hat{c}^{\mathrm{c}}(\theta, \alpha^{\mathrm{c}}), \hat{c}^{\mathrm{t}}\right\} \tag{27}$$



(a) Conditional Critical Value (b) Modified Conditional Critical Value

Figure 2: Example 8 - Rejection Region

The red curve corresponds to $\hat{c}^{\mathrm{c}}(\theta; \alpha^{\mathrm{c}})$. The blue solid line represents the lower truncation $\hat{c}^{\mathrm{t}}$. The grey region on the left panel denotes the rejection region for the test with the conditional critical value $\hat{c}^{\mathrm{c}}(0; \alpha^{\mathrm{c}})$, and the one on the right denotes the rejection region with the modified conditional critical value $\hat{c}^{\mathrm{m}}(\theta, \alpha)$. The two square regions filled with lines represent the rejection region of the simple test. In this example, $\alpha = 0.05$ and $\alpha^{\mathrm{c}} = 0.04$.

where $\hat{c}^{\mathrm{t}}$ is defined later in (31).

To illustrate the construction of the lower truncation, let $\widetilde{CI}^{\mathrm{m}}(c)$ be the confidence interval based on (10) with $\hat{c}^{\mathrm{m}}(\theta; \alpha)$ replaced by $\tilde{c}^{\mathrm{m}}(\theta, c; \alpha)$. Given a potential true parameter λ, the rejection rate at θ is

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

It suffices to define the lower truncation as the minimum value that achieves uniform size control, i.e.

$$c^{\mathrm{t}}(\theta) = \inf\left\{c \in \mathbb{R}_+ : \sup_{\lambda \in \Lambda_0(\theta)} p(c; \theta, \lambda) \leq \alpha\right\}, \tag{28}$$

where Λ_0 is the set of feasible λ satisfying H_0:

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

Note that $c^{\mathrm{t}}(\theta) \leq c^{\mathrm{sim}} = \Phi^{-1}(1 - \frac{\alpha}{2})$ because

$$p(c^{\mathrm{sim}}; \theta, \lambda) \leq \alpha$$

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

$$P\left(E_\ell \cup E_u; \lambda\right) \geq \frac{1 - \alpha}{1 - \alpha^c} \tag{29}$$

with $\alpha^{\mathrm{c}} < \alpha$, where E_ℓ, E_u are defined in (20). Thus we only need to consider truncation in more favorable DGPs deviating from (18), 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 $p(c; \theta, \lambda)$ 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 confidence interval 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 $p(c; \theta, \lambda)$ by

$$
\begin{aligned}
p(c;\theta,\lambda) \leq & \max\left\{P\left([\theta_\ell,\theta_m] \not\subseteq \widetilde{CI}^{\mathrm{m}}(c);\lambda\right), P\left([\theta_m,\theta_u] \not\subseteq \widetilde{CI}^{\mathrm{m}}(c);\lambda\right)\right\} \\
\leq & \max\left\{P\left(\hat{T}(\theta_\ell) > \tilde{c}^{\mathrm{m}}(\theta_\ell, c) \text{ or } \left\{\hat{T}(\theta_m) > \tilde{c}^{\mathrm{m}}(\theta_m, c) \text{ and } \hat{T}(\theta_u) > \tilde{c}^{\mathrm{m}}(\theta_u, c)\right\};\lambda\right),\right. \\
& \left. P\left(\hat{T}(\theta_u) > \tilde{c}^{\mathrm{m}}(\theta_u, c) \text{ or } \left\{\hat{T}(\theta_m) > \tilde{c}^{\mathrm{m}}(\theta_m, c) \text{ and } \hat{T}(\theta_\ell) > \tilde{c}^{\mathrm{m}}(\theta_\ell, c)\right\};\lambda\right)\right\} \\
=: & \bar{p}(c,\lambda). \qquad (30)
\end{aligned}
$$

Therefore, it is valid, but conservative, to replace $p(c; \theta, \lambda)$ in (28) with $\bar{p}(c, \lambda)$. In addition, to avoid maximization over an unbounded set, let $\hat{\Lambda}$ be a $1-\eta$ compact confidence set of λ, such as[7]

$$
\hat{\Lambda}_\eta = \left\{(\lambda_\ell, \lambda_u) \in \Lambda : \frac{\left|\hat{\lambda}_{\ell,b} - \lambda_{\ell,b}\right|}{\sigma_{\ell,b}} \leq \Phi\left(1 - \frac{\eta}{4|\mathcal{B}|}\right), \frac{\left|\hat{\lambda}_{u,b} - \lambda_{u,b}\right|}{\sigma_{u,b}} \leq \Phi\left(1 - \frac{\eta}{4|\mathcal{B}|}\right)\right\}
$$

with suggested value $\eta = 0.001$. In sum, it suffices to use

$$
\hat{c}^{\mathrm{t}} = \inf_c \left\{c \geq 0 : \sup_{\lambda \in \hat{\Lambda}_\eta} \bar{p}(c,\lambda) + \eta \leq \alpha\right\}, \qquad (31)
$$

and $\bar{p}(c, \lambda)$ is defined in (30). In terms of computation, $\bar{p}(c, \lambda)$ 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 (29). 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 decreases 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

[7]Section A.1 gives a more efficient $\hat{\Lambda}_\eta$ set.

research.

Remark 1. The relaxation in (30) to get $\bar{p}(c,\lambda)$ 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. For instance, we can add the quarter point $\theta_{1/4} = (3\theta_\ell + \theta_u)/4$ and three-quarter point $\theta_{3/4} = (\theta_\ell + 3\theta_u)/4$ in addition to θ_m, and bound $p(c;\theta,\lambda)$ by

$$\begin{aligned} p(c;\theta,\lambda) \leq \max\Big\{ & P\left(\left[\theta_\ell, \theta_{1/4}\right] \not\subseteq \widetilde{CI}^{\mathrm{m}}(c^{\mathrm{t}});\lambda\right), P\left(\left[\theta_{1/4}, \theta_m\right] \not\subseteq \widetilde{CI}^{\mathrm{m}}(c^{\mathrm{t}});\lambda\right) \\ & P\left(\left[\theta_m, \theta_{3/4}\right] \not\subseteq \widetilde{CI}^{\mathrm{m}}(c^{\mathrm{t}});\lambda\right), P\left(\left[\theta_{3/4}, \theta_u\right] \not\subseteq \widetilde{CI}^{\mathrm{m}}(c^{\mathrm{t}});\lambda\right)\Big\} \end{aligned}$$

which requires conducting the test at five points θ_ℓ, $\theta_{1/4}$, θ_m, $\theta_{3/4}$ and θ_u, but returns a weakly shorter CI. □

Example 8. (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 $c^{\mathrm{t}} = 1.06$. In Figure 2b, I plot the rejection region of the modified conditional test and the simple test. The blue dotted curve is the boundary corresponding to the lower truncation c^{t}, and the grey region is the rejection region for the modified conditional test. The rest are the same as in Figure 2a. 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) There are known $|\mathcal{B}| \times J$ matrices A_ℓ, A_u such that*

$$\lambda_\ell = A_\ell \delta_P,\ \lambda_u = A_u \delta_P \tag{32}$$

[8]To see this, note that by construction $\hat{c}^{\mathrm{m}} \geq 0$, so $\left[\hat{\lambda}_{\ell,\hat{b}_\ell}, \hat{\lambda}_{u,\hat{b}_u}\right] \subseteq CI^{\mathrm{m}}\left(\hat{\lambda}_n, \Sigma_n; \alpha\right)$. If $\sqrt{n}\left(\lambda_{u,b_u} - \lambda_{\ell,b_\ell}\right) \to \infty$, we have $P\left(\hat{\lambda}_{\ell,\hat{b}_\ell} \leq \theta_m \leq \hat{\lambda}_{u,\hat{b}_u}\right) \to 1$.

$$\hat{\lambda}_\ell = A_\ell \hat{\delta}_n, \ \hat{\lambda}_u = A_u \hat{\delta}_n$$
$$\hat{\Sigma}_n = \left[\begin{array}{cc} A'_\ell & A'_u \end{array} \right]' \hat{\Omega}_n \left[\begin{array}{cc} A'_\ell & A'_u \end{array} \right]$$

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. We 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) We have an estimator* $\hat{\Omega}_n$ *that is uniformly consitent for* Ω_P,

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

for all $\varepsilon > 0$.

Assumption 5. *(Confidence Set of* $(\lambda_\ell, \lambda_u)$*) 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 Rambachan and Roth (2023) in Appendix A.5. Assumption 5 requires that $\hat{\Lambda}$ is a uniformly valid $1-\eta$ confidence set of $(\lambda_\ell, \lambda_u)$, e.g. $\hat{\Lambda}$ implied by (69) in Section A.1.

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*

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

Remark 2. Theorem 1 shows that the modified conditional CI has uniform asymptotic coverage under a large set of DGPs. Moreover, with some modification, the proposed method can be applied to cases where Assumption 4 fails. The main idea is that we can rewrite the union bounds as the union of several sub-union bounds, with Assumption 4 holding in each sub-union bound. By taking the union of CIs for each sub-union bound, we can get a valid CI for θ. I illustrate this in Example 9 in Appendix A.4. □

Remark 3. The same inference procedure and coverage property apply when $\mathcal{B}$ is unknown but consistently estimable in the sense that $d_H\left(\hat{\mathcal{B}}, \mathcal{B}\right) \xrightarrow{p} 0$ uniformly where d_H is the Hausdorff distance and $\hat{\mathcal{B}} \subseteq \overline{\mathcal{B}}$ is the estimator for $\mathcal{B}$. $\overline{\mathcal{B}}$ is a finite outer set of $\mathcal{B}$. The reason is that for finite $\mathcal{B}$, there is $\varepsilon > 0$ such that

$$\liminf_n \inf_{P\in\mathcal{P}} P\left(\hat{\mathcal{B}} = \mathcal{B}\right) = \liminf_n \inf_{P\in\mathcal{P}} P\left(d_H\left(\hat{\mathcal{B}}, \mathcal{B}\right) < \varepsilon\right) = 1. \tag{33}$$

Therefore, asymptotically we can treat $\hat{\mathcal{B}}$ as the true set without adjusting the estimation uncertainty. Masten and Poirier (2021) and Apfel and Windmeijer (2022) implicitly assume (33) in their empirical applications, see Example 4 for more discussion. □

Next, I compare the modified CI to two existing approaches which are also uniformly valid: (i) the simple CI given in (14), and (ii) the adjusted bootstrap CI proposed in Ye et al. (2023).

Theorem 2. *(Power Comparison with Simple CI) Suppose Assumptions 1, 2, 3, and 4 hold. And* $\hat{\Lambda}_\eta$ *is defined as in (68). 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*

$$\limsup_{P\in\mathcal{P}_n} \max_{b_1\in\mathcal{B}} \min_{b_2\in\mathcal{B}} \rho_\ell(b_1, b_2) < \rho_1^*(\alpha, \alpha^c), \tag{34}$$

$$\limsup_{P\in\mathcal{P}_n} \rho_\ell(b_\ell, b_u) < \rho_2^*(\alpha, \eta), \tag{35}$$

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*

$$\mathcal{P}_n = \left\{P \in \mathcal{P} : \lambda_{u,b_u} - \lambda_{\ell,b_\ell} \geq \frac{\kappa_n}{\sqrt{n}}\right\}. \tag{36}$$

It holds that

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

$$\liminf_n \inf_{P\in\mathcal{P}_n} P\left(CI^m\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right) \subseteq CI^{sim}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha'\right)\right) = 1. \tag{37}$$

[9]Here $\hat{\rho}_\ell(b_1, b_2) = \hat{\rho}_u(b_1, b_2) = \hat{\rho}_{\ell u}(b_1, b_2)$, so I only impose restrictions on $\hat{\rho}_\ell$.

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*

$$\liminf_{a_n \to \infty} P_{a_n}\left(\theta_{a_n} \notin CI^m\left(\hat{\lambda}_{a_n}, \hat{\Sigma}_{a_n}/a_n; \alpha\right)\right) - P_{a_n}\left(\theta_{a_n} \notin CI^{sim}\left(\hat{\lambda}_{a_n}, \hat{\Sigma}_{a_n}/a_n; \alpha\right)\right) > 0. \tag{38}$$

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 Rambachan and Roth (2023). 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:

$$\begin{aligned} \rho_1^*(0.05, 0.04) &= 0.84, & \rho_1^*(0.10, 0.08) &= 0.83, \\ \rho_2^*(0.05, 0.001) &\approx 1, & \rho_2^*(0.10, 0.001) &\approx 1. \end{aligned}$$

The values are large and thus the restriction (34) 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).

Next, I compare the modified conditional CI with the adjusted bootstrap procedure proposed by Ye et al. (2023). Their bootstrap procedure relies on a random draw of a subsample with size $m = \frac{n}{\kappa_n}$, and thus the convergence rate of the confidence interval to the identified set is $\sqrt{m}$, slower than $\sqrt{n}$.

Theorem 3. *(Power Comparison with Ye et al. (2023)) 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*

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

Then

$$\liminf_{n \to \infty} \inf_{P \in \mathcal{P}} P\left(\theta_n \notin CI^m\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) = 1 \tag{39}$$

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

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 Inference with Infinite $\mathcal{B}$

In this section, I explore the inference procedure with potentially infinite $\mathcal{B}$. This approach has broad applications in counterfactual analysis within a structural model, where $\mathcal{B}$ represents the identified set of the structural parameters, and the target object θ is the counterfactual of interest.

The identified set of θ can be expressed as

$$\theta \in \bigcup_{\beta \in \mathcal{B}} \left[\lambda_{P,\ell}(\beta, \varphi_P),\ \lambda_{P,u}(\beta, \varphi_P)\right],$$

where $\mathcal{B} \subseteq \overline{\mathcal{B}}$ is formed by the moment inequalities:

$$E_P\left[m_j(X_i; \beta, \varphi_P)\right] \leq 0,\ j = 1, ..., J_1, \tag{40}$$

$$E_P\left[m_j(X_i; \beta, \varphi_P)\right] = 0,\ j = J_1 + 1, ..., J_1 + J_2, \tag{41}$$

as is often the case in empirical applications. Here, $\overline{\mathcal{B}}$ is the known parameter space of β, $\{X_i\}_{i=1}^n$ is an independent and identically distributed (i.i.d.) sequence of random variables with distribution P and $(m_1, ..., m_{J_1+J_2})' : \mathbb{R}^d \times \overline{\mathcal{B}} \times \Phi \to \mathbb{R}^{J_1+J_2}$ is a known measurable function of a finite dimensional parameter vector $(\beta, \varphi) \in \overline{\mathcal{B}} \times \Phi$. In turn, φ is a nuisance parameter with a point identified true value φ_P, which can be estimated in a preliminary step without using the moment inequalities in (40) and (41). Additionally, I assume that

$$\varphi_P = \varphi^\dagger\left(E_P\left[m_\varphi\left(X_i\right)\right]\right), \tag{42}$$

where m_φ and $\varphi^\dagger$ are known measurable functions. Given a potential true value β, the object θ is bounded by $\lambda_{P,\ell}(\beta, \varphi_P)$ and $\lambda_{P,u}(\beta, \varphi_P)$. For this, I assume that

$$\lambda_{P,k}(\beta, \varphi) = \lambda_k^\dagger\left(E_P\left[m_k\left(X_i; \beta, \varphi\right)\right]\right) \quad k = \ell, u, \tag{43}$$

where $m_k : \mathbb{R}^d \times \overline{\mathcal{B}} \times \Phi \to \mathbb{R}^{d_k}$, $\lambda_k^\dagger : \mathbb{R}^{d_k} \to \mathbb{R}$ are known and measurable functions. I give a detailed illustration in Section 5.2 based on Dickstein and Morales (2018).

It is important to note that (42) and (43) impose that estimands are functions of moments of the data. However, it suffices to have that the asymptotic distributions of the estimators for $(\lambda_{P,\ell}, \lambda_{P,u}, \varphi_P, E_P\left[m_j(X; \beta, \varphi)\right])$ can be easily approximated.

4.1 Notation

To begin, I define the estimators used for inference. I treat moment equalities as two opposing moment inequalities and define $J = J_1 + 2J_2$ moments by letting

$$m_{\mathcal{J}} = \left(m_1, ..., m_{J_1+J_2}, -m_{J_1+J_2+1}, ..., -m_{J_1+2J_2}\right).$$

Then, define

$$\begin{aligned}
\bar{m}_{\mathcal{J}}(\beta, \varphi) &= \frac{1}{n}\sum_{i=1}^{n} m_{\mathcal{J}}(X_i; \beta, \varphi), \\
\bar{m}_k(\beta, \varphi) &= \frac{1}{n}\sum_{i=1}^{n} m_k(X_i; \beta, \varphi), \quad k = \ell, u, \\
\bar{m}_{\varphi} &= \frac{1}{n}\sum_{i=1}^{n} m_{\varphi}(X_i).
\end{aligned}$$

Intuitive point estimators for $\lambda_{P,k}(\beta, \varphi_P)$, $k = \ell, u$, are given by

$$\hat{\lambda}_k(\beta) = \lambda_k^{\dagger}\left(\bar{m}_k(\beta, \hat{\varphi})\right), \tag{44}$$

with first step estimator $\hat{\varphi} = \varphi^{\dagger}(\bar{m}_{\varphi})$.

Next, I present the covariance matrix for the asymptotic distribution of

$$\mathbb{G}_n(\beta) = \sqrt{n}\begin{bmatrix} \bar{m}_{\mathcal{J}}(\beta, \hat{\varphi}) - E_P\left[m_{\mathcal{J}}(X; \beta, \varphi_P)\right] \\ \hat{\lambda}_\ell(\beta) - \lambda_{\ell,P}(\beta, \varphi_P) \\ \hat{\lambda}_u(\beta) - \lambda_{u,P}(\beta, \varphi_P) \end{bmatrix}. \tag{45}$$

Let

$$\Omega_P(\beta) = \text{var}_P\left(\begin{pmatrix} m'_{\mathcal{J}}(X; \beta, \varphi_P), & m'_\ell(X; \beta, \varphi_P), & m'_u(X; \beta, \varphi_P), & m'_\varphi(X) \end{pmatrix}'\right)$$

denote the covariance matrix for the moments evaluated at (β, φ_P). The variation of $\mathbb{G}_n(\beta)$ arises from two sources: (i) the use of sample averages for expectations, which comes from the variation in $m'_{\mathcal{J}}, m_\ell, m_u$; and (ii) the estimation uncertainty in $\hat{\varphi}$, which comes from the variation in m_φ. Define the Jacobian matrix of the estimator

$$\left(\bar{m}_{\mathcal{J}}(\beta, \hat{\varphi})', \hat{\lambda}_\ell(\beta), \hat{\lambda}_u(\beta)\right)'$$

with respect to the moments as

$$G_P(\beta) = \begin{bmatrix} \mathcal{I}_J, & 0, & 0, & \nabla_{\varphi'} E\left[m(X;\beta,\varphi_P)\right] \nabla'\varphi^\dagger \\ 0, & \nabla'\lambda_\ell^\dagger, & 0, & \nabla_{\varphi'}\lambda_{P,\ell}(\beta,\varphi_P)\nabla'\varphi^\dagger \\ 0, & 0, & \nabla'\lambda_u^\dagger, & \nabla_{\varphi'}\lambda_{P,u}(\beta,\varphi_P)\nabla'\varphi^\dagger \end{bmatrix}. \tag{46}$$

Then the first source of the variation of $\mathbb{G}_n(\beta)$ is captured by the first three block columns of G_P while the second one is captured by the last block column of G_P. In (46), $\nabla'\lambda_k^\dagger$ is short for $\nabla_{m'}\lambda_k^\dagger(m)\Big|_{E[m_k(\beta,\varphi_P)]}$ with $k=\ell,u$ and $\nabla'\varphi^\dagger$ is short for $\nabla_{m'}\varphi^\dagger(m)\big|_{E[m_\varphi(X)]}$. Under mild regularity conditions, the asymptotic variance of $\mathbb{G}_n(\beta)$ is given by

$$\Sigma_P(\beta) = G_P(\beta)\Omega_P(\beta)G'_P(\beta).$$

Throughout the paper, I assume that there exist consistent estimators $\hat{G}(\beta)$, $\hat{\Omega}(\beta)$ for $G_P(\beta)$, $\Omega_P(\beta)$, respectively. A computationally simple and intuitive estimator of $\hat{\Sigma}(\beta)$ is

$$\hat{\Sigma}(\beta) = \hat{G}(\beta)\hat{\Omega}(\beta)\hat{G}'(\beta).$$

To simplify notation, let

$$\hat{\sigma}_j(\beta) = \sqrt{\hat{\Sigma}_{jj}},\ \forall j=1,...,J, \qquad \hat{\sigma}_\ell(\beta) = \sqrt{\hat{\Sigma}_{J+1}(\beta)}, \qquad \hat{\sigma}_u(\beta) = \sqrt{\hat{\Sigma}_{J+2}(\beta)}, \tag{47}$$
$$\hat{\sigma}_{u\ell}(\beta) = \sqrt{\begin{bmatrix} 0_{1\times J}, & 1, & -1 \end{bmatrix} \hat{\Sigma}(\beta) \begin{bmatrix} 0_{1\times J}, & 1, & -1 \end{bmatrix}'}$$

where $\hat{\sigma}_{u\ell}(\beta)$ is the estimated variance of $\sqrt{n}\left(\hat{\lambda}_u(\beta) - \hat{\lambda}_\ell(\beta) - \lambda_{P,u}(\beta,\varphi_P) + \lambda_{P,\ell}(\beta,\varphi_P)\right)$. Define $\sigma_j(\beta)$, $\sigma_\ell(\beta)$, $\sigma_{u\ell}(\beta)$ as in (47) with $\hat{\Sigma}$ replaced by Σ.

Lastly, let

$$D_{P,k}(\beta) = \frac{\nabla_{\beta'}\lambda_{P,k}(\beta,\varphi_P)}{\sigma_{P,k}(\beta,\varphi_P)},\ k=\ell,u, \tag{48}$$
$$D_{P,j}(\beta) = \frac{\nabla_{\beta'}E\left[m_j(X;\beta,\varphi_P)\right]}{\sigma_{P,j}(\beta,\varphi_P)},\ j=1,...,J, \tag{49}$$

be the Jacobian of the objective and moment conditions with respect to the parameter β normalized by the standard deviation. Furthermore, assume the existence of consistent estimators $\hat{D}_k(\beta)$ and $\hat{D}_j(\beta)$.

4.2 The Confidence Interval

The confidence interval has the following form

$$CI_n = \left[\inf_{\beta \in \hat{\mathcal{B}}} \hat{\lambda}_\ell(\beta) - \frac{\hat{\sigma}_\ell(\beta)}{\sqrt{n}} \hat{c}_n(\beta), \quad \sup_{\beta \in \hat{\mathcal{B}}} \hat{\lambda}_u(\beta) + \frac{\hat{\sigma}_u(\beta)}{\sqrt{n}} \hat{c}_n(\beta) \right], \tag{50}$$

where $\hat{\mathcal{B}}$ is an estimator for $\mathcal{B}$ defined as

$$\hat{\mathcal{B}} = \left\{ \beta \in \overline{\mathcal{B}} : \frac{\sqrt{n} \bar{m}_j(\beta, \hat{\varphi})}{\hat{\sigma}_j(\beta)} \leq \hat{c}_n(\beta), \ \forall j = 1, ..., J \right\}, \tag{51}$$

and $\hat{c}(\beta)$ is the critical value specified below. The same critical value is used for both the moment restrictions and the bounds $\hat{\lambda}_\ell$, $\hat{\lambda}_u$. The main reason is that, for both the objective and moment conditions, the critical value provides an upper bound for a normal random variable with a non-positive expectation and unit variance, and consequently it is natural to use the same critical value.[10] In addition, it is computationally easier than calibrating multiple critical values.

Next, I define the critical value $\hat{c}_n(\beta)$. The critical value is given by

$$\hat{c}_n(\beta) = \max \left\{ \hat{c}_u(\beta), \ \hat{c}_\ell(\beta) \right\}, \tag{52}$$

where for $k = \ell, u$, $[\ell] = u$, $[u] = \ell$,

$$\hat{c}_k(\beta) = \inf \Bigg\{ c \in \mathbb{R}_+ : P \Bigg(\min_{\Delta \in \boldsymbol{\Delta}_n(\beta, c)} \mathbb{Z}^s_k(\beta) + \hat{D}_k(\beta)\Delta \leq c, \text{ and} \tag{53}$$

$$\min_{\Delta \in \boldsymbol{\Delta}_n(\beta, c)} \mathbb{Z}^s_{[k]}(\beta) + \hat{D}_{[k]}(\beta)\Delta + \hat{\zeta}_0(\beta) \leq c \Bigg| \{X_i\}_{i=1}^n \Bigg) \geq 1 - \alpha \Bigg\}, \tag{54}$$

and the random feasible set for Δ is

$$\begin{aligned} \boldsymbol{\Delta}_n(\beta, c) = \Big\{ &\Delta \in \sqrt{n}(\overline{\mathcal{B}} - \beta) \cap \rho \, [-1, 1]^d : \\ &\mathbb{Z}^s_j(\beta) + \hat{D}_j(\beta)\Delta + \hat{\zeta}_j(\beta) \leq c, j = 1, \ldots, J \Big\}. \end{aligned} \tag{55}$$

The elements in (53), (54), and (55) are listed below.

1. $\mathbb{Z}^s_{\mathcal{J}}(\cdot), \mathbb{Z}^s_\ell(\cdot), \mathbb{Z}^s_u(\cdot)$ is simulated from

$$\left(\mathbb{Z}^s_{\mathcal{J}}(\beta)', \mathbb{Z}^s_\ell(\beta), \mathbb{Z}^s_u(\beta) \right)' \Big| \{X_i\}_{i=1}^n \sim \mathcal{N} \left(0, \text{diag} \left(\hat{\Sigma}(\beta) \right)^{-1/2} \hat{\Sigma}(\beta) \, \text{diag} \left(\hat{\Sigma}(\beta) \right)^{-1/2} \right)$$

2. $\hat{\zeta}_j(\beta)$, $\hat{\zeta}_0(\beta)$ in (54) and (55) are generalized moment selection (GMS) type functions

[10] To see this, (60), (62), (63) have similar structure with non-positive last term $\frac{E[m_j(\beta, \varphi_P)]}{\sigma_{P,j}(\beta, \varphi_P)/\sqrt{n}}$, $\frac{\lambda_{P,\ell}(\beta) - \theta}{\sigma_{P,\ell}(\beta)/\sqrt{n}}$, $\frac{\theta - \lambda_{P,u}(\beta)}{\sigma_{P,u}(\beta)/\sqrt{n}}$ respectively.

defined as

$$\hat{\zeta}_j(\beta) = \begin{cases} 0 & \text{if } \frac{\sqrt{n}\bar{m}_j(\beta,\hat{\varphi})}{\kappa_n \hat{\sigma}_j(\beta)} \geq -1 \text{ or } j = J_1+1, ..., J \\ -\infty & \text{otherwise} \end{cases}$$

$$\hat{\zeta}_0(\beta) = \begin{cases} 0 & \text{if } \frac{\sqrt{n}\left(\hat{\lambda}_u(\beta)-\hat{\lambda}_\ell(\beta)\right)}{\kappa_n \max\{\hat{\sigma}_u(\beta),\hat{\sigma}_\ell(\beta),\hat{\sigma}_{u\ell}(\beta)\}} \leq 1 \\ -\infty & \text{otherwise} \end{cases}$$

where $\kappa_n = o(n^{\frac{1}{4}})$ and $\kappa_n \to \infty$, with a suggested value $\sqrt{\ln n}$. It is easy to see that if $\hat{\zeta}_0(\beta) = 0$, it holds numerically that $\hat{c}_\ell(\beta) = \hat{c}_u(\beta)$; and if $\hat{\zeta}_0(\beta) = -\infty$, the constraint with $\mathbb{Z}^s_{[k]}(\beta)$ in (54) is negligible.

3. $\rho > 0$ is a user-chosen tuning parameter with the recommended rule of thumb value

$$\rho = \Phi^{-1}\left(\frac{1}{2} + \frac{1}{2}\left(1 - \eta/\binom{J_1+J_2+2}{d}\right)^{1/d}\right) \tag{56}$$

with $\eta = 0.01$. The critical value $\hat{c}_n(\beta)$ is weakly decreasing in ρ, thus a larger ρ returns a shorter confidence interval. Nevertheless, uniform validity of inference requires $\rho < \infty$.

Note that (53), (54), (55) are linear in Δ. Therefore, with polyhedral $\overline{\mathcal{B}}$, which is a frequent scenario in practice, the computation of the critical value $\hat{c}(\beta)$ can be simplified to the process of solving linear programs.

The construction of the confidence interval uses the insights of Kaido et al. (2019), and the intuition is as follows. By definition, there is some "true value" $\beta \in \mathcal{B}$ such that

$$\theta \in \left[\lambda_{P,\ell}(\beta,\varphi_P), \lambda_{P,u}(\beta,\varphi_P)\right].$$

A sufficient condition for θ being covered is that there exists

$$\tilde{\beta} \in \overline{\mathcal{B}} \cap \left\{\beta + \frac{\rho}{\sqrt{n}}[-1,1]^d\right\} \tag{57}$$

such that (i) θ is covered

$$\hat{\lambda}_\ell(\tilde{\beta}) - \frac{\hat{\sigma}_\ell(\tilde{\beta})}{\sqrt{n}}\hat{c}_n(\tilde{\beta}) \leq \theta \leq \hat{\lambda}_u(\tilde{\beta}) + \frac{\hat{\sigma}_u(\tilde{\beta})}{\sqrt{n}}\hat{c}_n(\tilde{\beta}) \tag{58}$$

and (ii) the moment conditions are not rejected

$$\frac{\sqrt{n}\bar{m}_j(\tilde{\beta})}{\hat{\sigma}_j(\tilde{\beta})} \leq \hat{c}_n(\tilde{\beta}), \quad \forall j = 1, ..., J. \tag{59}$$

That said, we only need to calibrate $\hat{c}(\tilde{\beta})$ such that (58) and (59) hold with probability no less than $1-\alpha$ for some $\tilde{\beta}$ in the true value β's local neighborhood, without considering the coverage of β itself. This returns a much smaller critical value than uncalibrated projection, which covers the structural parameter β with prespecified probability in the first step.

In (57), $\tilde{\beta}$ is restricted to a local neighborhood of β, and the tuning parameter ρ defines its boundary. This allows us to linearize (58) and (59), so that they can be approximated by (53, 54) and (55), respectively. As for (59), a first-order Taylor expansion gives

$$\frac{\sqrt{n}\bar{m}_j(\tilde{\beta})}{\hat{\sigma}_j(\tilde{\beta})} \approx \frac{\mathbb{G}_{n,j}(\beta)}{\sigma_{P,j}(\beta)} + D_{P,j}(\beta)\Delta + \frac{\sqrt{n}E\left[m_j(\beta,\varphi_P)\right]}{\sigma_{P,j}(\beta)} \tag{60}$$

where $\Delta = \sqrt{n}(\tilde{\beta}-\beta)$, $\mathbb{G}_n$, D_P, and σ_P are defined in (45), (49) and below (47), respectively. In (55), I replace $\frac{\mathbb{G}_{n,j}(\beta)}{\sigma_{P,j}(\beta)}$ and $D_{P,j}(\beta)$ with their feasible analogs. In addition, it is well known that the last term $\frac{\sqrt{n}E[m_j(\beta,\varphi_P)]}{\sigma_{P,j}(\beta)}$ can not be consistently estimated, but can be conservatively approximated by the GMS function proposed in D. W. K. Andrews and Soares (2010). Therefore, (55) is a valid approximation of the moment restrictions in (59). Similarly, (58) is equivalent to

$$\frac{\hat{\lambda}_\ell(\tilde{\beta})-\theta}{\hat{\sigma}_\ell(\tilde{\beta})/\sqrt{n}} \leq \hat{c}(\tilde{\beta}), \qquad \frac{\theta-\hat{\lambda}_u(\tilde{\beta})}{\hat{\sigma}_u(\tilde{\beta})/\sqrt{n}} \leq \hat{c}_n(\tilde{\beta}). \tag{61}$$

Local approximations of the standardized $\hat{\lambda}_\ell(\beta)$, $\hat{\lambda}_u(\beta)$ are

$$\frac{\hat{\lambda}_\ell(\tilde{\beta})-\theta}{\hat{\sigma}_\ell(\tilde{\beta})/\sqrt{n}} \approx \frac{\mathbb{G}_{n,\ell}(\beta)}{\sigma_{P,\ell}(\beta)} + D_{P,\ell}(\beta)\Delta + \frac{\lambda_{P,\ell}(\beta)-\theta}{\sigma_{P,\ell}(\beta)/\sqrt{n}} \tag{62}$$

$$\frac{\theta-\hat{\lambda}_u(\tilde{\beta})}{\hat{\sigma}_u(\tilde{\beta})/\sqrt{n}} \approx \frac{\mathbb{G}_{n,u}(\beta)}{\sigma_{P,u}(\beta)} + D_{P,u}(\beta)\Delta + \frac{\theta-\lambda_{P,u}(\beta)}{\sigma_{P,u}(\beta)/\sqrt{n}}. \tag{63}$$

The approximations for (62, 63) are similar to the moment restrictions, except for the last term $\frac{\lambda_{P,\ell}(\beta)-\theta}{\sigma_{P,\ell}(\beta)/\sqrt{n}}$ and $\frac{\theta-\lambda_{P,u}(\beta)}{\sigma_{P,u}(\beta)/\sqrt{n}}$. Here, θ is unknown and partially identified, thus the GMS function does not directly apply. However, observe that if

$$\frac{\sqrt{n}\left(\lambda_{P,u}(\beta)-\lambda_{P,\ell}(\beta)\right)}{\kappa_n \max\left\{\sigma_{P,u}(\beta),\sigma_{P,\ell}(\beta),\sigma_{P,u\ell}(\beta)\right\}} \to \infty, \tag{64}$$

either $\frac{\lambda_{P,\ell}(\beta)-\theta}{\sigma_{P,\ell}(\beta)/\sqrt{n}}$ or $\frac{\theta-\lambda_{P,u}(\beta)}{\sigma_{P,u}(\beta)/\sqrt{n}}$ would go to $-\infty$ and at least one inequality in (61) is not binding with probability approaching one. Otherwise, we can replace both terms with zero if (64) does not hold, which introduces a conservative distortion, which in turn leads to the restriction in (53) and (54).

Remark 4. The construction of $\hat{c}_n$ uses the insights from Kaido et al. (2019) but with two main differences. First, Kaido et al. (2019), as well as many moment inequality papers, assume that all parameters are jointly estimated by a set of moment conditions, while in this paper, I allow for first-step plug-in estimator $\hat{\varphi}$ and suggest an easy adjustment for its estimation uncertainty. Second, the target object of Kaido et al. (2019) has form $\theta = \lambda(\beta)$ with a known function $\lambda(\cdot)$, while my paper assumes that $\theta \in [\lambda_\ell(\beta),\ \lambda_u(\beta)]$ with $\lambda_\ell(\cdot)$ and $\lambda_u(\cdot)$ estimated. In addition, the definition of Jacobian $D_{P,j}$ in (49) is different from Kaido et al. (2019), where $D_{P,j}^{KMS}(\cdot) = \nabla_{\beta'}\left\{E\left[m_j(X;\cdot)\right]/\sigma_{P,j}(\cdot)\right\}$. Note that under mild conditions, $D_{P,j}(\cdot)$ and $D_{P,j}^{KMS}(\cdot)$ are asymptotically equivalent for $\beta_n \in \overline{\mathcal{B}}$ such that the moment is close to binding, i.e. $E\left[m_j(X;\beta_n)\right] = o(1)$. If $E\left[m_j(X;\beta_n)\right]$ is bounded away from zero, $D_{P,j}(\cdot)$ and $D_{P,j}^{KMS}(\cdot)$ become irrelevant, since in this case, moment j is either slack or rejected with probability approaching one. I use the definition in (49) as it is easier to calculate. Moreover, the analogous Jacobian for λ_k defined in (48) follows directly from the Taylor expansion in (62)-(63), where the denominators are fixed. □

Remark 5. The construction of the CI relies on test inversion over the structural parameter space, which can be time-consuming. However, the E-A-M algorithm proposed by Kaido et al. (2019) can enhance computational efficiency with certain modifications. Refer to Appendix B.1 for further discussion. □

4.3 Asymptotic Results

I present next the assumptions that are important for proper coverage.

Let $\overline{\mathcal{B}}^{\varepsilon}$ be the ε expansion of $\overline{\mathcal{B}}$

$$\overline{\mathcal{B}}^{\varepsilon} = \left\{\beta \in \mathbb{R}^d : d(\beta, \overline{\mathcal{B}}) \le \varepsilon\right\}.$$

Assumption 6. *$\overline{\mathcal{B}} \subset \mathbb{R}^d$ is a compact hyperrectangle with nonempty interior.*

Assumption 6 restricts the shape of the parameter space. It is satisfied in most applications and guarantees that the calculation in (53) can be obtained with simple linear programs.

Assumption 7. *Let*

$$D_P = \left[\begin{array}{cccccc} D_{P,1}(\beta)' & \ldots & D_{P,J}(\beta)' & D_{P,\ell}(\beta)' & D_{P,u}(\beta)' \end{array}\right]'.$$

There is $M < \infty$, $\varepsilon > 0$ such that for all $P \in \mathcal{P}$, for all $\beta, \tilde{\beta} \in \overline{\mathcal{B}}^{\varepsilon}$

$$\begin{aligned} \left\|G_P(\beta) - G_P(\tilde{\beta})\right\| &= M\left\|\beta - \tilde{\beta}\right\|, \\ \left\|D_P(\beta) - D_P(\tilde{\beta})\right\| &\le M\left\|\beta - \tilde{\beta}\right\|, \\ \|G_P(\beta)\| \le M, \qquad &\|D_P(\beta)\| \le M. \end{aligned}$$

Assumption 8. *The following conditions hold uniformly over $\mathcal{P}$.*

1. *There exists estimator* $\hat{D}(\beta)$ *such that*

$$\sup_{\beta\in\overline{\mathcal{B}}^{\epsilon}}\left\|\hat{D}(\beta)-D_P(\beta)\right\|=o_P(1).$$

2. *For all* $\varepsilon>0$, *the estimator* $\hat{G}(\beta)$ *satisfies*

$$\liminf_{n}\inf_{P\in\mathcal{P}}P\left(\sup_{m\geq n}\sup_{\beta\in\overline{\mathcal{B}}}\left\|\hat{G}(\beta)-G_P(\beta)\right\|\leq\epsilon\right)=1.$$

Assumption 7 imposes that the derivatives D_P and G_P are sufficiently smooth in their arguments, and Assumption 8 requires consistent estimators for the Jacobian D_P and G_P. Under these two assumptions, we can approximate the Taylor expansion in (60), (62), (63) with feasible analogs.

Assumption 9. *For some constants* $\omega>0$, $\varepsilon>0$, *all distributions* $P\in\mathcal{P}$ *satisfy the following condition. Let*

$$\mathcal{J}_1(P,\beta;\varepsilon)=\left\{j\in\{1,\cdots,J_1\}:\frac{E_P\left[m_j\left(X_i;\beta,\varphi_P\right)\right]}{\sigma_{P,j}(\beta,\varphi_P)}\geq-\varepsilon\right\},\tag{65}$$

$$\mathcal{J}_k(P,\beta;\varepsilon)=\begin{cases}\{u,\ell,J_1+1,...,J_1+J_2\} & \text{if } \tau_{0,P,\beta}\geq-\varepsilon\\ \{k,J_1+1,...,J_1+J_2\} & \text{otherwise}\end{cases}\tag{66}$$

where

$$\tau_{0,P,\beta}=\frac{\lambda_{P,\ell}\left(\beta,\varphi_P\right)-\lambda_{P,u}\left(\beta,\varphi_P\right)}{\max\left\{\sigma_{P,\ell}(\beta),\sigma_{P,u}(\beta),\sigma_{P,u\ell}(\beta)\right\}}.$$

Then for $k=\ell,u$,

$$\inf_{\beta\in\mathcal{B}}\operatorname{eig}\left(\Sigma_{\mathcal{J}_1(P,\beta;\varepsilon)\cup\mathcal{J}_k(P,\beta;\varepsilon)}\right)\geq\omega,$$

where $\operatorname{eig}(\Sigma)$ *is the smallest eigenvalue of* Σ.

Note that $\mathcal{J}_1(P,\beta;\varepsilon)\cup\mathcal{J}_k(P,\beta;\varepsilon)$ is the collection of moments that are close to binding. Assumption 9 imposes that the covariance matrix for those moments is non-singular. In Assumption 11 in Appendix D.2, I relax this assumption by allowing the covariance matrix of paired moments to have a singular limit at the cost that the sum of two paired moments should be non-positive for all samples. This is an analog of Kaido et al. (2019) Assumption E3.2.

Theorem 4. *Suppose Assumptions 7, 8, 6, 10 hold. In addition, suppose Assumption 9 or 11 holds. Let* $0<\alpha<1/2$. *Then*

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{P}}\inf_{\theta\in\bigcup_{\beta\in\mathcal{B}}\left[\lambda_{P,\ell}(\beta),\lambda_{P,u}(\beta)\right]}P\left(\theta\in CI_n\right)\geq1-\alpha.$$

Assumptions 10, which is a regularity condition for the moments, can be found in Appendix D.2.

5 Simulation

In this section, I study the size and power properties of the proposed procedures and compare them to several alternatives.

5.1 When $\mathcal{B}$ is Finite

When $\mathcal{B}$ is finite, I conduct simulations in the context of Example 1, i.e. relaxation of the parallel trends assumption as in Rambachan and Roth (2023). Besides the modified conditional confidence interval proposed in Section 3, I consider two existing procedures for union bounds: (i) the adjusted bootstrap in Ye et al. (2023), (ii) the simple confidence interval in (14), and (iii) the inference procedure in Rambachan and Roth (2023).[11] 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 all DGPs. In sum, I consider $4 \times 3 = 12$ empirically motivated DGPs. Note that the simulation results of the modified conditional CI, the Rambachan and Roth (2023) CI, and the simple CI are invariant to n, while YKHS depends on n because of a subsampling step. In this simulation, 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

[11] Ye et al. (2023) propose two CIs for the parameter of interest in their 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 Rambachan and Roth (2023), I use their hybrid conditional CI with tuning parameter $\eta = \frac{\alpha}{10}$, which is the default choice in their code.

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.

Rambachan and Roth (2023) CI is plotted in blue dashed curves. The performance of Rambachan and Roth (2023) 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, Rambachan and Roth (2023) may perform 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 Rambachan and Roth (2023) 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 and Theorem 3.

In Table 3 in Appnedix A.2, I report the median CIs.[12] I compare the differences between the length of median CIs and the length of the union bound estimates, as a measure of efficiency.[13] 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 Rambachan and Roth (2023) (resp. YKHS, simple CI) by a proportion up to 43% (resp. 32%, 37%).

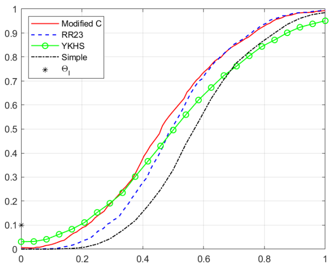
5.2 When $\mathcal{B}$ is Infinite

When $\mathcal{B}$ is infinite, I use the setting of the simulation in Dickstein and Morales (2018). Consider a simple trade model with $i = 1, ..., N$ firms deciding whether to export to a foreign market f, $f = 1, ..., K$. The revenues in home and foreign markets are determined by

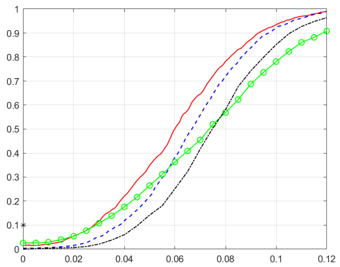
$$r_h = X_1 + X_2 + X_3,$$
$$r_f = \varphi_f r_h + e.$$

[12]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.

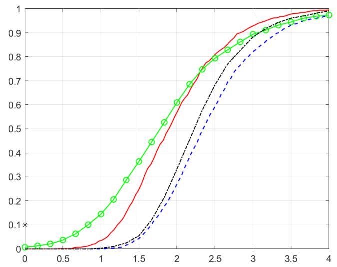
[13]The consistent union bound estimate is $\left[\min_{b\in\mathcal{B}} \hat{\lambda}_{\ell,b},\ \max_{b\in\mathcal{B}} \hat{\lambda}_{u,b}\right]$.



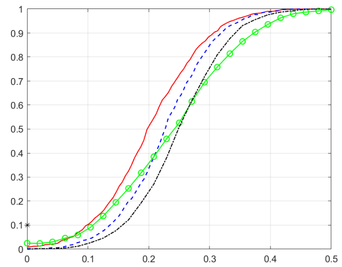
(a) Parallel Trends with Σ_1



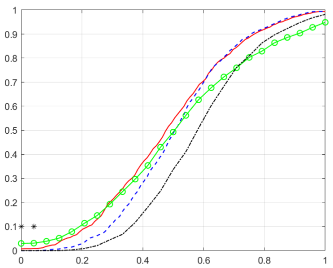
(b) Parallel Trends with Σ_2



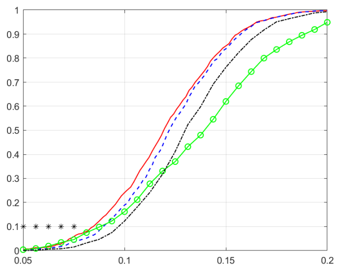
(c) Parallel Trends with Σ_3



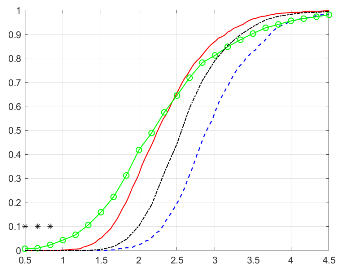
(d) Parallel Trends with Σ_4



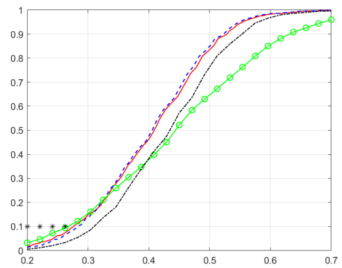
(e) Small Violation with (γ_1, Σ_1)



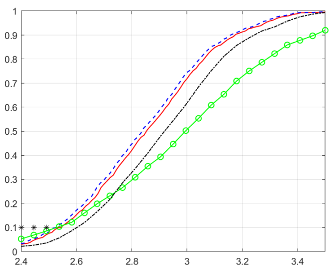
(f) Small Violation with (γ_2, Σ_2)



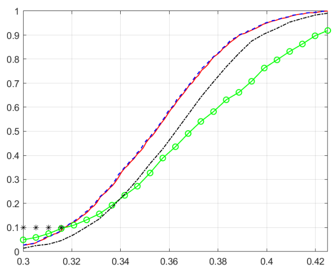
(g) Small Violation with (γ_3, Σ_3)



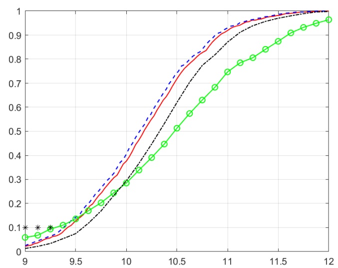
(h) Small Violation with (γ_4, Σ_4)



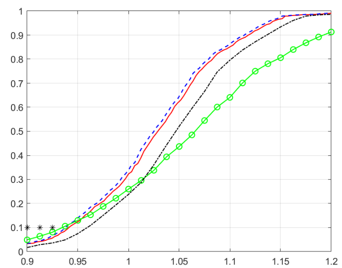
(i) Large Violation with Σ_1



(j) Large Violation with Σ_2



(k) Large Violation with Σ_3



(l) Large Violation with Σ_4

Figure 3: Simulation Results for fiinte $\mathcal{B}$ - Rejection Rate with $\alpha = 0.1$

The horizontal axis is θ, while the vertical axis is the rate that θ is not included in the CI. For instance, in 3a, point $(0.5, 0.33)$ on the black curve implies that $\theta = 0.5$ is not covered by the simple CI with probability 33%. Here (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 unknown $\mathcal{B}$ with $\alpha = 0.1$

# markets	Identified Set	Calibrated Projection CI (%)		Bonferroni CI (%)	
		Median CI	Coverage	Median CI	Coverage
$K = 1$	[2.344, 5.359]	[0.247, 10.176]	[99.8, 99.7]	[0.178, 11.457]	[100, 99.9]
$K = 4$	[2.344, 5.359]	[0.233, 10.226]	[99.8, 99.8]	[0.161, 11.559]	[100, 99.9]
$K = 8$	[2.344, 5.359]	[0.237, 10.356]	[99.9, 99.7]	[0.158, 11.682]	[100, 99.9]

The expected profit of export is given by

$$p_f = \eta^{-1} E\left[r_f \middle| \mathcal{J} \right] - \tilde{\beta}_2 - \tilde{\beta}_1 \nu,$$

where η is the elasticity, $\mathcal{J}$ is the information set the firms have when making the export decision, and $(X_1, X_2, X_3, \nu) \sim \mathcal{N}(0, \mathcal{I}_4)$. Then the decision is given by

$$\begin{aligned} d &= \mathbf{1}\left[\eta^{-1} E\left[r_f \middle| \mathcal{J} \right] \frac{1}{\tilde{\beta}_1} - \frac{\tilde{\beta}_2}{\tilde{\beta}_1} - \nu \geq 0 \right] \\ &= \mathbf{1}\left[\eta^{-1} E\left[\varphi_f r_h \middle| \mathcal{J} \right] \beta_1 - \beta_2 \geq \nu \right] \end{aligned}$$

where $\beta_1 = \tilde{\beta}_1^{-1}$, $\beta_2 = \tilde{\beta}_1^{-1} \tilde{\beta}_2$. This reparameterization is just for simplicity. The researcher observes (X, r_h, d, dr_f) and a subset of the information set $Z \subseteq \mathcal{J}$. The parameters are $(\varphi_f, \beta_1, \beta_2)$, and $\eta = 2$ is known to both the researcher and firms.[14] The counterfactual of interest is the change in the number of exporters if the information set changes from $\mathcal{J}_{\text{small}} = \{X_1\}$ to $\mathcal{J}_{\text{large}} = \{X_1, X_2, X_3\}$. The details of moment conditions and counterfactual outcomes are given in Appendix B.2.

In this simulation, I set the sample size $NK = 2000$, and the simulation is based on $S = 1000$ sample draws. The nominal rejection rate is $\alpha = 0.1$. The true value is $(\beta_1, \beta_2) = (1, 0.5)$ and $\varphi_f = 0.5$ for all $f = 1, ..., K$. In Table 1, I report the calibrated projection CI using the inference proposed in Section 4, as well as a Bonferroni type CI defined as

$$CI^{\text{Bon}} = \left[\min_{\beta \in \widetilde{\mathcal{B}}_{1-\frac{\alpha}{2}}} \hat{\lambda}_\ell(\beta, \hat{\varphi}) - \Phi^{-1}(1 - \frac{\alpha}{4}) \hat{\sigma}_\ell(\beta), \quad \max_{\beta \in \widetilde{\mathcal{B}}_{1-\frac{\alpha}{2}}} \hat{\lambda}_u(\beta, \hat{\varphi}) + \Phi^{-1}(1 - \frac{\alpha}{4}) \hat{\sigma}_u(\beta) \right],$$

where $\widetilde{\mathcal{B}}_{1-\frac{\alpha}{2}}$ is a $1 - \frac{\alpha}{2}$ confidence set of $\mathcal{B}$ calculated based on D. W. K. Andrews and Soares (2010). This Bonferroni CI is a valid alternative to the new CI, but less efficient, especially when the dimension of β is large. Both the Bonferroni CI and the calibrated projection CI have proper coverage, but the length of the calibrated CI is smaller than the Bonferroni CI.

[14] Note that this is in fact a normalization since η and β_1 cannot be identified separately.

6 Empirical Illustration

6.1 Sensitivity Analysis for Effects of the Minimum Wage

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 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

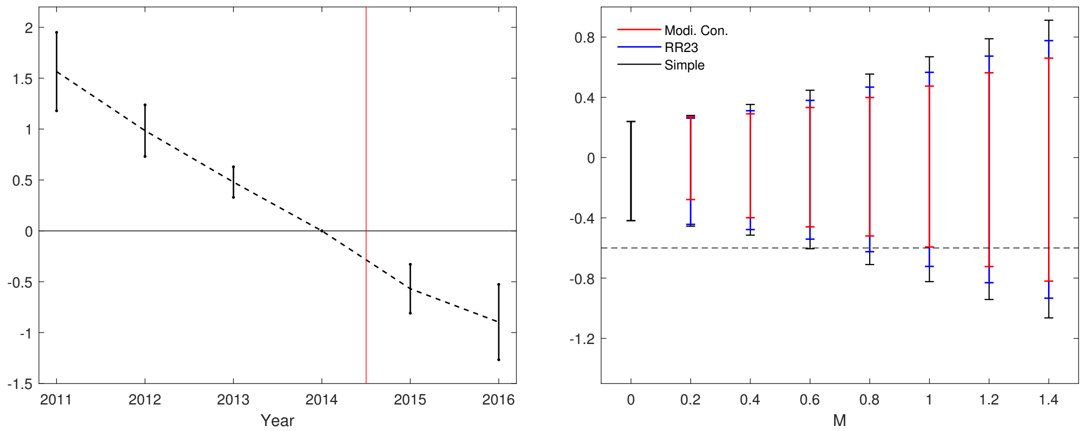
$$\log(\text{emp}_{rt}) = \sum_{\tau=2011,\tau\neq 2014}^{2016} \gamma_\tau \overline{GAP}_r \mathbf{1}\left[\tau = t\right] + \alpha_r + \xi_t + \varepsilon_{rt} \tag{67}$$

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 4a shows the estimated coefficients $\{\hat{\gamma}_\tau\}$ from specification (67). 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 4a 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 Rambachan and Roth (2023), 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 4a. The employment effect of interest is quantified as $\gamma_{2015} - \xi_{2015}$. That is, with one unit increase in GAP and other covariates fixed, the employment rate will increase by $100(\gamma_{2015} - \xi_{2015})\%$



(a) Estimated Employment Effect $\hat{\gamma}_\tau$ (b) Sensitivity Analysis under SDRM

Figure 4: Empirical Illustration for finite $\mathcal{B}$ with $\alpha = 0.05$

in expectation.

In Figure 4b, I report the 95% confidence intervals for different values of M constructed based on three different methods: the modified conditional CI proposed in Section 3, the hybrid CI in Rambachan and Roth (2023), and the simple CI in (14).[15] We can clearly 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 Rambachan and Roth (2023) 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 (67) 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 4b. 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.

6.2 Counterfactual Analysis for Exporter's Information Set

Dickstein and Morales (2018) study how the information possessed by potential exporters influences their export decisions. A challenge in modeling firms' decisions lies in the fact that these decisions are contingent upon firms' expectations of export profit, which are

[15] The estimated coefficient and covariance are available but the data for regression is confidential, thus I can not implement the Ye et al. (2023) bootstrap procedure.

rarely directly observable by researchers. Previous literature often makes the strong assumption that firms' expectations are rational and depend on a set of variables available in the data, which carries the risk of model misspecification. Different from earlier research, Dickstein and Morales (2018) do not require full knowledge of an exporter's information set. Instead, they specify only a subset of the variables that agents use to form their expectations. The trade-off of this approach is that it results in partial identification of the structural parameters and counterfactuals. The empirical results show that, on average, the number of exporters decreases when firms have access to better information. However, this change varies with firm size.

The model details are given in Appendix B.3. In this model, the structural parameter β is partially identified by moment inequalities as in (40) with $J_1 = 48$, $J_2 = 0$. The moment conditions contain 220 plug-in estimands $\varphi_{P,1}$,...,$\varphi_{P,220}$, with each corresponding to a specific market. The plug-in estimands are determined independently of the moment inequalities. The counterfactual of interest, denoted as θ, is the percentage change in the number of exporters if the information set changes from the smallest available to perfect foresight. The smallest information set includes the firm's own domestic sales in the previous year, sectoral aggregate exports in the previous year, and the distance variable. Given a specific structural parameter β, there exists a point-identified pair $\lambda_\ell(\beta, \varphi_P)$ and $\lambda_u(\beta, \varphi_P)$ that bounds the counterfactual outcome.

The original 95% CI in Dickstein and Morales (2018), denoted as CI^{DM18}, is computed as follows:

$$CI^{\text{DM18}} = \left[\min_{\beta \in \widetilde{\mathcal{B}}_{1-\alpha}} \hat{\lambda}_\ell(\beta), \quad \max_{\beta \in \widetilde{\mathcal{B}}_{1-\alpha}} \hat{\lambda}_u(\beta) \right].$$

Here $\widetilde{\mathcal{B}}_{1-\alpha}$ is a $1-\alpha$ confidence set for β, calculated based on D. W. K. Andrews and Soares (2010) treating the estimator $\hat{\varphi}$ as the true value. This interval is reported in the first row in Table 2, and it gives statistically significant results in all three subsamples. However, CI^{DM18} does not account for the estimation uncertainty in $\hat{\lambda}_\ell$, $\hat{\lambda}_u$ and $\hat{\varphi}$, making it potentially invalid. Additionally, it is unclear how CI^{DM18} compares to the new confidence interval. On the one hand, CI^{DM18} lacks validity and may be too narrow; on the other hand, it is a projection of the confidence set $\hat{\mathcal{B}}_{1-\alpha}$, which can be unnecessarily wide. To decompose these two effects, I first validate CI^{DM18} by appropriately adjusting for estimation uncertainty with

$$CI^{\text{DM18 adj}} = \left[\min_{\beta \in \hat{\mathcal{B}}_{1-\frac{\alpha}{2}}} \hat{\lambda}_\ell(\beta) - \Phi^{-1}(1 - \frac{\alpha}{4})\hat{\sigma}_\ell(\beta), \quad \max_{\beta \in \hat{\mathcal{B}}^a_{1-\frac{\alpha}{2}}} \hat{\lambda}_u(\beta) + \Phi^{-1}(1 - \frac{\alpha}{4})\hat{\sigma}_u(\beta) \right].$$

Here $\hat{\mathcal{B}}^a_{1-\frac{\alpha}{2}}$ is the $1 - \frac{\alpha}{2}$ confidence set of β that takes into account the estimation uncertainty in $\hat{\varphi}$. $CI^{\text{DM18 adj}}$ employs the same projection method used in Dickstein and Morales (2018), but with a Bonferroni-type adjustment, ensuring its validity. This inter-

Table 2: Percentage Change in Number of Exporters, $\alpha = 0.05$

Firm	All	Large	Small
DM18	[-10.2, -6.1]	[-17.3, -12.7]	[0.3, 0.5]
DM18 - Bonferroni ajusted	[-22.9, 1.5]	[-30.4, -0.8]	[-0.1, 1.4]
New	[-20.4, -1.6]	[-27.4, -8.7]	[0, 1.2]

val is reported in the second row. $CI^{\text{DM18 adj}}$ is considerably wider and crosses zero for all-firm and small-firm samples. In the third line, I report the new CI calculated by the calibrated projection method proposed in Section 4. The calibrated projection method, which is not only valid but also more efficient, is shorter than $CI^{\text{DM18 adj}}$ and restores statistical significance, signing the effect of the change in the number of exporters under perfect foresight.

7 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 truncated conditional critical value, which significantly improves upon existing procedures over a large set of DGPs. Empirical examples include sensitivity analysis in DiD and RDD, bunching strategies to identify the elasticity of taxable income, and misspecification in instrumental variable models. When the union is taken over an infinite set, I propose a calibrated projection CI, which is computationally attractive and applicable to structural counterfactual analysis in general moment inequality settings.

There are a few directions for future work. For finite $\mathcal{B}$, 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, for both finite and infinite $\mathcal{B}$, my inference procedures assume a correct specification that the union bound is non-empty. If the model is misspecified, the confidence interval 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 D. W. K. Andrews and Kwon (2023).

References

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

Andrews, D. W. K., & Guggenberger, P. (2010). Asymptotic size and a problem with subsampling and with the m out of n bootstrap. *Econometric Theory*, *26*(2), 426–468.

Andrews, D. W. K., & Kwon, S. (2023, March). Misspecified Moment Inequality Models: Inference and Diagnostics. *The Review of Economic Studies*.

Andrews, D. W. K., & Shi, X. (2013). Inference based on conditional moment inequalities. *Econometrica*, *81*(2), 609–666.

Andrews, D. W. K., & Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, *78*(1), 119–157.

Andrews, I., Kitagawa, T., & McCloskey, A. (2019). *Inference on winners* (Tech. Rep.). National Bureau of Economic Research.

Andrews, I., Kitagawa, T., & McCloskey, A. (2021). Inference after estimation of breaks. *Journal of Econometrics*, *224*(1), 39–59.

Andrews, I., & Mikusheva, A. (2016). Conditional inference with a functional nuisance parameter. *Econometrica*, *84*(4), 1571–1612.

Andrews, I., Roth, J., & Pakes, A. (2023). Inference for Linear Conditional Moment Inequalities. *The Review of Economic Studies*.

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

Ban, K., & Kedagni, D. (2022). Generalized Difference-in-differences Models: Robust Bounds. *arXiv preprint arXiv:2211.06710*.

Bei, X. (2024). Local linearization based subvector inference in moment inequality models. *Journal of Econometrics*, *238*(1), 105549.

Benzarti, Y., & Carloni, D. (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.

Berry, S., Eizenberg, A., & Waldfogel, J. (2016). Optimal product variety in radio markets. *The RAND Journal of Economics*, *47*(3), 463–497.

Blomquist, S., Newey, W. K., Kumar, A., & Liang, C.-Y. (2021). On bunching and identification of the taxable income elasticity. *Journal of Political Economy*, *129*(8), 2320–2343.

Bombardini, M., Li, B., & Trebbi, F. (2023). Did US Politicians Expect the China Shock? *American Economic Review*, *113*(1), 174–209.

Bugni, F. A., Canay, I. A., & Shi, X. (2015). Specification tests for partially identified models defined by moment inequalities. *Journal of Econometrics*, *185*(1), 259–282.

Bugni, F. A., Canay, I. A., & Shi, X. (2017). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics*, *8*(1), 1–38.

Casella, G., & Berger, R. L. (2021). *Statistical inference*. Cengage Learning.

Chen, X., Christensen, T. M., & Tamer, E. (2018). Monte Carlo confidence sets for identified sets. *Econometrica*, *86*(6), 1965–2018.

Chernozhukov, V., Hong, H., & Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models 1. *Econometrica*, *75*(5), 1243–1284.

Chernozhukov, V., Lee, S., & Rosen, A. M. (2013). Intersection bounds: estimation and inference. *Econometrica*, *81*(2), 667–737.

Chernozhukov, V., Newey, W. K., & Santos, A. (2023). Constrained conditional moment restriction models. *Econometrica*, *91*(2), 709–736.

Cho, J., & Russell, T. M. (2023). Simple Inference on Functionals of Set-Identified Parameters Defined by Linear Moments. *Journal of Business & Economic Statistics*, *0*(0), 1–16.

Christensen, P., Keiser, D. A., & Lade, G. E. (2023). Economic effects of environmental crises: Evidence from Flint, Michigan. *American Economic Journal: Economic Policy*, *15*(1), 196–232.

Ciliberto, F., Murry, C., & Tamer, E. (2021). Market structure and competition in airline markets. *Journal of Political Economy*, *129*(11), 2995–3038.

Conley, T. G., Hansen, C. B., & Rossi, P. E. (2012). Plausibly exogenous. *Review of Economics and Statistics*, *94*(1), 260–272.

Cox, G., & Shi, X. (2023). Simple adaptive size-exact testing for full-vector and subvector inference in moment inequality models. *The Review of Economic Studies*, *90*(1), 201–228.

Crawford, G. S., & Yurukoglu, A. (2012). The welfare effects of bundling in multichannel television markets. *American Economic Review*, *102*(2), 643–685.

Dickstein, M. J., & Morales, E. (2018). What do exporters know? *The Quarterly Journal of Economics*, *133*(4), 1753–1801.

Dustmann, C., Lindner, A., Schönberg, U., Umkehrer, M., & Vom Berge, P. (2022). Reallocation effects of the minimum wage. *The Quarterly Journal of Economics*, *137*(1), 267–328.

Eizenberg, A. (2014). Upstream innovation and product variety in the us home pc market. *Review of Economic Studies*, *81*(3), 1003–1045.

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

Fang, Z., & Santos, A. (2019). Inference on directionally differentiable functions. *The*

Review of Economic Studies, *86*(1), 377–412.

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

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

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

Imbens, G. W., & Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, *72*(6), 1845–1857.

Jia, P. (2008). What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica*, *76*(6), 1263–1316.

Johnson, C. R., & Horn, R. A. (1985). *Matrix analysis*. Cambridge university press Cambridge.

Kaido, H., Molinari, F., & Stoye, J. (2019). Confidence intervals for projections of partially identified parameters. *Econometrica*, *87*(4), 1397–1432.

Kalouptsidi, M., Kitamura, Y., Lima, L., & Souza-Rodrigues, E. (2021). *Counterfactual Analysis for Structural Dynamic Discrete Choice Models* (Tech. Rep.). Working paper, Harvard University.

Kireyev, P. (2020). Markets for ideas: Prize structure, entry limits, and the design of ideation contests. *The RAND Journal of Economics*, *51*(2), 563–588.

Kleibergen, F. (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica*, *73*(4), 1103–1123. (Publisher: Wiley Online Library)

Kolesár, M., & Rothe, C. (2018). Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, *108*(8), 2277–2304.

Lee, J. D., Sun, D. L., Sun, Y., & Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, 907–927.

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

Manski, C. F., & Pepper, J. V. (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, M. A., & Poirier, A. (2021). Salvaging falsified instrumental variable models. *Econometrica*, *89*(3), 1449–1469.

Mogstad, M., Santos, A., & Torgovitsky, A. (2016). Using instrumental variables for inference about policy relevant treatment parameters.

Mogstad, M., Santos, A., & Torgovitsky, A. (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica*, *86*(5), 1589–1619.

Molchanov, I. S., & Molchanov, I. S. (2005). *Theory of random sets* (Vol. 19) (No. 2). Springer.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica*, *71*(4), 1027–1048. (Publisher: Wiley Online Library)

Pata, V., & others. (2019). *Fixed point theorems and applications* (Vol. 116). Springer.

Ponomarev, K. (2022). Efficient Estimation of Directionally Differentiable Functionals.

Rambachan, A., & Roth, J. (2023, February). A More Credible Approach to Parallel Trends. *The Review of Economic Studies*, *90*(5), 2555–2591.

Rockafellar, R. (1970). *Convex Analysis princeton university press princeton.*

Rockafellar, R. T., & Wets, R. J.-B. (2009). *Variational analysis* (Vol. 317). Springer Science & Business Media.

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

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

Saez, E. (2010). Do taxpayers bunch at kink points? *American economic Journal: economic policy*, *2*(3), 180–212.

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

Stoye, J. (2020). *Simple misspecification adaptive inference for interval identified parameters* (Tech. Rep.). cemmap working paper.

Van Der Vaart, A. W., & Wellner, J. A. (1996). Weak convergence. In *Weak convergence and empirical processes* (pp. 16–28).

Wollmann, T. G. (2018). Trucks without bailouts: Equilibrium product characteristics for commercial vehicles. *American Economic Review*, *108*(6), 1364–1406.

Yang, C. (2020). Vertical structure and innovation: A study of the SoC and smartphone industries. *The RAND Journal of Economics*, *51*(3), 739–785.

Ye, T., Keele, L., Hasegawa, R., & Small, D. S. (2023). A negative correlation strategy for bracketing in difference-in-differences. *Journal of the American Statistical Association*(just-accepted), 1–24.

Appendix

A Additional Results for Finite $\mathcal{B}$

A.1 Implementation Details

Below I give a step by step implementation algorithm.

1. Input $\hat{\delta}_n$, $\hat{\Omega}_n$, α, α^c, A_ℓ, A_u, η, ε, where ε is the computational tolerence. I suggest using $\alpha^c = \frac{4}{5}\alpha$ and $\eta = 0.001$.

2. Construct a $1-\eta$ confidence set $\hat{\Delta}$ for δ:

$$\hat{\Lambda} = \left\{(A_\ell\delta,\ A_u\delta) \in \Lambda : \delta \in \hat{\Delta}\right\}, \tag{68}$$

$$\hat{\Delta} = \left\{\delta : \frac{|\hat{\delta}_j - \delta_j|}{\hat{\omega}_j} \leq \hat{c}_\eta, \forall j = 1, ..., J\right\} \tag{69}$$

   where $\hat{\omega}_j = \sqrt{\hat{\Omega}_{jj}}$ and

$$\hat{c}_\eta = Q\left(\max_{j=1,\ldots,J} |Z_j^*|, 1-\eta\right)$$

   with $Z^* \sim \mathcal{N}\left(0, \text{diag}(\hat{\Omega})^{-\frac{1}{2}}\hat{\Omega}\text{diag}(\hat{\Omega})^{-\frac{1}{2}}\right)$.

3. Calculate $\hat{c}^t$:

   (a) Initialize $\underline{c} = 0$, $\bar{c} = \Phi^{-1}(1-\frac{\alpha}{2})$.

   (b) Let $c = (\underline{c} + \bar{c})/2$ and calculate

$$p = \inf_{\delta \in \hat{\Delta}} \bar{p}\left(c, (A_\ell\delta, A_u\delta)\right).$$

   If $p < \alpha$, $\bar{c} = c$; if $p \geq \alpha$, $\underline{c} = c$.

   (c) If $\bar{c} - \underline{c} > \varepsilon$, go to 3b;
   If $\bar{c} - \underline{c} \leq \varepsilon$, $\hat{c}^t = (\underline{c} + \bar{c})/2$.

4. Construct the confidence interval by a grid search over an outer set of CI^{h} given in (10).

   (a) Initialize $\theta = \min_{b\in\mathcal{B}_\ell} \hat{\lambda}_{\ell,b} - \Phi^{-1}(1-\eta)\hat{\sigma}_{\ell,b}$.

   (b) Calculate $\hat{T}(\theta)$ and $\hat{c}^{\text{h}}(\theta;\alpha)$.

   (c) If $\hat{T}(\theta) > \hat{c}^{\text{m}}(\theta;\alpha)$, let $\theta = \theta + \varepsilon$.

      - If $\theta \leq \max_{b\in\mathcal{B}_u} \hat{\lambda}_{u,b} + \Phi^{-1}(1-\eta)\hat{\sigma}_{u,b}$ Go to Step 4b
      - If $\theta > \max_{b\in\mathcal{B}_u} \hat{\lambda}_{u,b} + \Phi^{-1}(1-\eta)\hat{\sigma}_{u,b}$. The confidence interval is $CI^{\text{m}} = \emptyset$, exit the algorithm.

   If $\hat{T}(\theta) \leq \hat{c}^{\text{m}}(\theta;\alpha)$, $\theta_1 = \theta$.

(d) Initialize $\theta = \max_{b \in \mathcal{B}_u} \hat{\lambda}_{u,b} + \Phi^{-1}(1-\eta)\hat{\sigma}_{u,b}$.

(e) Calculate $\hat{T}(\theta)$ and $\hat{c}^{\mathrm{m}}(\theta; \alpha)$.

(f) If $\hat{T}(\theta) > \hat{c}^{\mathrm{m}}(\theta; \alpha)$, $\theta = \theta - \varepsilon$. Go to Step 4e
If $\hat{T}(\theta) \leq \hat{c}^{\mathrm{m}}(\theta; \alpha)$, $\theta_2 = \theta$.

5. The confidence interval is $CI^{\mathrm{m}} = [\theta_1, \theta_2]$.

It is possible for the confidence interval to be empty, indicated by an output of $\emptyset$. This can occur when the model is misspecified and the lower bound exceeds the upper bound. However, if the realized minimizer is consistently smaller than the maximizer, which is the case in Rambachan and Roth (2023), Kolesár and Rothe (2018), and Masten and Poirier (2021), the modified conditional confidence interval is non-empty. It would also be interesting to consider misspecification robust inference for general union bounds, in the spirit of Stoye (2020) and I. Andrews et al. (2019), but this is outside the scope of this paper and I leave it for future research.

A.2 More Simulation Results

Median Confidence Intervals

In Table 3, I report the median CIs.[16] In the second row of each panel, I report the differences between the length of median CIs and the length of median point estimates, i.e. $\left[\max_{b \in \mathcal{B}} \hat{\lambda}_{\ell,b},\ \max_{b \in \mathcal{B}} \hat{\lambda}_{u,b}\right]$. The difference is a measure of the efficiency, net of the effect of identified set length. The difference of the modified conditional CI is the shortest, or slightly larger than the shortest, in all DGPs. It reduced the value of Rambachan and Roth (2023) (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).

Different Values of α^{c}

In Figure 5, I report the rejection rate with three different values of $\alpha^{\mathrm{c}} \in (\frac{\alpha}{2}, \alpha)$. The rejection rates are not overly sensitive to the tuning parameter α^{c}, and under all choice of α^{c}, the rejection rate of modified conditional CI is higher than the simple CI. We can also see that α^{c} trades off the power under more and less favorable DGPs. When the bounds are well separated, e.g. Figure 5i - 5l, the rejection rate increases in α^{c}. In contrast, when the bounds are close to each others, e.g. Figure 5c and 5g, smaller α^{c} gives higher power for alternatives closer to the null.

Different Sample Sizes

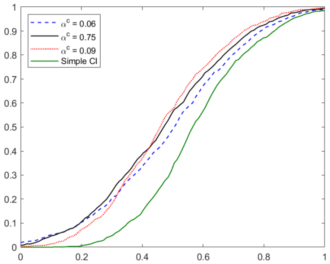
The sample is simulated from

$$Y_{it} = \sum_{s \neq 0} \xi_s \mathbf{1}[t = s] D_i + \xi_{X,1} X_{1it} + \xi_{X,2} X_{2it} + \varepsilon_{it}, \tag{70}$$

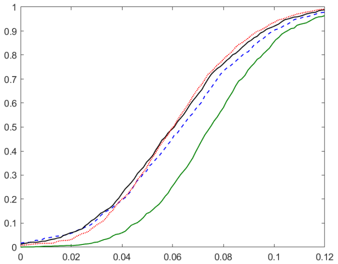
[16] A median CI is the median lower bound of the $1-\alpha$ CI to the median upper bound, and the median is taken over S samples.

Table 3: Simulation Results for known $\mathcal{B}$ - Median CI

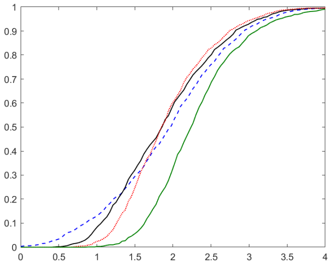
		Point	Modi. Con.	RR23	YKHS	Simple
		Dustmann et al. (2022) $\underline{T} = 3$				
Parallel	CI	[-0.188, 0.188]	[-0.470, 0.456]	[-0.505, 0.492]	[-0.486, 0.502]	[-0.578,0.565]
$[0, 0]$	Diff.		0.550	0.621	0.612	0.768
Small Vio.	CI	[-0.196, 0.195]	[-0.467, 0.475]	[-0.504, 0.505]	[-0.492, 0.504]	[-0.576,0.581]
$[-0.080, 0.080]$	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



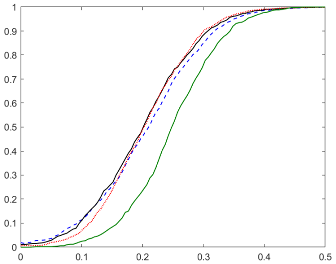
(a) Parallel Trends with Σ_1



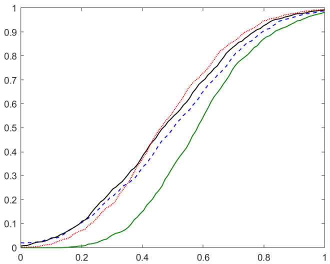
(b) Parallel Trends with Σ_2



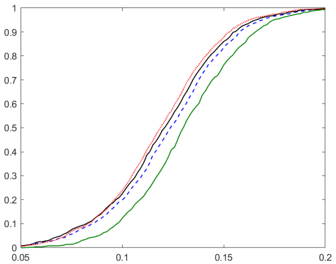
(c) Parallel Trends with Σ_3



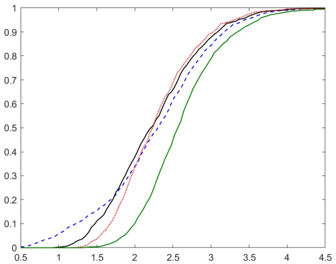
(d) Parallel Trends with Σ_4



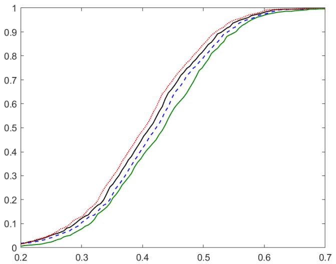
(e) Small Violation with (γ_1, Σ_1)



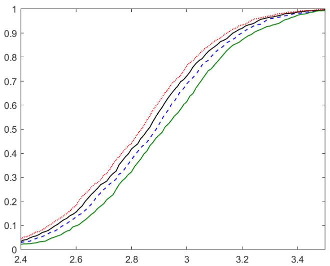
(f) Small Violation with (γ_2, Σ_2)



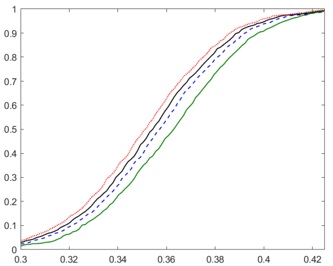
(g) Small Violation with (γ_3, Σ_3)



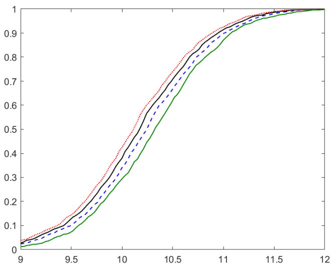
(h) Small Violation with (γ_4, Σ_4)



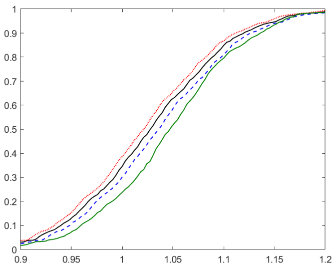
(i) Large Violation with Σ_1



(j) Large Violation with Σ_2



(k) Large Violation with Σ_3



(l) Large Violation with Σ_4

Figure 5: Rejection Rate Different α^c

The horizontal axis is θ, while the vertical axis is the rate that θ is not included in the CI. Here (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 4: Rejection Rate (%) Under Different Sample Sizes $\alpha = 0.1$

ξ	Parallel Trends				Small Violation				Large Violation			
Σ	Σ_1	Σ_2	Σ_3	Σ_4	Σ_1	Σ_2	Σ_3	Σ_4	Σ_1	Σ_2	Σ_3	Σ_4
$n = 125$	1.3	2.2	0	0.9	3.4	3.7	5.5	11.3	10.3	11.9	9	10
$n = 500$	1.2	1.5	0	1.1	4	5.8	8.8	9.2	8.5	8.6	8.6	9.4
$n = 2000$	1.6	2	0	1.1	8.3	8.9	9.4	8.4	9.2	8.6	9.7	9.8

with $i = 1, ..., n$, $t = 1, ..., T$. Let $E[D] = r = 0.6$, $X_{1,it} \sim \mathcal{N}(0,1)$, $X_{2,it} \sim t(5)$, $\varepsilon_i \sim \mathcal{N}(0, \Sigma_\varepsilon)$ and $\Sigma_\varepsilon = \Sigma/\Sigma_{TT} r$. $E[\xi]$ and Σ are defined as in Section 5.1, and $E[\xi_X] = 0$. I consider three sample sizes, $n = 125, 500, 2000$. The maximum rejection rate of θ in the identified set is reported in Table 4. There are only slight over rejection with $n = 125$.

A.3 Potentially Non Connected Union Bounds

In this section, I illustrate how to apply the modified conditional inference idea to potentially non-connected union bounds defined by

$$\theta \in \bigcup_{b \in \mathcal{B}} [\lambda_{\ell,b},\ \lambda_{u,b}] .$$

The confidence interval is constructed by inverting the test of the hypothesis

$$H_0 : \min_{b \in \mathcal{B}} \max \{\lambda_{\ell,b} - \theta,\ \theta - \lambda_{u,b}\} \leq 0.$$

The corresponding $1 - \alpha$ confidence interval is

$$CI^{\mathrm{m}}(\hat{\lambda}_n, \Sigma_n; \alpha) = \left\{\theta : \hat{T}(\theta) \leq \hat{c}^{\mathrm{m}}(\theta; \alpha)\right\} .$$

I illustrate with a normally distributed estimator $\hat{\lambda}_n = \left(\hat{\lambda}_\ell, \hat{\lambda}_u\right)$ satisfying (8). To simplify notation, assume that $\theta \in [\lambda_{\ell,b^*},\ \lambda_{u,b^*}]$.

The test statistic is defined as

$$\hat{T}(\theta) = \min_{b \in \mathcal{B}} \max \left\{ \frac{\hat{\lambda}_{\ell,b} - \theta}{\sigma_{\ell,b}},\ \frac{\theta - \hat{\lambda}_{u,b}}{\sigma_{u,b}} \right\} .$$

The simple CI uses the same test statistic and a simple critical value $c^{\mathrm{sim}} = \Phi^{-1}(1 - \frac{\alpha}{2})$, which gives confidence interval

$$CI^{\mathrm{sim}} = \bigcup_{b \in \mathcal{B}} \left[\hat{\lambda}_{\ell,b} - \sigma_{\ell,b} c^{\mathrm{sim}},\ \hat{\lambda}_{u,b} + \sigma_{u,b} c^{\mathrm{sim}}\right] .$$

The simple critical value is conservative because

$$P\left(\hat{T}(\theta) > \Phi^{-1}(1 - \frac{\alpha}{2})\right)$$

$$
\begin{aligned}
=&P\left(\min_{b\in\mathcal{B}}\max\left\{\frac{\hat{\lambda}_{\ell,b}-\theta}{\sigma_{\ell,b}},\ \frac{\theta-\hat{\lambda}_{u,b}}{\sigma_{u,b}}\right\}>\Phi^{-1}(1-\frac{\alpha}{2})\right)\\
\leq&P\left(\max\left\{\frac{\hat{\lambda}_{\ell,b^*}-\theta}{\sigma_{\ell,b^*}},\ \frac{\theta-\hat{\lambda}_{u,b^*}}{\sigma_{u,b^*}}\right\}>\Phi^{-1}(1-\frac{\alpha}{2})\right)\\
\leq&P\left(\frac{\hat{\lambda}_{\ell,b^*}-\theta}{\sigma_{\ell,b}}>\Phi^{-1}(1-\frac{\alpha}{2})\right)+P\left(\frac{\theta-\hat{\lambda}_{u,b^*}}{\sigma_{u,b}}>\Phi^{-1}(1-\frac{\alpha}{2})\right)\\
\leq&\frac{\alpha}{2}+\frac{\alpha}{2}\leq\alpha
\end{aligned}
$$

The first inequality is conservative if the bounds are not well separated, i.e. the Hausdorff distance between bounds is much larger than the standard deviations. The third inequality is conservative if the bound is relatively large. However, c^{sim} is not overly conservative under the less favorable DGPs where the bounds are well separated and short, i.e.

$$
\min_{b\in\mathcal{B}\setminus\{b^*\}} d\left([\lambda_{\ell,b^*},\lambda_{u,b^*}],[\lambda_{\ell,b},\lambda_{u,b}]\right)\gg\max_{b\in\mathcal{B}}\{\sigma_{\ell,b},\sigma_{u,b}\},\qquad \lambda_{\ell,b^*}\approx\lambda_{u,b^*}. \tag{71}
$$

Under the less favorable DGPs in (71), the probability

$$
P\left(\hat{T}(\theta)=\max\left\{\frac{\hat{\lambda}_{\ell,b^*}-\theta}{\sigma_{\ell,b^*}},\ \frac{\theta-\hat{\lambda}_{u,b^*}}{\sigma_{u,b^*}}\right\}\right)
$$

is close to one and therefore, I will construct the conditional critical value based on

$$
T(\theta)\,\Big|\,T(\theta)=\hat{\mathcal{Z}}_{\ell,b^*}\ \text{ or } T(\theta)\,\Big|\,T(\theta)=\hat{\mathcal{Z}}_{u,b^*}\ .
$$

Specifically, let

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

where

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

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

$$
\begin{aligned}
t_{\ell,2}(\theta,b)=\min\Bigg\{&\min_{\tilde{b}\in\mathcal{B}_\ell(b)\setminus\mathcal{B}_{\ell u}(b)}\frac{\mathcal{Z}_{\ell,\tilde{b}}-\rho_\ell(b,\tilde{b})\mathcal{Z}_{\ell,b}}{1-\rho_\ell(b,\tilde{b})},\\
&\min_{\tilde{b}\in\mathcal{B}_{\ell u}(b)}\max\left\{\frac{\mathcal{Z}_{\ell,\tilde{b}}-\rho_\ell(b,\tilde{b})\mathcal{Z}_{\ell,b}}{1-\rho_\ell(b,\tilde{b})},\ \frac{\mathcal{Z}_{u,\tilde{b}}+\rho_{\ell u}(b,\tilde{b})\mathcal{Z}_{\ell,b}}{1+\rho_{\ell u}(b,\tilde{b})}\right\}\Bigg\}
\end{aligned}
$$

$$t_{u,2}(\theta,b) = \min\left\{ \min_{\tilde{b}\in\mathcal{B}_u(b)\setminus\mathcal{B}_{u\ell}(b)} \frac{\mathcal{Z}_{u,\tilde{b}} - \rho_u(b,\tilde{b})\mathcal{Z}_{u,b}}{1-\rho_u(b,\tilde{b})} \right.$$
$$\left. \min_{\tilde{b}\in\mathcal{B}_{u\ell}(b)} \max\left\{ \frac{\mathcal{Z}_{u,\tilde{b}} - \rho_u(b,\tilde{b})\mathcal{Z}_{u,b}}{1-\rho_u(b,\tilde{b})}, \frac{\mathcal{Z}_{\ell,\tilde{b}} + \rho_{u\ell}(\tilde{b},b)\mathcal{Z}_{u,b}}{1+\rho_{u\ell}(\tilde{b},b)} \right\}\right\}$$

with

$$\mathcal{B}_\ell(b) = \left\{\tilde{b}\in\mathcal{B} : \rho_\ell(b,\tilde{b}) = 1\right\},$$
$$\mathcal{B}_u(b) = \left\{\tilde{b}\in\mathcal{B} : \rho_u(b,\tilde{b}) = 1\right\},$$
$$\mathcal{B}_{\ell u}(b) = \left\{\tilde{b}\in\hat{\mathcal{B}}_\ell : \rho_{\ell u}(b,\tilde{b}) = -1\right\},$$
$$\mathcal{B}_{u\ell}(b) = \left\{\tilde{b}\in\hat{\mathcal{B}}_u : \rho_{\ell u}(\tilde{b},b) = -1\right\}.$$

The modified conditional critical value is defined as

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

The lower truncation $\hat{c}^{\mathrm{t}}$ as the minimum value that achieves uniform size control, i.e.

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

where Λ_0 is the set of feasible λ satisfying H_0:

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

A.4 Violation of Assumption 4

As previously mentioned, when Assumption 4 fails, we can rewrite the union bounds as the union of several sub-union bounds, with Assumption 4 holding in each sub-union bound. Then we can apply the union principle by taking the union of CIs for each sub-union bound to get a valid CI for θ.

Example 9. Assume that $\mathcal{B} = \{1,2\}$, $A_\ell = \begin{bmatrix} \mathcal{I} & 0_{2\times 2}\end{bmatrix}$, $A_u = \begin{bmatrix} 0_{2\times 2} & \mathcal{I}\end{bmatrix}$, and

$$\Omega_n = \begin{bmatrix} 1 & 1-\frac{1}{n} & \\ 1-\frac{1}{n} & 1 & \\ & & \mathcal{I}_{2\times 2}\end{bmatrix}.$$

In this case, the limit of Ω_n is singular and consequently Assumption 4 fails. However, we can write

$$\theta \in [\min\{\lambda_{\ell,1},\lambda_{\ell,2}\},\ \max\{\lambda_{u,1},\lambda_{u,2}\}] = \Theta_1 \cup \Theta_2$$

where $\delta^1 = (\delta_1, \delta_3, \delta_4)'$, $\delta^2 = (\delta_2, \delta_3, \delta_4)'$, $\Omega_n^1 = \Omega_n^2 = \mathcal{I}_3$,

$$\Theta_1 = [\lambda_{\ell,1},\ \max\{\lambda_{u,1}, \lambda_{u,2}\}] = \left[\min_{b\in\mathcal{B}} \tilde{A}_{\ell,b}\delta^1,\ \max_{b\in\mathcal{B}} \tilde{A}_{u,b}\delta^1\right],$$
$$\Theta_2 = [\lambda_{\ell,2},\ \max\{\lambda_{u,1}, \lambda_{u,2}\}] = \left[\min_{b\in\mathcal{B}} \tilde{A}_{\ell,b}\delta^2,\ \max_{b\in\mathcal{B}} \tilde{A}_{u,b}\delta^2\right],$$
$$\tilde{A}_\ell = \left[\begin{array}{cc} 1_{2\times 1} & 0_{2\times 2} \end{array}\right], \qquad \tilde{A}_u = \left[\begin{array}{cc} 0_{2\times 1} & \mathcal{I}_2 \end{array}\right].$$

Therefore, Assumption 4 holds for Θ_1 and Θ_2, and we can get uniformly valid $1-\alpha$ CI $\widetilde{CI}^{\mathrm{m},1}$, $\widetilde{CI}^{\mathrm{m},2}$ separately. Then, it is easy to verify that $\widetilde{CI}^{\mathrm{m},1} \cup \widetilde{CI}^{\mathrm{m},2}$ is a uniformly valid $1-\alpha$ CI for θ, though at the cost of a potential efficiency loss. □

A.5 Union Bounds in Rambachan and Roth (2023)

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

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

The target object $\theta = \iota'\tau$ is the weighted average of ATT of post policy years, and ξ is a bias from a difference in trend. Here $\xi^{pre} = \left(\xi^{pre}_{-\underline{T}}, ..., \xi^{pre}_{-1}\right)$, $\xi^{post} = \left(\xi^{post}_1, ..., \xi^{post}_{\overline{T}}\right)$ and $\gamma_0 = \xi^{pre}_0$ is normalized to zero. In this section, I show that under relative magnitude relaxation and second differences relative magnitude relaxation, the identified set of the target object is a union bound.

Relative Magnitudes

Under the relative magnitude relaxation, we assume that the violation of parallel trends at time $t \geq 1$ is bounded above by the maximum pre-policy trend difference

$$\left|\xi^{post}_t - \xi^{post}_{t-1}\right| \leq M \max_{s=-1,\ldots,-\underline{T}} \left|\xi^{pre}_{s+1} - \xi^{pre}_s\right|, \tag{72}$$

Note that

$$\xi^{post} = \begin{bmatrix} \xi^{post}_1 \\ \xi^{post}_2 - \xi^{post}_1 + \xi^{post}_1 \\ \vdots \\ \xi^{post}_{\overline{T}} - \xi^{post}_{\overline{T}-1} + \cdots + \xi^{post}_2 - \xi^{post}_1 + \xi^{post}_1 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ & & \ddots & \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{pmatrix} \xi^{post}_1 \\ \xi^{post}_2 - \xi^{post}_1 \\ \vdots \\ \xi^{post}_{\overline{T}} - \xi^{post}_{\overline{T}-1} \end{pmatrix} = L\Delta^{post}$$

where

$$L = \underbrace{\begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ & & \ddots & \\ 1 & 1 & \cdots & 1 \end{bmatrix}}_{\overline{T}\times\overline{T}}, \Delta^{post} = \begin{pmatrix} \xi^{post}_1 \\ \xi^{post}_2 - \xi^{post}_1 \\ \vdots \\ \xi^{post}_{\overline{T}} - \xi^{post}_{\overline{T}-1} \end{pmatrix}.$$

The parameter of interest is

$$\begin{aligned}\iota'\left(\gamma^{post}-\xi^{post}\right) &= \iota'\gamma^{post}-M\iota' L\Delta^{post}\\ &\in\left[\iota'\gamma^{post}-M\left|\iota' L\right|\mathbf{1}_{\overline{T}\times 1}\bar{\Delta},\quad \iota'\gamma^{post}+M\left|\iota' L\right|\mathbf{1}_{\overline{T}\times 1}\bar{\Delta}\right],\end{aligned}$$

where $|\iota' L|$ is the vector of absolute value of each element in $\iota' L$, where

$$\bar{\Delta}=\max_{s=-1,\ldots,-\underline{T}}\left|\xi_{s+1}^{pre}-\xi_{s}^{pre}\right|.$$

Let

$$\Delta^{pre}=\begin{pmatrix}\xi_{1-\underline{T}}^{pre}-\xi_{-\underline{T}}^{pre}\\ \vdots\\ \xi_{-1}^{pre}-\xi_{-2}^{pre}\\ \xi_{0}^{pre}-\xi_{-1}^{pre}\end{pmatrix},\qquad \underbrace{\delta}_{\underline{T}+1}=\begin{pmatrix}\iota'\gamma^{post}\\ \Delta^{pre}\end{pmatrix}.$$

And we have

$$\theta\in\left[\min_{b=1,..,2\underline{T}}\lambda_{\ell,b},\quad \max_{b=1,..,2\underline{T}}\lambda_{u,b}\right]$$

where

$$\begin{aligned}\lambda_{\ell}=\lambda_{u}&=A\delta,\\ A&=\left[\begin{array}{ll}\mathbf{1}_{\underline{T}\times 1}, & M\left|\iota' L\right|\mathbf{1}_{\overline{T}\times 1}\mathcal{I}_{\underline{T}}\\ \mathbf{1}_{\underline{T}\times 1}, & -M\left|\iota' L\right|\mathbf{1}_{\overline{T}\times 1}\mathcal{I}_{\underline{T}}\end{array}\right].\end{aligned}$$

Second Differences Relative Magnitudes

Under the second difference relative magnitude relaxation, we assume that the violation of linear trend at time $t\geq 1$ is bounded above by the maximum pre-policy linear trend violation

$$\left|(\xi_t-\xi_{t-1})-(\xi_{t-1}-\xi_{t-2})\right|\leq M\max_{s=-1,\ldots,-\underline{T}}\left|(\xi_{s+1}-\xi_{s})-(\xi_{s}-\xi_{s-1})\right|. \tag{73}$$

For $t\geq 1$, let

$$\Delta_t=(\xi_t-\xi_{t-1})-(\xi_{t-1}-\xi_{t-2})$$

and it is easy to get

$$\begin{aligned}\xi_t-\xi_{t-1}&=-\xi_{-1}+\sum_{j=1}^{t}\Delta_j,\\ \Rightarrow \xi_t&=-t\xi_{-1}+\sum_{k=1}^{t}\sum_{j=1}^{k}\Delta_j,\end{aligned}$$

For $t=1,...,\overline{T}$, let

$$L_t=\frac{t(t+1)}{2},\quad H_t=t.$$

The target object is bounded by

$$\begin{aligned}\iota'\tau &= \iota'\gamma^{post} - \iota'\xi^{post} \\ &\in \left[\iota'\gamma^{post} + \iota' H\xi_{-1} - \iota' LM\bar{\Delta},\ \iota'\gamma^{post} + \iota' H\xi_{-1} + \iota' LM\bar{\Delta}\right]\end{aligned}$$

In this example,

$$\theta \in \left[\min_{b=1,..,2\underline{T}} \lambda_{\ell,b}, \quad \max_{b=1,..,2\underline{T}} \lambda_{u,b}\right]$$

where

$$\begin{aligned}\lambda_\ell &= \lambda_u = A\delta, \\ A_\ell &= A_u = \begin{bmatrix} 1, & \iota' LM\mathcal{I}_{T^{pre}-1} \\ 1, & \iota' LM\mathcal{I}_{T^{pre}-1} \end{bmatrix}, \\ \delta &= \begin{pmatrix} \iota'\gamma^{post} + \iota' H\xi_{-1} \\ \Delta_{-T^{pre}+2} \\ \vdots \\ \Delta_0 \end{pmatrix}.\end{aligned}$$

B Additional Results for Infinite $\mathcal{B}$

B.1 A Modified E-A-M Algorithm based on Kaido et al. (2019)

Kaido et al. (2019) provide an E-A-M algorithm to improve the computational efficiency of the grid search. For completeness, I provide below the full version of the E-A-M algorithm. This follows closely from Kaido et al. (2019), with the main difference in the M step.

I illustrate with the upper bound of the CI, and the lower bound is symmetric. The goal is to solve the optimization problem

$$\sup_{\beta\in\overline{\mathcal{B}}} \hat{\lambda}_u(\beta) + \frac{\hat{\sigma}_u(\beta)}{\sqrt{n}}\hat{c}_n(\beta), \qquad \text{s.t. } \frac{\sqrt{n}\bar{m}_{n,j}(\beta)}{\hat{\sigma}_j(\beta)} \leq \hat{c}_n(\beta).$$

Initialization: Draw randomly (uniformly) over $\overline{\mathcal{B}}$ a set$\left(\beta^{(1)}, \ldots, \beta^{(k)}\right)$ of initial evaluation points. Evaluate $\hat{c}_n\left(\beta^{(\ell)}\right)$ for $\ell = 1, \ldots, k-1$. Initialize $L = k$.

E-step: Record the tentative optimal value

$$\begin{aligned}&\hat{\lambda}_u(\beta^{*,L}) + \frac{\hat{\sigma}_u(\beta^{*,L})}{\sqrt{n}}\hat{c}_n(\beta^{*,L}) \\ =& \max\left\{\hat{\lambda}_u(\beta^{(\ell)}) + \frac{\hat{\sigma}_u(\beta^{(\ell)})}{\sqrt{n}}\hat{c}_n(\beta^{(\ell)}) : \ell \in \{1, \ldots, L\}, \bar{g}(\beta^{(\ell)}) \leq \hat{c}_n\left(\beta^{(\ell)}\right)\right\},\end{aligned}$$

with $\bar{g}(\beta) = \max_{j=1,\ldots,J} \frac{\sqrt{n}\bar{m}_{n,j}(\beta)}{\hat{\sigma}_j(\beta)}$.

A-step: Approximate $\beta \mapsto \hat{c}_n(\beta)$ by a flexible auxiliary model. We use a Gaussian-process regression model, denoted by a mean-zero Gaussian process $\zeta(\cdot)$ indexed by β and with constant variance ς^2:

$$Y^{(\ell)} = \mu + \zeta\left(\beta^{(\ell)}\right), \quad \ell = 1, \ldots, L,$$
$$\text{Corr}\left(\zeta(\beta), \zeta\left(\beta'\right)\right) = K_\gamma\left(\beta - \beta'\right), \quad \beta, \beta' \in \overline{\mathcal{B}},$$

where $Y^{(\ell)} = \hat{c}_n\left(\beta^{(\ell)}\right)$ and K_γ is a kernel with parameter vector $\gamma \in \prod_{h=1}^{d}\left[\underline{\gamma}_h, \bar{\gamma}_h\right] \subset \mathbb{R}_{++}^d$. For instance, $K_\gamma\left(\beta - \beta'\right) = \exp\left(-\sum_{h=1}^{d}\left|\beta_h - \beta_h'\right|^2/\gamma_h\right)$. The unknown parameters $\left(\mu, \varsigma^2\right)$ can be estimated by a GLS regression of $\boldsymbol{Y} = \left(Y^{(1)}, \ldots, Y^{(L)}\right)'$ on a constant with the given correlation matrix. The unknown parameters γ can be estimated using a (concentrated) MLE. The predictor of the critical value is given by

$$c_L(\beta) = \hat{\mu} + \mathbf{r}_L(\beta)'\mathbf{R}_L^{-1}(\boldsymbol{Y} - \hat{\mu}\mathbf{1}),$$

where $\mathbf{r}_L(\beta)$ is a vector whose ℓ th component is $\text{Corr}\left(\zeta(\beta), \zeta\left(\beta^{(\ell)}\right)\right)$ as given above with estimated parameters, and $\mathbf{R}_L$ is an L-by-L matrix whose (ℓ, ℓ') entry is $\text{Corr}\left(\zeta(\beta^{(\ell)}), \zeta(\beta^{(\ell')})\right)$ with estimated parameters. The uncertainty left in $\hat{c}_n(\cdot)$ is captured by the variance

$$\hat{\varsigma}^2 s_L^2(\beta) = \hat{\varsigma}^2\left(1 - \mathbf{r}_L(\beta)'\mathbf{R}_L^{-1}\mathbf{r}_L(\beta) + \frac{\left(1 - \mathbf{1}'\mathbf{R}_L^{-1}\mathbf{r}_L(\beta)\right)^2}{\mathbf{1}'\mathbf{R}_L^{-1}\mathbf{1}}\right).$$

M-step: With probability $1 - \epsilon$, obtain the next evaluation point $\beta^{(L+1)}$ as

$$\begin{aligned}\beta^{(L+1)} &\in \arg\max_{\beta \in \overline{\mathcal{B}}} E_L(\theta) \\ &= \arg\max_{\beta \in \overline{\mathcal{B}}} \sigma(\beta)\left(\phi\left(x^*(\beta)\right)\hat{\varsigma} s_L(\beta) + \left(\frac{\hat{\lambda}_u(\beta) - \hat{\lambda}_u(\beta^{*,L}) - n^{-\frac{1}{2}}\hat{\sigma}_u(\beta^{*,L})\hat{c}_n(\beta^{*,L})}{\hat{\sigma}_u(\beta)} + c_L(\beta)\right)\Phi\left(x^*(\beta)\right)\right),\end{aligned}$$

where $E_L(\beta)$ is the expected improvement function and

$$x^*(\beta) = (\hat{\varsigma} s_L(\beta))^{-1}\left(c_L(\beta) - \max\left\{\bar{g}(\beta), \ \frac{\hat{\lambda}_u(\beta^{*,L}) + n^{-\frac{1}{2}}\hat{\sigma}_u(\beta^{*,L})\hat{c}_n(\beta^{*,L}) - \hat{\lambda}_u(\beta)}{\hat{\sigma}_u(\beta)}\right\}\right).$$

This step can be implemented by standard nonlinear optimization solvers, for example, MATLAB's fmincon. With probability ϵ, draw $\beta^{(L+1)}$ randomly from a uniform distribution over $\overline{\mathcal{B}}$. Set $L \leftarrow L + 1$ and return to the E-step.

B.2 Simulation Details

Following Dickstein and Morales (2018), $\varphi = (\varphi_1, ..., \varphi_K)$ is identified by

$$\varphi_f = \frac{\text{cov}\left(r_f, r_h | d = 1\right)}{\text{var}\left(r_h | d = 1\right)} = \varphi^\dagger\left(E\left[m_{\varphi_f}(W)\right]\right),$$

where

$$\varphi^\dagger\left(a_1, a_2, a_3, a_4\right) = \frac{a_1 - a_3 a_4}{a_2 - a_4^2},$$
$$m_{\varphi_f}(W) = \left(d r_h r_{f'}, d r_h^2, d r_f, d r_h\right)\mathbf{1}[f' = f].$$

Importantly, (β_1, β_2) is partially identified by the set of moment conditions:

$$
\begin{aligned}
m_1(W;\beta,\varphi) &= -d\frac{1-\Phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)}{\Phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)} + (1-d)\\
m_2(W;\beta,\varphi) &= -(1-d)\frac{\Phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)}{1-\Phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)} + d\\
m_3(W;\beta,\varphi) &= (1-d)\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right) - d\frac{\phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)}{\Phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)}\\
m_4(W;\beta,\varphi) &= -d\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right) - (1-d)\frac{\phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)}{1-\Phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)}
\end{aligned}
$$

The moment conditions are given by

$$
E\left[g(Z)\otimes m\right] \le 0.
$$

The counterfactual of interest for given (β, φ) is given by

$$
\begin{aligned}
\lambda_\ell(\beta,\varphi) = \lambda_u(\beta,\varphi) &= \frac{E\left[\Phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)\right]}{E\left[\Phi\left(\eta^{-1}\varphi_f X_1 \beta_1 - \beta_2\right)\right]} - 1\\
&= \lambda^\dagger\left(E\left[\Phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)\right], E\left[\Phi\left(\eta^{-1}\varphi_f X_1 \beta_1 - \beta_2\right)\right]\right)
\end{aligned}
$$

where

$$
\lambda^\dagger(a_1, a_2) = \frac{a_1}{a_2}.
$$

B.3 Empirical Details with Dickstein and Morales (2018)

All firms located in a country h, indexed by $i = 1, \ldots, N$, decide whether to sell in each export market j with $j = 1, \ldots, J$. In the first period, firms determine the set of countries to which they intend to export. In the second period, upon entering a foreign market, all firms optimally set their prices and realize the associated export profits.

In the second period, the revenue firm i would obtain if it were to sell in market j is r_{ij} and

$$
r_{ij} = \varphi_j r_{ih} + e_{ij} \tag{74}
$$

Firms do not know e_{ij} when deciding whether to export to market j and

$$
E_{jt}\left[e_{ij} \mid \mathcal{J}_{ij}, r_{ih}, f_{ij}\right] = 0 \tag{75}
$$

where f_{ij} is the fixed cost and $\mathcal{J}_{ij}$ is the information available to firm i when deciding whether to participate in market j. The export profits that i would obtain in j is

$$
\pi_{ij} = \eta^{-1} r_{ij} - f_{ij},
$$

where η is the demand elasticity. The fixed export costs is

$$f_{ij} = \tilde{\beta}_1 + \tilde{\beta}_2 dist_j + v_{ij},$$

where $dist$ denotes the distance from country h to country j, and the term v_{ijt} represents determinants of f_{ijt} that the researcher does not observe. Firms know f_{ijt} when deciding whether to export to j at t and

$$v_{ij} \mid (\mathcal{J}_{ij},\ \text{dist}\ _j) \sim \mathcal{N}\left(0, \tilde{\beta}_3^2\right)$$

In the first period, a risk-neutral firm i will decide to export to j if and only if

$$d_{ij} = \mathbf{1}\left\{\eta^{-1} E\left[\varphi_j r_{ih} \mid \mathcal{J}_{ij}\right] \beta_1 - \beta_2 - \beta_3 dist_j \geqslant v_{ij}/\tilde{\beta}_3\right\} \tag{76}$$

where $\beta = (\beta_1, \beta_2, \beta_2) = (\frac{1}{\tilde{\beta}_3}, \frac{\tilde{\beta}_1}{\tilde{\beta}_3}, \frac{\tilde{\beta}_2}{\tilde{\beta}_3})$ and the probability that i exports to j at t conditional on $\mathcal{J}_{ij}$ and $dist_j$ is

$$E\left[d_{ij} \mid \mathcal{J}_{ij}, dist_j\right] = \Phi\left(\eta^{-1} E\left[\varphi_j r_{ih} \mid \mathcal{J}_{ijt}\right] \beta_1 - \beta_2 - \beta_3 dist_j\right) \quad . \tag{77}$$

Moment conditions

Odds-Based Moment Inequalities: For any $Z_{ij} \subseteq (\mathcal{J}_{ij}, dist_j)$, we define the conditional odds-based moment inequalities as

$$\mathcal{M}^{ob}\left(Z_{ij}; \beta, \varphi\right) = E\left[\begin{array}{c} m_l^{ob}\left(d_{ij}, r_{ij}^o, dist_j; \theta\right) \\ m_u^{ob}\left(d_{ij}, r_{ij}^o, dist_j; \theta\right) \end{array} \mid Z_{ij}\right] \geqslant 0$$

where $r_{ij}^o = E\left[\varphi_j r_{ih} | \mathcal{J}_{ij}\right]$ and the two moment functions are defined as

$$m_l^{ob}(\cdot) = d_{ij} \frac{1 - \Phi\left(\eta^{-1} r_{ij}^o \beta_1 - \beta_2 - \beta_3 dist_j\right)}{\Phi\left(\eta^{-1} r_{ij}^o \beta_1 - \beta_2 - \beta_3 dist_j\right)} - (1 - d_{ij}),$$

$$m_u^{ob}(\cdot) = (1 - d_{ij}) \frac{\Phi\left(\eta^{-1} r_{ij}^o \beta_1 - \beta_2 - \beta_3 dist_j\right)}{1 - \Phi\left(\eta^{-1} r_{ij}^o \beta_1 - \beta_2 - \beta_3 dist_j\right)} - d_{ij}.$$

Revealed Preference Moment Inequalities: For any $Z_{ij} \subseteq (\mathcal{J}_{ij}, dist_j)$, we define a conditional revealed preference moment inequality as

$$\mathcal{M}^{r}\left(Z_{ijt}; \beta, \varphi\right) = E\left[\begin{array}{c} m_l^{r}\left(d_{ijt}, r_{ijt}^o, dist_j; \theta\right) \\ m_u^{r}\left(d_{ijt}, r_{ijt}^o, dist_j; \theta\right) \end{array} \mid Z_{ijt}\right] \geqslant 0,$$

where the two moment functions are defined as

$$m_l^r(\cdot) = -(1-d_{ij})\left(\eta^{-1}r_{ij}^o\beta_1 - \beta_2 - \beta_3 dist_j\right) + d_{ij}\frac{\phi\left(\eta^{-1}r_{ij}^o\beta_1 - \beta_2 - \beta_3 dist_j\right)}{\Phi\left(\eta^{-1}r_{ij}^o\beta_1 - \beta_2 - \beta_3 dist_j\right)},$$
$$m_u^r(\cdot) = d_{ij}\left(\eta^{-1}r_{ij}^o\beta_1 - \beta_2 - \beta_3 dist_j\right) + (1-d_{ij})\frac{\phi\left(\eta^{-1}r_{ij}^o\beta_1 - \beta_2 - \beta_3 dist_j\right)}{1-\Phi\left(\eta^{-1}r_{ij}^o\beta_1 - \beta_2 - \beta_3 dist_j\right)}.$$

Counterfactuals

1. Changes in Information Sets $\mathcal{J}_{ij}^c$. The number of exporter

$$N_{c1}^{ex} = \sum_i \Phi\left(\eta^{-1}r_{ij}^{oc}\beta_1 - \beta_2 - \beta_3 dist_j\right)$$

where

$$r_{ij}^{oc} = E\left[\alpha_j r_{ih} | \mathcal{J}_{ij}^c\right]$$

We are interested in the change of exporter numbers $\theta = \frac{E[N_{c1}^{ex}]}{E[N^{ex}]}$.

2. Changes in Fixed Export Costs: a reduction in exporters' fixed costs of 40%. Suppose $Z_{ijt} \subseteq \mathcal{J}_{ijt}$ and, for any $\beta \in \overline{\mathcal{B}}$, define

$$N_{c2}^{ex} = \sum_i \Phi\left(\eta^{-1}\underbrace{E\left[\alpha_j r_{ih} \mid \mathcal{J}_{ij}\right]}_{\text{not point identified}}\beta_1 - \beta_2 - \beta_3 dist_j\right)$$

Then,

$$\sum_i \frac{1}{1+B^l\left(Z_{ij};\beta\right)} \leqslant N_{c2}^{ex} \leqslant \sum_i \frac{B^u\left(Z_{ij};\beta\right)}{1+B^u\left(Z_{ij};\beta\right)},$$

where

$$B^l\left(Z_{ij};\beta,\varphi\right) = E\left[\frac{1-\Phi\left(\eta^{-1}r_{ij}^o\beta_1 - 0.6(\beta_2+\beta_3 dist_j)\right)}{\Phi\left(\eta^{-1}r_{ij}^o\beta_1 - 0.6(\beta_2+\beta_3 dist_j)\right)} \mid Z_{ij}\right],$$
$$B^u\left(Z_{ij};\beta,\varphi\right) = E\left[\frac{\Phi\left(\eta^{-1}r_{ij}^o\beta_1 - 0.6(\beta_2+\beta_3 dist_j)\right)}{1-\Phi\left(\eta^{-1}r_{ij}^o\beta_1 - 0.6(\beta_2+\beta_3 dist_j)\right)} \mid Z_{ij}\right].$$

We are interested in the change of exporter numbers $\theta = \frac{E[N_{c2}^{ex}]}{E[N^{ex}]}$.

C Proofs for Section 3

C.1 Notation

For simplicity, 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. For instance, if

$$A_\ell = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad A_u = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$$

then $\mathcal{B}_\ell = \{1, 2\}$ and $\mathcal{B}_u = \{1, 2, 3\}$. Intuitively, $\mathcal{B}_\ell$ and $\mathcal{B}_u$ remove the redundant rows and is (possibly an outer set of) the support of $\hat{b}_\ell$ and $\hat{b}_u$.

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

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

Let

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

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

$$T_k = \max\left\{\min_{b \in \mathcal{B}} Z_{\ell,b} + \lambda_{k\ell,b},\ \min_{b \in \mathcal{B}} Z_{u,b} + \lambda_{ku,b}\right\} \tag{78}$$

be the asymptotic analog of $\hat{T}(\theta_{P_n,k})$, where $(\lambda_{k\ell}, \lambda_{ku})$ are specified in Lemma 3. And let $b_{k\ell}$, b_{ku} be the asymptotic analog of $\hat{b}_\ell(\theta_k)$ and $\hat{b}_u(\theta_k)$, with support $\mathcal{B}_{k\ell}$, $\mathcal{B}_{ku}$:

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

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

$$t_{k\ell,1}(b) = \begin{cases} \min_{\tilde{b} \in \mathcal{B}} \left(1 + \rho_{\ell u}(b, \tilde{b})\right)^{-1} \left(Z_{u,\tilde{b}} + \rho_{\ell u}(b, \tilde{b}) Z_{\ell,b} + t^\dagger_{k\ell,1}(b, \tilde{b})\right), & \text{if } \min_{\tilde{b} \in \mathcal{B}} \rho_{\ell u}(b, \tilde{b}) > -1 \\ -\infty & \text{elsewhere} \end{cases} \tag{79}$$

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

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

$$t_{ku,2}(b)=\begin{cases}\min_{\tilde b\in\mathcal{B}:\rho_{u}(\tilde b,b)<1}\left(1-\rho_{u}(\tilde b,b)\right)^{-1}\left(Z_{u,\tilde b}-\rho_{u}(\tilde b,b)Z_{u,b}+t^{\dagger}_{ku,2}(b,\tilde b)\right) & \text{if } \min_{\tilde b\in\mathcal{B}}\rho_{u}(b,\tilde b)<1\\ +\infty & \text{elsewhere}\end{cases}$$

where

$$\begin{aligned}t^{\dagger}_{k\ell,1}(b,\tilde b)&=\lambda_{ku,\tilde b}+\rho_{\ell u}(b,\tilde b)\lambda_{k\ell,b}\\ t^{\dagger}_{ku,1}(b,\tilde b)&=\lambda_{k\ell,\tilde b}+\rho_{\ell u}(\tilde b,b)\lambda_{ku,b}\\ t^{\dagger}_{k\ell,2}(b,\tilde b)&=\lambda_{k\ell,\tilde b}-\rho_{\ell}(b,\tilde b)\lambda_{k\ell,b}\\ t^{\dagger}_{ku,2}(b,\tilde b)&=\lambda_{ku,\tilde b}-\rho_{u}(\tilde b,b)\lambda_{ku,b}.\end{aligned}$$

Note that if $\left|\lambda_{ku,\tilde b}\right|\to\infty$ and $|\lambda_{k\ell,b}|\to\infty$, $t^{\dagger}_{k\ell,1}(b,\tilde b)$ may not be well defined. However, as we will see later, this case is irrelevant for the proof. Same applies to $t^{\dagger}_{ku,1}$, $t^{\dagger}_{k\ell,2}$ and $t^{\dagger}_{ku,2}$. And let

$$c^{\mathrm{c}}_k=\begin{cases}\Phi^{-1}\left(\alpha^{\mathrm{c}}\Phi\left(t_{k\ell,1}(b_{k\ell})\right)+(1-\alpha^{\mathrm{c}})\Phi\left(t_{k\ell,2}(b_{k\ell})\right)\right) & \text{if } Z_{\ell,b_{k\ell}}+\lambda_{k\ell,b_{k\ell}}\ge Z_{u,b_{ku}}+\lambda_{ku,b_{ku}}\\ \Phi^{-1}\left(\alpha^{\mathrm{c}}\Phi\left(t_{ku,1}(b_{ku})\right)+(1-\alpha^{\mathrm{c}})\Phi\left(t_{ku,2}(b_{ku})\right)\right) & \text{if } Z_{\ell,b_{k\ell}}+\lambda_{k\ell,b_{k\ell}}< Z_{u,b_{ku}}+\lambda_{ku,b_{ku}}\end{cases}\tag{80}$$

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

$$\begin{aligned}p(c)=\max\{&P\left(T_\ell>c^{\mathrm{m}}_\ell(c)\text{ or }\{T_m>c^{\mathrm{m}}_m(c)\text{ and }T_u>c^{\mathrm{m}}_u(c)\}\right),\\ &P\left(T_u>c^{\mathrm{m}}_u(c)\text{ or }\{T_m>c^{\mathrm{m}}_m(c)\text{ and }T_\ell>c^{\mathrm{m}}_\ell(c)\}\right)\},\end{aligned}\tag{81}$$

where

$$c^{\mathrm{m}}_k(c)=\max\{c^{\mathrm{c}}_k,c\},$$

and T_k, c^{c}_k are defined in (78), (80). Lastly, let

$$c^t=\inf_c\{c\ge 0: p(c)\le\alpha-\eta\},\tag{82}$$

be the asymptotic analog of $\hat c^{\mathrm{t}}$ defined (31).

I use Φ for the CDF of $\mathcal{N}(0,1)$ and $\Phi_2(x_1,x_2;\rho)$ for the CDF of $\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1&\rho\\ \rho&1\end{bmatrix}\right)$.

C.2 Proofs for Theorems and Propositions

Proof of Lemma 1.

Proof. Let b_1 satisfy $\lambda_{\ell,b_1} \leq \theta$, and I show that

$$\frac{\Phi\left(\hat{T}(\theta)\right) - \Phi\left(t_{\ell,1}(\theta, b_1)\right)}{\Phi\left(t_{\ell,2}(\theta, b_1)\right) - \Phi\left(t_{\ell,1}(\theta, b_1)\right)} \Big| \left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\right\} \overset{\text{FOSD}}{\preceq} \text{Unif}(0,1). \tag{83}$$

The proof mainly uses Theorem 5.2 and Lemma A.1 in Lee et al (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

$$\left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\right\} = \bigcup_{s\in\mathcal{B}} \left\{A_s \mathcal{Z}_{\ell,b_1} \leq b_s\right\}. \tag{84}$$

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

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

and

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

Therefore,

$$\left\{A_s\mathcal{Z}_{\ell,b_1} \leq b_s\right\} = \left\{\mathcal{V}_s^- \leq \mathcal{Z}_{\ell,b_1} \leq \mathcal{V}^+, \mathcal{V}^0 \geq 0\right\} \tag{85}$$

where

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

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

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

Note that

$$\mathcal{Z}_{\ell,b_1} \perp \left\{\mathcal{V}^+, \{\mathcal{V}_s^-, \mathcal{V}_s^0\}_{s\in\mathcal{B}}\right\}$$

by construction. We can easily verify that

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

where $t_{\ell,1}(\theta,b_1)$ and $t_{\ell,2}(\theta,b_1)$ are defined in Lemma 1.

Let

$$\mu = E\left[\mathcal{Z}_{\ell,b_1}\right] = \frac{\lambda_{\ell,b_1} - \theta}{\sigma_{\ell,b_1}} \leq 0$$

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

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

Then by Theorem 5.3 in Lee, Sun, Sun, and Taylor (2016),

$$F_\mu\left(\mathcal{Z}_{\ell,b_\ell}; t_{\ell,1}(\theta,b_1), t_{\ell,2}(\theta,b_1)\right) \Big| \bigcup_{s\in\mathcal{B}} \{A_s\mathcal{Z}_{\ell,b_1} \leq b_s\} \sim \text{Unif}(0,1), \tag{86}$$

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

$$F_0\left(z; t_{\ell,1}(\theta,b_1), t_{\ell,2}(\theta,b_1)\right) \leq F_\mu\left(z; t_{\ell,1}(\theta,b_1), t_{\ell,2}(\theta,b_1)\right). \tag{87}$$

Therefore, we have

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

□

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 $b_1 \in \mathcal{B}_{\ell 0}$, by Lemma 1, it holds that

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

where the second line follows from $F_0(x; t_1, t_2)$ strictly increasing in x, the inequality follows from (87) and by construction

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

and the last line follows from (86).

Let $b_2 \in \mathcal{B}_{u0}$. Similar argument gives

$$P\left(\hat{T}(\theta) > \hat{c}^{\mathrm{c}}(\theta,\alpha) \middle| \hat{T}(\theta) = \mathcal{Z}_{u,b_2}\right) \leq \alpha. \tag{88}$$

Therefore, we have

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

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 (23), and the inequality follows from (83) and (88). □

Proof of Theorem 1.

Proof. By Lemma 3, we only focus on the subsequence s_n, and for simplicity, I write n for subsequence s_n.

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

$$\bar{p}\left(c, \lambda_{P_n}, \hat{\Sigma}_n/n\right) \xrightarrow{p} p(c) \tag{89}$$

where $\bar{p}\left(c, \lambda_{P_n}, \hat{\Sigma}_n/n\right)$ is defined in (30) and $p(c)$ is defined in (81). Note that by Assumption 4 and (110), $\hat{\Omega}_n \xrightarrow{p} \Omega_0$, thus there is $\tau_n = o(1)$ such that

$$\hat{\Omega}_n = \Omega_0 + o_p(\tau_n). \tag{90}$$

Let

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

To show (89), note that for all $\varepsilon > 0$,

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

where the first inequality follows from $P(A) \leq P(A \cap B) + P(B^c)$, the second inequality follows from (90), and the last line is by $\bar{p}\left(c, \lambda_{P_n}, \Sigma/n\right)$ and $p(c)$ are non-random. Thus it suffices to show

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

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

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

and it suffices to show

$$\limsup_n \bar{p}\left(c, \lambda_{P_n}, \Sigma_n/n\right) = p(c). \tag{91}$$

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 λ_{P_n}, note that

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

is bounded and continuous on D with

$$P(D^c) = P\left(T_\ell = c^{\mathrm{m}}(c_\ell^{\mathrm{c}}, c) \text{ or } T_m = c^{\mathrm{m}}(c_m^{\mathrm{c}}, c) \text{ or } T_u = c^{\mathrm{m}}(c_u^{\mathrm{c}}, c)\right) = 0,$$

because (i) (T_ℓ, T_m, T_u) is continuously distributed and (ii) $T_\ell \perp c^{\rm c}_\ell$, $T_m \perp c^{\rm c}_m$, $T_u \perp c^{\rm c}_u$ by construction. Thus (91) follows from Lemma 4.

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

$$\tilde{p}(c, \lambda_{P_n}, \Sigma_n/n) = \max\Big\{P\left(\hat{T}(\lambda_{P_n,\ell,b_\ell}) > \hat{c}^{\rm m}(\lambda_{P_n,\ell,b_\ell}, c; \alpha^{\rm c}); \mathcal{N}(\lambda_{P_n}, \Sigma_n)\right), \\ P\left(\hat{T}(\lambda_{P_n,u,b_u}) > \hat{c}^{\rm m}(\lambda_{P_n,u,b_u}, c; \alpha^{\rm c}); \mathcal{N}(\lambda_{P_n}, \Sigma_n)\right)\Big\} \tag{92}$$

and we have

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

where the last equality follows from Lemma 4. And

$$\begin{aligned}\tilde{p}(c, \lambda_{P_n}, \Sigma_n/n) =& \max\Big\{P\left(\hat{T}(\theta_{n,\ell}) > \tilde{c}^{\rm m}(\theta_{n,\ell}, c; \alpha^{\rm c}); \mathcal{N}(\lambda_{P_n}, \Sigma_n)\right), \\ & \qquad P\left(\hat{T}(\theta_{n,u}) > \tilde{c}^{\rm m}(\theta_{n,u}, c; \alpha^{\rm c}); \mathcal{N}(\lambda_{P_n}, \Sigma_n)\right)\Big\} \\ \to& \max\left\{P\left(T_\ell > c^{\rm m}_\ell(c)\right),\ P\left(T_u > c^{\rm m}_u(c)\right)\right\} \qquad (93) \\ =& p(c)\end{aligned}$$

(93) follows from continuous mapping theorem.

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

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

where $c^{\rm t}$ is defined in (82) and $\hat{c}^t_{P_n}$ is defined in (112). Note that by definition

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

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

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

where the last equation is by

$$\bar{p}\left(c^t - \varepsilon, \lambda_{P_n}, \hat{\Sigma}_n/n\right) \xrightarrow{p} p\left(c^t - \varepsilon\right) < \alpha - \eta, \tag{94}$$

and (94) follows from Step 1 (89).

Step 3. For all $\varepsilon > 0$, it holds that

$$\begin{aligned}&\limsup_{n\to\infty} \max\Big\{P_n\left(\hat{T}(\theta_\ell) > \hat{c}^{\rm m}(\theta_\ell, \hat{c}^t) \vee \left\{\hat{T}(\theta_m) > \hat{c}^{\rm m}(\theta_m, \hat{c}^t) \wedge \hat{T}(\theta_u) > \hat{c}^{\rm m}(\theta_u, \hat{c}^t)\right\}\right), \\ &\qquad P_n\left(\hat{T}(\theta_u) > \hat{c}^{\rm m}(\theta_u, \hat{c}^t) \vee \left\{\hat{T}(\theta_m) > \hat{c}^{\rm m}(\theta_m, \hat{c}^t) \wedge \hat{T}(\theta_\ell) > \hat{c}^{\rm m}(\theta_\ell, \hat{c}^t)\right\}\right)\Big\} \\ =& \limsup_{n\to\infty} \max\Big\{P_n\left(\hat{T}(\theta_\ell) > \tilde{c}^{\rm m}(\theta_\ell, \hat{c}^t) \vee \left\{\hat{T}(\theta_m) > \tilde{c}^{\rm m}(\theta_m, \hat{c}^t) \wedge \hat{T}(\theta_u) > \tilde{c}^{\rm m}(\theta_u, \hat{c}^t)\right\}\right),\end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad P_n\left(\hat{T}(\theta_u) > \tilde{c}^{\mathrm{m}}(\theta_u, \hat{c}^t) \vee \left\{\hat{T}(\theta_m) > \tilde{c}^{\mathrm{m}}(\theta_m, \hat{c}^t) \wedge \hat{T}(\theta_\ell) > \tilde{c}^{\mathrm{m}}(\theta_\ell, \hat{c}^t)\right\}\right)\Big\} \\
&\leq \limsup_{n\to\infty} \max\Big\{P_n\left(\hat{T}(\theta_\ell) > \tilde{c}^{\mathrm{m}}(\theta_\ell, c^t-\varepsilon) \vee \left\{\hat{T}(\theta_m) > \tilde{c}^{\mathrm{m}}(\theta_m, c^t-\varepsilon) \wedge \hat{T}(\theta_u) > \tilde{c}^{\mathrm{m}}(\theta_u, c^t-\varepsilon)\right\}\right), \\
&\qquad\qquad P_n\left(\hat{T}(\theta_u) > \tilde{c}^{\mathrm{m}}(\theta_u, c^t-\varepsilon) \vee \left\{\hat{T}(\theta_m) > \tilde{c}^{\mathrm{m}}(\theta_m, c^t-\varepsilon) \wedge \hat{T}(\theta_\ell) > \tilde{c}^{\mathrm{m}}(\theta_\ell, c^t-\varepsilon)\right\}\right)\Big\} \\
&\quad + \limsup_{n\to\infty} P_n\left(\hat{c}^t_{P_n} \leq c^t - \varepsilon\right) \\
&= p(c^t-\varepsilon). \qquad (95)
\end{aligned}
$$

Here I omit the subcript P_n in θ_ℓ, θ_m, θ_u and α in $\hat{c}^{\mathrm{m}}$, $\tilde{c}^{\mathrm{m}}$ for simplicity. Since (95) holds at all $\varepsilon > 0$, we can take a sequence of $\varepsilon \to 0$, then by Lemma 9,

$$
\begin{aligned}
&\limsup_{n\to\infty} \max\Big\{P_n\left(\hat{T}(\theta_\ell) > \hat{c}^{\mathrm{m}}(\theta_\ell;\alpha) \vee \left\{\hat{T}(\theta_m) > \hat{c}^{\mathrm{m}}(\theta_m;\alpha) \wedge \hat{T}(\theta_u) > \hat{c}^{\mathrm{m}}(\theta_u;\alpha)\right\}\right), \\
&\qquad\qquad P_n\left(\hat{T}(\theta_u) > \hat{c}^{\mathrm{m}}(\theta_u,\alpha) \vee \left\{\hat{T}(\theta_m) > \hat{c}^{\mathrm{m}}(\theta_m;\alpha) \wedge \hat{T}(\theta_\ell) > \hat{c}^{\mathrm{m}}(\theta_\ell;\alpha)\right\}\right)\Big\} \\
&\leq \lim_{\varepsilon\to 0} p(c^t-\varepsilon) = p(c^t).
\end{aligned}
$$

By construction,

$$p(c^{\mathrm{t}}) \leq \alpha - \eta.$$

Thus by Lemma 3,

$$\lim_n P_n\left(\theta \notin CI^{\mathrm{m}}\right) \leq \alpha.$$

□

Proof of Theorem 2.

Proof. Part I. Symmetric Bounds. Since $\lambda_\ell = \lambda_u$ and $\hat{\lambda}_\ell = \hat{\lambda}_u$, I will omit the subscript ℓ, u in this proof. Let

$$
\begin{aligned}
\mathcal{Z}_{u,b} &= \mathcal{Z}_b = \frac{\hat{\lambda}_b - \theta}{\hat{\sigma}_b/\sqrt{n}}, \ \forall b = 1, .., |\mathcal{B}|, \\
\mathcal{Z}_{\ell,b} &= -\mathcal{Z}_b.
\end{aligned}
$$

And the test statistic is

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

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

$$\liminf_n \inf_{P\in\mathcal{P}} P\left(\hat{c}^{\mathrm{t}} \leq \Phi^{-1}(1-\frac{\alpha_1'}{2})\right) = 1. \qquad (96)$$

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

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

Let $\alpha' = \min\{\alpha_1', \alpha_2'\} > \alpha$, and then (37) follows from

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

<u>Step 2.</u> I show (38) with $\theta_n = \theta_\ell - \frac{\kappa}{\sqrt{n}}$. Note that by (37), there is $\alpha' > \alpha$ such that

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

Under P_n, we can show that there is a subsequence P_{a_n} such that (110), (111) hold and

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

And let

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

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

To simplify notation, let $\varepsilon = \frac{1}{4}\left(\Phi^{-1}(1-\frac{\alpha}{2}) - \Phi^{-1}(1-\frac{\alpha'}{2})\right)$, and $c_1 = \Phi^{-1}(1-\frac{\alpha'}{2}) + \varepsilon$, $c_2 = \Phi^{-1}(1-\frac{\alpha}{2}) - \varepsilon$. We have

$$\liminf_n P_n\left(\hat T(\theta_n) \in \left(\Phi^{-1}(1-\frac{\alpha'}{2}), \Phi^{-1}(1-\frac{\alpha}{2})\right)\right) \geq P\left(T^* \in (c_1, c_2)\right).$$

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. $\sigma_{b^*} \geq \max_{b\in\bar B}\sigma_b$. Then

$$\begin{aligned}
&P\left(T^* \in (c_1, c_2)\right)\\
\geq&P\left(c_2 \geq Z_{b^*} + \bar\lambda_{b^*} + \frac{\kappa}{\sigma_{b^*}} \geq c_1, Z_b + \bar\lambda_b + \frac{\kappa}{\sigma_b} \geq c_1, b \in \bar{\mathcal{B}}\backslash\{b^*\}\right)
\end{aligned}$$

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

where

$$\mathbb{E}_b=Z_b-\rho(b, b^*) Z_{b^*}.$$

There is $\kappa \in \mathbb{R}$ such that

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

and therefore

$$P\left(T^* \in\left(\bar{c}, \Phi^{-1}(1-\frac{\alpha}{2})\right)\right)>0.$$

Part II. Large Bounds. By Lemma 12, there is α_1' such that

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

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

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

And the proof for the lower bound is symmetric.

Let $\kappa_n' \rightarrow \infty$ 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 (36),

$$
\begin{aligned}
\mathcal{Z}_{\ell, \hat{b}_\ell} \leq \mathcal{Z}_{\ell, b_\ell} &=\frac{\hat{\lambda}_{\ell, b_\ell}-\theta}{\hat{\sigma}_{\ell, b_\ell}/\sqrt{n}}\\
&\leq \frac{\hat{\lambda}_{\ell, b_\ell}-\hat{\lambda}_{u, b_u}-\frac{\hat{\sigma}_{u, b_u}}{\sqrt{n}} c_1}{\hat{\sigma}_{\ell, b_\ell}/\sqrt{n}}\\
&=\frac{\hat{\lambda}_{\ell, b_\ell}-\hat{\lambda}_{u, b_u}}{\hat{\sigma}_{\ell, b_\ell}/\sqrt{n}}-\frac{\hat{\sigma}_{u, b_u}}{\hat{\sigma}_{\ell, b_\ell}} c_1 \rightarrow-\infty
\end{aligned}
\tag{97}
$$

and

$$\mathcal{Z}_{u,\hat{b}_u} = \frac{\theta - \hat{\lambda}_{u,\hat{b}_u}}{\hat{\sigma}_{u,\hat{b}_u}} \in \left(c_1,\ \kappa_n'\right]. \tag{98}$$

Thus, with probability approaching one,

$$\hat{T}(\theta) = \mathcal{Z}_{u,\hat{b}_u} > c_1,$$

$$\hat{c}^{\mathrm{c}}(\theta, \alpha^{\mathrm{c}}) = \Phi^{-1}\left(\alpha^{\mathrm{c}}\Phi\left(t_{u,1}(\theta, \hat{b}_u)\right) + (1-\alpha^{\mathrm{c}})\Phi\left(t_{u,2}(\theta, \hat{b}_u)\right)\right).$$

Moreover, (97) and (98) implies that

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

Therefore,

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

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

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

and θ is rejected.

The proof for (38) is similar to Part I Step 2. □

Proof of Theorem 3.

Proof. By Lemma 2, it is easy to see that

$$\begin{aligned}\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n} P\left(\theta_{\ell,n} \notin CI^{\mathrm{m}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) &= 1 \\ \liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n} P\left(\theta_{u,n} \notin CI^{\mathrm{m}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) &= 1.\end{aligned} \tag{99}$$

Ye et al. (2023) confidence interval has form

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

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

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

$\hat{p} \xrightarrow{p} p^* \in \left[\frac{\alpha}{2}, \alpha\right]$. The upper bound is defined symmetrically. First consider the rejection of $\theta_{\ell,n}$. Note that

$$\begin{aligned}
&P\left(\theta_{\ell,n} \notin CI^{\text{YKHS}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) \\
=&P\left(\theta_{\ell,n} < \hat{\lambda}_{m,\min} - \sqrt{\frac{n}{m}} Q^*\left(\hat{\lambda}^*_{n,\min} - \hat{\lambda}_{n,\min}, \hat{p}\right)\right) \\
&+ P\left(\theta_{\ell,n} > \hat{\theta}_{m,\max} - \sqrt{\frac{n}{m}} Q^*\left(\hat{\lambda}^*_{n,\max} - \hat{\lambda}_{n,\max}, 1-\hat{p}\right)\right) \\
=&P\left(Q^*\left(\sqrt{n}\left(\hat{\lambda}^*_{n,\min} - \hat{\lambda}_{n,\min}\right), \hat{p}\right) < \sqrt{m}(\hat{\lambda}_{m,\min} - \theta_{\ell,n})\right) \quad (100) \\
&+ P\left(Q^*\left(\sqrt{n}(\hat{\lambda}^*_{n,\max} - \hat{\lambda}_{n,\max}), 1-\hat{p}\right) > \sqrt{m}(\hat{\lambda}_{m,\max} - \theta_{\ell,n})\right). \quad (101)
\end{aligned}$$

As for (100), note that

$$\begin{aligned}
\sqrt{m}(\hat{\lambda}_{m,\min} - \theta_{\ell,n}) &= \sqrt{m}(\hat{\lambda}_{m,\min} - \lambda_{\ell,b_\ell}) + \sqrt{m}(\lambda_{\ell,b_\ell} - \theta_{\ell,n}) \\
&= \sqrt{m}(\hat{\lambda}_{m,\min} - \lambda_{\ell,b_\ell}) + \kappa'_n \frac{\sqrt{m}}{\sqrt{n}} a \\
&= \sqrt{m}(\hat{\lambda}_{m,\min} - \lambda_{\ell,b_\ell}) + o_p(1) \\
&\xrightarrow{d} \min_{b \in \mathcal{B}} Z_{\ell,b} + \tau_{\ell,b}
\end{aligned}$$

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

$$(100) = P\left(Q^*\left(\sqrt{n}\left(\hat{\lambda}^*_{n,\min} - \hat{\lambda}_{n,\min}\right), \hat{p}\right) < \sqrt{m}(\hat{\lambda}_{m,\min} - \lambda_{\ell,b_\ell})\right) + o(1).$$

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

$$(101) = P\left(Q^*\left(\sqrt{n}(\hat{\lambda}^*_{n,\max} - \hat{\lambda}_{n,\max}), 1-\hat{p}\right) > \sqrt{m}(\hat{\lambda}_{m,\max} - \lambda_{\ell,b_\ell})\right) + o(1) \quad (102)$$

with similar argument. And if $\sqrt{m}(\lambda_{u,b_u} - \theta_{\ell,n}) \to \infty$, (102) still holds since both side of the equation is $o(1)$. In sum, we have

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

thus by Theorem 2(d) in Ye et al. (2023), it holds that

$$\limsup_n \sup_P P\left(\theta_{\ell,n} \notin CI^{\text{YKHS}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) \leq \alpha. \quad (103)$$

(99) and (103) gives (39). □

C.3 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

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

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 the following two conditions hold

$$P_{a_n}\left(\hat{T}(\theta) > \hat{c}^{\mathrm{m}}(\theta;\alpha) \text{ for all } \theta < \theta_\ell - \frac{\bar{\kappa}}{\sqrt{a_n}}\right) \geq 1 - \frac{\varepsilon}{2}, \tag{104}$$

$$P_{a_n}\left(\hat{T}(\theta) > \hat{c}^{\mathrm{m}}(\theta;\alpha) \text{ for all } \theta < \theta_\ell - \frac{\bar{\kappa}}{\sqrt{a_n}}\right) \geq 1 - \frac{\varepsilon}{2}. \tag{105}$$

I will show (104), and the proof of (105) 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

$$\liminf_n P_n\left(A_n(\bar{\kappa}_1)\right) \geq 1 - \frac{\varepsilon}{6}, \tag{106}$$

where

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

To see this,

$P_n\left(A_n(\bar{\kappa}_1)\right)$

$$
\begin{aligned}
=&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)\\
=&P_n\left(\frac{\hat{\lambda}_{\ell,\hat{b}_\ell}-\lambda_{\ell,\hat{b}_\ell}}{\hat{\sigma}_{\ell,\hat{b}_\ell}/\sqrt{n}} + \frac{\lambda_{\ell,\hat{b}_\ell}-\theta_\ell}{\hat{\sigma}_{\ell,\hat{b}_\ell}/\sqrt{n}} + \frac{\theta_\ell-\theta}{\hat{\sigma}_{\ell,\hat{b}_\ell}/\sqrt{n}} \geq \frac{\theta-\theta_\ell}{\hat{\sigma}_{u,b_u}/\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.\\
&\left.\frac{\hat{\lambda}_{\ell,\hat{b}_\ell}-\lambda_{\ell,\hat{b}_\ell}}{\hat{\sigma}_{\ell,\hat{b}_\ell}/\sqrt{n}} + \frac{\lambda_{\ell,\hat{b}_\ell}-\theta_\ell}{\hat{\sigma}_{\ell,\hat{b}_\ell}/\sqrt{n}} + \frac{\theta_\ell-\theta}{\hat{\sigma}_{\ell,\hat{b}_\ell}/\sqrt{n}} > \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{\bar{\kappa}_1}{\hat{\sigma}_{u,b_u}} + \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}} + \frac{\bar{\kappa}_1}{\hat{\sigma}_{\ell,\hat{b}_\ell}} > \Phi^{-1}(1-\frac{\alpha-\eta}{2})\right)\\
=&P_n\left(\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), \bar{\kappa}_1 > \hat{\sigma}_{\ell,\hat{b}_\ell}\Phi^{-1}(1-\frac{\alpha}{2}) - \sqrt{n}\left(\hat{\lambda}_{\ell,\hat{b}_\ell} - \lambda_{\ell,\hat{b}_\ell}\right)\right)
\end{aligned}
$$

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

$$
\begin{aligned}
\left(\frac{1}{\hat{\sigma}_{\ell,\hat{b}_\ell}} + \frac{1}{\hat{\sigma}_{u,b_u}}\right)^{-1}\left(\frac{\theta_u-\hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}} - \frac{\hat{\lambda}_{\ell,\hat{b}_\ell}-\lambda_{\ell,\hat{b}_\ell}}{\hat{\sigma}_{\ell,\hat{b}_\ell}/\sqrt{n}}\right) &= O_P(1),\\
\hat{\sigma}_{\ell,\hat{b}_\ell}\Phi^{-1}(1-\frac{\alpha-\eta}{2}) - \sqrt{n}\left(\hat{\lambda}_{\ell,\hat{b}_\ell} - \lambda_{\ell,\hat{b}_\ell}\right) &= O_P(1).
\end{aligned}
$$

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

$$
\liminf_n P_n(B_n) \geq 1 - \frac{\varepsilon}{6}, \tag{107}
$$

where

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

because

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

and the existence of $\bar{M}$ follows from

$$\frac{1}{\xi}\left|\frac{\lambda_{u,b_u}-\hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}}\right| = O_P(1).$$

And by similar argument in (106), there is $\bar{\kappa}_2$ such that

$$\liminf_n P_n\left(C_n(\bar{\kappa}_2)\right) \geq 1-\frac{\varepsilon}{6}, \tag{108}$$

where

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

where $\bar{z}$ is defined in Lemma 5 with $\bar{M}$ given above (107).

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

$$D_n = \left\{\hat{T}(\theta_n) > \hat{c}^{\mathrm{c}}(\theta_n, 1-\alpha^{\mathrm{c}}) \text{ for all } \theta \leq \theta_\ell - \frac{\bar{\kappa}}{\sqrt{n}}\right\},$$

we have

$$\begin{aligned}
&\liminf_n P_n\left(D_n\right)\\
\geq&\liminf_n P_n\left(A_n(\bar{\kappa})\cap B_n \cap C_n(\bar{\kappa})\cap D_n\right)\\
=&\liminf_n P_n\left(A_n(\bar{\kappa})\cap B_n \cap C_n(\bar{\kappa})\right)\\
\geq&\liminf_n P_n\left(A_n(\bar{\kappa})\right)+P_n\left(B_n\right)+P_n\left(C_n(\bar{\kappa})\right)-2\\
\geq&\liminf_n P_n\left(A_n(\bar{\kappa}_1)\right)+P_n\left(B_n\right)+P_n\left(C_n(\bar{\kappa}_2)\right)-2\\
\geq&1-\frac{\varepsilon}{2}.
\end{aligned}$$

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 (106), (107) and (108). □

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

$$\begin{aligned}
\limsup_{n\to\infty}\max\Big\{&P_n\left(\hat{T}(\theta_{P_n,\ell}) > \tilde{c}^m(\theta_{P_n,\ell},\hat{c}^t_{P_n};\alpha) \text{ or } \hat{T}(\theta_{P_n,m}) > \tilde{c}^m(\theta_{P_n,m},\hat{c}^t_{P_n};\alpha)\right),\\
&P_n\left(\hat{T}(\theta_{P_n,u}) > \tilde{c}^m(\theta_{P_n,u},\hat{c}^t_{P_n};\alpha) \text{ or } \hat{T}(\theta_{P_n,m}) > \tilde{c}^m(\theta_{P_n,m},\hat{c}^t_{P_n};\alpha)\right)\Big\} \leq \alpha-\eta,
\end{aligned}\tag{109}$$

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

1. *The convergence of Ω,*

$$\Omega(P_n)\to\Omega_0\in\mathcal{S}. \tag{110}$$

2. The convergence of

$$(\lambda_{n,k\ell},\lambda_{n,ku})=\left(\left(\frac{\lambda_{P_n,\ell,b}-\theta_{P_n,k}}{\sigma_{P_n,\ell,b}/\sqrt{n}}\right)_{b\in\mathcal{B}},\left(\frac{\theta_{P_n,k}-\lambda_{P_n,u,b}}{\sigma_{P_n,u,b}/\sqrt{n}}\right)_{b\in\mathcal{B}}\right)\to(\lambda_{k\ell},\lambda_{ku}), \tag{111}$$

with $\lambda_{\ell\ell}\in\Lambda_0$*,* $\lambda_{uu}\in\Lambda_0$*,* $\lambda_{\ell,u},\lambda_{mu}\in\Lambda_-$ *and* $\lambda_{m\ell},\lambda_{u\ell}\in\Lambda_-$*,* $\sigma_{0,\ell,b}=\sqrt{A_{\ell,b}\Omega_0A'_{\ell,b}}$*,* $\sigma_{0,u,b}=\sqrt{A_{u,b}\Omega_0A'_{u,b}}$*,*

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

where

$$\hat{c}^t_{P_n}=\inf_c\left\{c\geq 0:\bar{p}\left(c,\lambda_{P_n},\hat{\Sigma}_n/n\right)+\eta\leq\alpha\right\}. \tag{112}$$

Recall that $\tilde{c}^h(\theta,c;\alpha)$ *is defined in (27) and* $\bar{p}\left(c,\lambda_{P_n},\hat{\Sigma}_n/n\right)$ *is defined in (30).*

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

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}}\sup_{\theta\in[\theta_{P,\ell},\theta_{P,u}]}P\left(\theta\notin CI^{\mathrm{m}}\left(\hat{\lambda}_n,\hat{\Sigma}_n/n;\alpha\right)\right)=\lim_{n_a}P_{n_a}\left(\theta_{n_a}\notin CI^{\mathrm{m}}\left(\hat{\lambda}_n,\hat{\Sigma}_n/n;\alpha\right)\right). \tag{113}$$

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 (111) holds. We have found a subsequence n_s such that (110) and (111) hold. And, by (113), we have

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

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

$$\begin{aligned}
&P_{n_s}\left(\theta_{n_s}\notin CI^{\mathrm{m}}\left(\hat{\lambda}_{n_s},\hat{\Sigma}_{n_s}/n_s;\alpha\right)\right)\\
\leq&P_n\left(\theta_n\notin CI^{\mathrm{m}}\left(\hat{\lambda}_n,\hat{\Sigma}_n/n;\alpha\right),\lambda_{P_n}\in\hat{\Lambda}_n\right)+P_n\left(\lambda_{P_n}\notin\hat{\Lambda}_n\right)\\
\leq&\max\left\{P_n\left([\theta_{P_n,\ell},\theta_{P_n,m}]\nsubseteq CI^{\mathrm{m}}\left(\hat{\lambda}_n,\hat{\Sigma}_n/n;\alpha\right),\lambda_{P_n}\in\hat{\Lambda}_n\right),\right.\\
&\left.P_n\left([\theta_{P_n,m},\theta_{P_n,u}]\nsubseteq CI^{\mathrm{m}}\left(\hat{\lambda}_n,\hat{\Sigma}_n/n;\alpha\right),\lambda_{P_n}\in\hat{\Lambda}_n\right)\right\}+P_{n_s}\left(\lambda_{P_n}\notin\hat{\Lambda}_n\right)
\end{aligned}$$

$$\leq \max\Big\{P_n\left(\hat{T}(\theta_{P_n,\ell}) > \hat{c}^{\mathrm{m}}(\theta_{P_n,\ell};\alpha) \text{ or } \hat{T}(\theta_{P_n,m}) > \hat{c}^{\mathrm{m}}(\theta_{P_n,m};\alpha), \lambda_{P_n} \in \hat{\Lambda}_n\right),$$
$$P_n\left(\hat{T}(\theta_{P_n,m}) > \hat{c}^{\mathrm{m}}(\theta_{P_n,m};\alpha) \text{ or } \hat{T}(\theta_{P_n,u}) > \hat{c}^{\mathrm{m}}(\theta_{P_n,u};\alpha), \lambda_{P_n} \in \hat{\Lambda}_n\right)\Big\} + \eta + o(1)$$
$$\leq \max\Big\{P_n\left(\hat{T}(\theta_{P_n,\ell}) > \tilde{c}^{\mathrm{m}}(\theta_{P_n,\ell}, \hat{c}^{\mathrm{t}}_{P_n};\alpha) \text{ or } \hat{T}(\theta_{P_n,m}) > \tilde{c}^{\mathrm{m}}(\theta_{P_n,m}, \hat{c}^{\mathrm{t}}_{P_n};\alpha)\right),$$
$$P_{n_s}\left(\hat{T}(\theta_{P_n,m}) > \tilde{c}^{\mathrm{m}}(\theta_{P_n,m}, \hat{c}^{\mathrm{t}}_{P_n};\alpha) \text{ or } \hat{T}(\theta_{P_n,u}) > \tilde{c}^{\mathrm{m}}(\theta_{P_n,u}, \hat{c}^{\mathrm{t}}_{P_n};\alpha)\right)\Big\} + \eta + o(1).$$

Recall that $\hat{c}^{\mathrm{t}}$ is defined in (31) and $\hat{c}^{\mathrm{t}}_{P_n}$ is defined in (112), 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 (109). □

Lemma 4. *Under Assumptions 1, 2, 3, 4, 5, under sequences (110) and (111), if*

$$\min_{b\in\mathcal{B}} \sigma_{0,u,b}\lambda_{\ell u,b} \in \mathbb{R}, \tag{114}$$

it holds that

$$\left(\hat{T}(\theta_{P_n,k}), \hat{c}^c(\theta_{P_n,k},\alpha^c)\right)_{k=\ell,m,u} \xrightarrow{d} (T_k, c^c_k)_{k=\ell,m,u}. \tag{115}$$

If

$$\min_{b\in\mathcal{B}} \sigma_{0,u,b}\lambda_{\ell u,b} = -\infty, \tag{116}$$

it holds that

$$\left(\hat{T}(\theta_{P_n,k}), \hat{c}^c(\theta_{P_n,k},\alpha^c)\right)_{k=\ell,u} \xrightarrow{d} (T_k, c^c_k)_{k=\ell,u}, \tag{117}$$

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

$$P_n\left(\hat{T}(\theta_{P_n,m}) \geq c\right) \to 0. \tag{118}$$

Proof. Note that

$$\begin{aligned}
\lim_n \sqrt{n}\,(\lambda_{P_n,\ell,b_\ell} - \lambda_{P_n,u,b_u}) &= \lim_n \sigma_{P_n,u,b}\frac{\lambda_{P_n,\ell,b_\ell} - \max_{b\in\mathcal{B}}\lambda_{P_n,u,b}}{\sigma_{P_n,u,b}/\sqrt{n}}\\
&= \lim_n \min_{b\in\mathcal{B}} \sigma_{P_n,u,b}\frac{\theta_{P_n,\ell} - \lambda_{P_n,u,b}}{\sigma_{P_n,u,b}/\sqrt{n}}\\
&= \min_{b\in\mathcal{B}} \lim_n \sigma_{P_n,u,b}\frac{\theta_{P_n,\ell} - \lambda_{P_n,u,b}}{\sigma_{P_n,u,b}/\sqrt{n}} = \min_{b\in\mathcal{B}} \sigma_{0,u,b}\lambda_{\ell u,b}.
\end{aligned} \tag{119}$$

Thus the two cases in (114) and (116) correspond to whether the length of the identified set of θ is large asymptotically. I will show (115) under (114 in Step 1 and 2, then show (117) and (118) under (116) in Step 3.

Step 1. Show that under (114),

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

where

$$t_{k\ell}(\mathcal{B}_{k\ell}) = (t_{k\ell,1}(b),\ t_{k\ell,2}(b))_{b\in\mathcal{B}_{k\ell}},$$
$$t_{ku}(\mathcal{B}_{ku}) = (t_{ku,1}(b),\ t_{ku,2}(b))_{b\in\mathcal{B}_{ku}}.$$

Step 1.1. Note that

$$\begin{aligned}
\hat{T}(\theta_{P_n,k}) &= \max\left\{\min_{b\in\mathcal{B}} \frac{\hat{\lambda}_{\ell,b}-\theta_{P_n,k}}{\hat{\sigma}_{\ell,b}/\sqrt{n}},\ \min_{b\in\mathcal{B}} \frac{\theta_{P_n,k}-\hat{\lambda}_{u,b}}{\hat{\sigma}_{u,b}/\sqrt{n}}\right\} \\
&= \max\left\{\min_{b\in\mathcal{B}_\ell} \frac{\hat{\lambda}_{\ell,b}-\lambda_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}/\sqrt{n}} + \frac{\sigma_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}}\lambda_{n,k\ell,b},\ \min_{b\in\mathcal{B}_u} \frac{\lambda_{P_n,u,b}-\hat{\lambda}_{u,b}}{\hat{\sigma}_{P_n,u,b}/\sqrt{n}} + \frac{\sigma_{P_n,u,b}}{\hat{\sigma}_{u,b}}\lambda_{n,ku,b}\right\} \\
&= \max\left\{\min_{b\in\mathcal{B}_{k\ell}} \frac{\hat{\lambda}_{\ell,b}-\lambda_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}/\sqrt{n}} + \frac{\sigma_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}}\lambda_{n,k\ell,b},\ \min_{b\in\mathcal{B}_{ku}} \frac{\lambda_{P_n,u,b}-\hat{\lambda}_{u,b}}{\hat{\sigma}_{u,b}/\sqrt{n}} + \frac{\sigma_{P_n,u,b}}{\hat{\sigma}_{u,b}}\lambda_{n,ku,b}\right\} \text{ w.p.a. } 1 \\
&\xrightarrow{d} \max\left\{\min_{b\in\mathcal{B}_{k\ell}} Z_{\ell,b}+\lambda_{k\ell,b},\ \min_{b\in\mathcal{B}_{ku}} Z_{u,b}+\lambda_{ku,b}\right\}. \qquad (120)
\end{aligned}$$

The first line is by definition, the second line simply rearranges terms with $\lambda_{n,k\ell,b}$, $\lambda_{n,ku,b}$ defined in (111) and $\mathcal{B}_\ell$, $\mathcal{B}_u$ defined in Section C.1 first paragraph. To see the third line, note that by Assumption 1, 2, 3, 4, we have

$$\left(\left(\frac{\hat{\lambda}_{\ell,b}-\lambda_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}/\sqrt{n}}\right)_{b\in\mathcal{B}}, \left(\frac{\lambda_{P_n,u,b}-\hat{\lambda}_{u,b}}{\hat{\sigma}_{u,b}/\sqrt{n}}\right)_{b\in\mathcal{B}}, \left(\frac{\sigma_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}}\right)_{b\in\mathcal{B}}, \left(\frac{\sigma_{P_n,u,b}}{\hat{\sigma}_{u,b}}\right)_{b\in\mathcal{B}}\right) \xrightarrow{d} \left(Z_\ell, Z_u, \mathbf{1}_{2|\mathcal{B}|}\right). \qquad (121)$$

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

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

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) (121), (ii) Slustsky's Lemma and (iii) the limit distribution is well defined because

$$\begin{aligned}
\lambda_{k\ell,b} &= \lim_n \frac{\lambda_{P_n,\ell,b}-\theta_{P_n,k}}{\sigma_{P_n,\ell,b}/\sqrt{n}} \geq \lim_n \frac{\lambda_{P_n,\ell,b_\ell}-\lambda_{P_n,u,b_u}}{\sigma_{P_n,\ell,b}/\sqrt{n}} = \frac{\min_{b\in\mathcal{B}}\sigma_{0,u,b}\lambda_{\ell u,b}}{\sigma_{0,\ell,b}} \in \mathbb{R}, \qquad (122)\\
\lambda_{ku,b} &= \lim_n \frac{\theta_{P_n,k}-\lambda_{P_n,u,b}}{\sigma_{P_n,u,b}/\sqrt{n}} \geq \lim_n \frac{\lambda_{P_n,\ell,b_\ell}-\lambda_{P_n,u,b_u}}{\sigma_{P_n,u,b}/\sqrt{n}} = \frac{\min_{b\in\mathcal{B}}\sigma_{0,u,b}\lambda_{\ell u,b}}{\sigma_{0,u,b}} \in \mathbb{R}.
\end{aligned}$$

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 (79). And 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 $\min_{\tilde{b}\in\mathcal{B}}\rho_{\ell u}(b,\tilde{b}) > -1$, where

$$\Phi\left(t_{\ell,1}(\theta_k,b)\right) = \min_{\tilde{b}\in\mathcal{B}}\Phi\left(\left(1+\hat{\rho}_{\ell u}(b,\tilde{b})\right)^{-1}\left(\hat{Z}_{u,\tilde{b}} + \frac{\sigma_{u,\tilde{b}}}{\hat{\sigma}_{u,\tilde{b}}}\lambda_{P_n,ku,\tilde{b}} + \lambda_{k\ell}\hat{\rho}_{\ell u}(b,\tilde{b})\left(\hat{Z}_{\ell,b} + \frac{\sigma_{\ell,b}}{\hat{\sigma}_{\ell,b}}\lambda_{P_n,k\ell,b}\right)\right)\right)$$

By (122) and the definition of $\mathcal{B}_{k\ell}$, we have $\lambda_{k\ell,b}\in\mathbb{R}$, and thus

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

Thus

$$\Phi\left(t_{\ell,1}(\theta_k,\mathcal{B}_{k\ell})\right) \xrightarrow{d} \Phi\left(t_{k\ell,1}(\mathcal{B}_{k\ell})\right). \tag{123}$$

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

<u>Step 1.3.</u> Let

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

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, and thus by continuous mapping theorem, it holds that

$$g\left(\mathcal{Z}_{\ell,b_1},\mathcal{Z}_{\ell,b_2}\right) \xrightarrow{d} g\left(Z_{\ell,b_1}+\lambda_{k\ell,b_1}, Z_{\ell,b_2}+\lambda_{k\ell,b_2}\right). \tag{124}$$

Similarly, we have

$$g\left(\mathcal{Z}_{u,b_2},\mathcal{Z}_{\ell,b_1}\right) \xrightarrow{d} g\left(Z_{u,b_2}+\lambda_{ku,b_2}, Z_{\ell,b_1}+\lambda_{k\ell,b_1}\right). \tag{125}$$

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, and we have

$$g\left(\mathcal{Z}_{\ell,b_1},\mathcal{Z}_{u,b_2}\right) \xrightarrow{d} g\left(Z_{\ell,b_1}+\lambda_{k\ell,b_1}, Z_{u,b_2}+\lambda_{ku,b_2}\right). \tag{126}$$

(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 (120), (123), (124), (125), (126) holds jointly.

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

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

And

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

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

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 (123), (124), (125), (126).

Step 3. Now assume (116) holds. We can show that

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

with similar argument as Step 1 and 2. Regarding (118), 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 (116), (119) 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 (118) holds. □

Lemma 5. *Assume that (i) $\theta \leq \theta_\ell$; (ii)*

$$\hat{T}(\theta) = \mathcal{Z}_{\ell,\hat{b}_\ell} > \Phi^{-1}(1 - \frac{\alpha - \eta}{2}); \tag{127}$$

(iii) either

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

or

$$\left(1 + \hat{\rho}_{\ell u}(\hat{b}_\ell, b_u)\right)^{-1} \frac{\lambda_{u,b_u} - \hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}} \leq \bar{M}, \tag{128}$$

$$\hat{T}(\theta) > \bar{z}, \tag{129}$$

where $\bar{M} \in \mathbb{R}$, $\bar{z}$ is defined in Lemma 6 with $\bar{M}$ given in (128). Then

$$\hat{T}(\theta) > \hat{c}^m(\theta, 1 - \alpha^c). \tag{130}$$

Proof. Note that $c^t \leq \Phi^{-1}(1 - \frac{\alpha-\eta}{2})$ by construction, thus under (127), $\hat{T}(\theta) > c^t$ and (130) is equivalent to

$$\hat{T}(\theta) > \hat{c}^c(\theta, 1 - \alpha^c). \tag{131}$$

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

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

In this case, (131) holds trivially. If $\min_{\tilde{b}\in\mathcal{B}} \rho_{\ell u}(\hat{b}_\ell, \tilde{b}) > -1$, we have

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

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

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

where $H(z,\bar{M})$ is defined in Lemma 6. $H(\mathcal{Z}_{\ell,\hat{b}_\ell}, \bar{M}) > 0$ follows from (129) 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 $\bar{z} \in \mathbb{R}$ such that $H(z) > 0$ for all $z \ge \bar{z}$.

Proof. Note that

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

and thus there is $\bar{z} \in \mathbb{R}$ such that $H'(z) < 0$ for all $z \ge \bar{z}$. Also note that

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

Therefore, for all $z \ge \bar{z}$, we have $H(z) > 0$. □

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*

$$\begin{aligned}
H\left(c,\Delta,\rho\right) &= \Phi_2\left(-c, \Delta - c; \rho\right) + \Phi\left(-\frac{\Delta}{2} - c\right), \\
\rho_2^*(\alpha,\eta) &= \sup_{\rho}\left\{\rho : \sup_{\Delta\ge 0} H\left(c^{sim},\Delta,\rho\right) \le \alpha - \eta\right\}.
\end{aligned} \tag{132}$$

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

$$\sup_{\rho \le \rho_2^*(\alpha,\eta)-\xi}\ \sup_{\Delta\ge 0} H\left(\bar{c},\Delta,\rho\right) < \alpha - \eta. \tag{133}$$

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

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

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

$$\sup_{\Delta\ge 0} H\left(c^{\mathrm{sim}},\Delta,1\right) = 2\Phi\left(-c\right) = \alpha > \alpha - \eta,$$

thus $\rho_2^*(\alpha,\eta) < 1$.

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

$$\sup_{\Delta \geq 0} H(c, \Delta, \rho) = \sup_{\Delta \in [0, \bar{\Delta}]} H(c, \Delta, \rho) \tag{134}$$

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

$$\begin{aligned}
\frac{\mathrm{d}H(c, \Delta, \rho)}{\mathrm{d}\Delta} &= \phi(\Delta - c)\left[\Phi\left(\frac{(\rho - 1)c - \rho\Delta}{\sqrt{1 - \rho^2}}\right) - \frac{1}{2}\exp\left(\frac{3}{8}\Delta(\Delta - 4c)\right)\right] \\
&\leq \phi(\Delta - c)\left[1 - \frac{1}{2}\exp\left(\frac{3}{8}\Delta(\Delta - 4c^{\text{sim}})\right)\right] \leq 0
\end{aligned}$$

Therefore, (134) holds for all $c \in (0, c^{\text{sim}}]$.

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

$$\begin{aligned}
\alpha - \eta &\geq \sup_{\Delta \in [0, \bar{\Delta}]} H\left(c^{\text{sim}}, \Delta, \rho_2^*(\alpha, \eta)\right) \\
&= \sup_{\Delta \in [0, \bar{\Delta}]} H\left(c^{\text{sim}}, \Delta, \bar{\rho}\right) + \frac{\mathrm{d}H\left(c^{\text{sim}}, \Delta, \tilde{\rho}(\Delta)\right)}{\mathrm{d}\rho}\xi \\
&\geq \sup_{\Delta \in [0, \bar{\Delta}]} H\left(c^{\text{sim}}, \Delta, \bar{\rho}\right) + a\xi
\end{aligned} \tag{135}$$

where

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

Rewrite (135) we get

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

Lastly,

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

Let

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

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

$$\begin{aligned}
\sup_{\Delta \in [0, \bar{\Delta}]} H\left(c^{\text{sim}}, \Delta, \rho\right) &= \sup_{\Delta \in [0, \bar{\Delta}]} H(\bar{c}, \Delta, \rho) + \frac{\mathrm{d}H(\tilde{c}(\Delta), \Delta, \rho)}{\mathrm{d}c}(c^{\text{sim}} - \bar{c}) \\
&\geq \sup_{\Delta \in [0, \bar{\Delta}]} H(\bar{c}, \Delta, \rho) - b(c^{\text{sim}} - \bar{c}).
\end{aligned} \tag{136}$$

In sum, for all $\rho \leq \rho_2^*(\alpha,\eta) - \xi$, by (134), (135), (136),

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

□

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*

$$\sup_{P \in \mathcal{P}} \rho_\ell(b_\ell, b_u) < \rho_2^*(\alpha, \eta), \tag{137}$$

where $\rho_2^*(\alpha,\eta)$ *is defined in Lemma 7 equation (132). Then there is* $\alpha' > \alpha$ *such that*

$$\liminf_n \inf_{P \in \mathcal{P}} P\left(\hat{c}^t \leq \Phi^{-1}(1 - \frac{\alpha'}{2})\right) = 1. \tag{138}$$

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 (133) holds. To show (138), note that

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

Recall that

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

I will show that

$$\sup_{\lambda \in \Lambda} P^s\left(\hat{T}(\theta_\ell) > \bar{c} \text{ or } \hat{T}(\theta_m) > \bar{c}; (\lambda, \hat{\Sigma})\right) \leq \alpha - \eta \text{ w.p.a. } 1,$$

and similarly we can show that

$$\sup_{\lambda \in \Lambda} P^s\left(\hat{T}(\theta_m) > \bar{c} \text{ or } \hat{T}(\theta_u) > \bar{c}; (\lambda, \hat{\Sigma})\right) \leq \alpha - \eta \text{ w.p.a. } 1.$$

To see this,

$$P^s\left(\hat{T}(\theta_\ell) > \bar{c} \text{ or } \hat{T}(\theta_m) > \bar{c}; (\lambda, \hat{\Sigma})\right)$$

$$
\begin{aligned}
=&P^s\left(\max\left\{\min_{b\in\mathcal{B}}\frac{\hat{\lambda}^s_b-\theta_\ell}{\hat{\sigma}_b/\sqrt{n}},\ \min_{b\in\mathcal{B}}\frac{\theta_\ell-\hat{\lambda}^s_b}{\hat{\sigma}_b/\sqrt{n}}\right\}>\bar{c}\text{ or }\max\left\{\min_{b\in\mathcal{B}}\frac{\hat{\lambda}^s_b-\theta_m}{\hat{\sigma}_b/\sqrt{n}},\ \min_{b\in\mathcal{B}}\frac{\theta_m-\hat{\lambda}^s_b}{\hat{\sigma}_b/\sqrt{n}}\right\}>\bar{c};(\lambda,\hat{\Sigma})\right)\\
\le&P^s\left(\max\left\{\min_{b\in\mathcal{B}}\frac{\hat{\lambda}^s_b-\theta_\ell}{\hat{\sigma}_b/\sqrt{n}},\ \min_{b\in\mathcal{B}}\frac{\theta_m-\hat{\lambda}^s_b}{\hat{\sigma}_b/\sqrt{n}}\right\}>\bar{c};(\lambda,\hat{\Sigma})\right)\\
\le&P^s\left(\min\left\{\frac{\hat{\lambda}^s_{b_\ell}-\theta_\ell}{\hat{\sigma}_{b_\ell}/\sqrt{n}},\frac{\hat{\lambda}^s_{b_u}-\theta_\ell}{\hat{\sigma}_{b_u}/\sqrt{n}}\right\}>\bar{c},\ \text{or}\ \frac{\theta_m-\hat{\lambda}^s_{b_u}}{\hat{\sigma}_{b_u}/\sqrt{n}}>\bar{c};(\lambda,\hat{\Sigma})\right)\\
=&P^s\left(\min\left\{\mathbb{Z}^s_{b_\ell},\mathbb{Z}^s_{b_u}+\frac{\theta_u-\theta_\ell}{\hat{\sigma}_{b_u}/\sqrt{n}}\right\}>\bar{c},\ \text{or}\ \frac{\theta_m-\theta_u}{\hat{\sigma}_{b_u}/\sqrt{n}}-\mathbb{Z}^s_{b_u}>\bar{c};(\lambda,\hat{\Sigma})\right)\\
\le&P^s\left(\mathbb{Z}^s_{b_\ell}>\bar{c},\ \mathbb{Z}^s_{b_u}+\frac{\theta_u-\theta_\ell}{\hat{\sigma}_{b_u}/\sqrt{n}}>\bar{c};(\lambda,\hat{\Sigma})\right)+P\left(\frac{\theta_m-\theta_u}{\hat{\sigma}_{b_u}/\sqrt{n}}-\mathbb{Z}^s_{b_u}>\bar{c}\right)\\
=&\Phi\left(-\bar{c},\frac{\theta_u-\theta_\ell}{\hat{\sigma}_{b_u}/\sqrt{n}}-\bar{c};\hat{\rho}_{\ell u}(b_\ell,b_u)\right)+\Phi\left(\frac{\theta_m-\theta_u}{\hat{\sigma}_{b_u}/\sqrt{n}}-\bar{c}\right)\\
\le&\Phi\left(-\bar{c},\Delta-\bar{c};\hat{\rho}_{\ell u}(b_\ell,b_u)\right)+\Phi\left(-\frac{\Delta}{2}-\bar{c}\right)\\
=&H\left(\bar{c},\Delta,\hat{\rho}_\ell(b_\ell,b_u)\right) \qquad (139)
\end{aligned}
$$

where $\Delta=\frac{\theta_u-\theta_\ell}{\hat{\sigma}_{b_u}/\sqrt{n}}\ge 0$,

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

Under (137) and Assumptions 1, 2, 3, 4, and 5, it holds that

$$
\hat{\rho}_\ell(b_\ell,b_u)\le\rho^*_2(\alpha,\eta)-\xi\text{ w.p.a. 1.}
$$

Thus (139) gives w.p.a. 1,

$$
\begin{aligned}
P\left(\hat{T}(\theta_\ell)>\bar{c}\text{ or }\hat{T}(\theta_m)>\bar{c};(\lambda,\hat{\Sigma})\right)&\le H\left(\bar{c},\Delta,\hat{\rho}_\ell(b_\ell,b_u)\right)\\
&\le\sup_{\rho\le\rho^*_2(\alpha,\eta)-\xi}\sup_{\Delta\ge 0}H\left(\bar{c},\Delta,\rho\right)<\alpha-\eta
\end{aligned}
$$

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

Lemma 9. *$p(c)$ is continuous at $c\ge 0$.*

Proof. For $\varepsilon>0$, let

$$
p_k(c,\varepsilon):=P\left(c^{\mathrm{m}}_k(c)\ge T_k>c^{\mathrm{m}}_k(c-\varepsilon)\right)\le P\left(c-\varepsilon<T_k\le c\right).
$$

Then

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

for all $c\ge 0$ since (i) under (114) and $k=\ell,m,u$, or under (116) and $k=\ell,u$, T_k is continuously distributed, (ii) under (116) and $k=m$,

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

But then

$$\begin{aligned}
&p(c-\varepsilon)-p(c)\\
&\leq \max\Big\{P\left(T_\ell > c^{\mathrm{h}}_\ell(c-\varepsilon) \text{ or } T_m > c^{\mathrm{h}}_m(c-\varepsilon)\right) - P\left(T_\ell > c^{\mathrm{h}}_\ell(c) \text{ or } T_m > c^{\mathrm{h}}_m(c)\right),\\
&\quad P\left(T_m > c^{\mathrm{h}}_m(c-\varepsilon) \text{ or } T_u > c^{\mathrm{h}}_u(c-\varepsilon)\right) - P\left(T_m > c^{\mathrm{h}}_m(c) \text{ or } T_u > c^{\mathrm{h}}_u(c)\right)\Big\}\\
&\leq \max\left\{p_\ell(c,\varepsilon)+p_m(c,\varepsilon),\ p_u(c,\varepsilon)+p_m(c,\varepsilon)\right\}\\
&\xrightarrow{\varepsilon\to 0} \to 0
\end{aligned}$$

Thus $p(c)$ is continous at $c \geq 0$. □

Lemma 10. *Let*

$$H(\rho,\alpha^c) = (1-\alpha^c)\Phi\left((1+\rho)\Xi\right) + \alpha^c\Phi\left((\rho-1)\Xi\right),$$

where

$$\Xi = \sqrt{\frac{1}{2\rho}\log\left(\frac{(1-\alpha^c)(1+\rho)}{\alpha^c(1-\rho)}\right)}.$$

It holds that

1. *for all $\alpha \in (0,\frac{1}{2})$, $\alpha^c \in (\frac{\alpha}{2},\alpha)$, there is unique solution $\rho^*(\alpha^c,\alpha) \in (0,1)$ such that*

$$H\left(\rho^*(\alpha^c,\alpha),\alpha^c\right) = 1-\frac{\alpha}{2}. \tag{140}$$

2. *Let $\xi > 0$. For all $\rho \in [0,\rho^*(\alpha^c,\alpha)-\xi]$, there is $\varepsilon > 0$ such that*

$$H\left(\rho,\alpha^c\right) \leq 1-\frac{\alpha}{2}-\varepsilon.$$

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

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

And note that

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

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 $\alpha^{\mathrm{c}} \in (\frac{\alpha}{2}, \alpha)$. Thus $H(\rho, \alpha^{\mathrm{c}})$ is strictly increasing in $\rho \in (0,1)$ and there is unique solution that $H(\rho^*) = 1-\frac{\alpha}{2}$. □

Lemma 11. *Suppose Assumptions 1, 2, 3, 4, and 5 hold. If*

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

then there is $\alpha' > \alpha$ such that

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

Proof. 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

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

The lower bound is

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

and the upper bound is

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

This θ is rejected if

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

Since by construction, $\mathcal{Z}_2 \ge \mathcal{Z}_1$, it suffices to show that

$$\Phi(\mathcal{Z}_1) > \sup_{z_2\ge\mathcal{Z}_1} (1-\alpha^{\mathrm{c}})\Phi\left(\frac{z_2-\rho_{12}\mathcal{Z}_1}{1-\rho_{12}}\right) + \alpha^{\mathrm{c}}\Phi\left(\frac{\rho_{12}\mathcal{Z}_1 - z_2}{1+\rho_{12}}\right). \tag{141}$$

Let

$$H(z_2) = (1-\alpha^{\mathrm{c}})\Phi\left(\frac{z_2-\rho_{12}\mathcal{Z}_1}{1-\rho_{12}}\right) + \alpha^{\mathrm{c}}\Phi\left(\frac{\rho_{12}\mathcal{Z}_1 - z_2}{1+\rho_{12}}\right),$$

and it is easy to see that

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

and

$$H(\mathcal{Z}_1) = (1-\alpha^{\mathrm{c}})\Phi\left(\mathcal{Z}_1\right) + \alpha^{\mathrm{c}}\Phi\left(\frac{\rho_{12}-1}{1+\rho_{12}}\mathcal{Z}_1\right)$$

$$< (1-\alpha^{\mathrm{c}})\Phi\left(\mathcal{Z}_1\right) + \frac{\alpha^{\mathrm{c}}}{2} < \Phi\left(\mathcal{Z}_1\right),$$

where the second line follows from $\mathcal{Z}_1 \geq \bar{c} > 0$. The first order derivative of $H(z_2)$ with respect to z_2 is

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

And $h(z_2) \geq 0$ is equivalent to

$$\log\left(\frac{1-\alpha^{\mathrm{c}}}{\alpha^{\mathrm{c}}}\frac{1+\rho_{12}}{1-\rho_{12}}\right) \geq \log\left(\phi\left(\frac{\rho_{12}\mathcal{Z}_1 - z_2}{1+\rho_{12}}\right)\right) - \log\left(\phi\left(\frac{z_2 - \rho_{12}\mathcal{Z}_1}{1-\rho_{12}}\right)\right)$$
$$= \frac{2\rho_{12}\left(z_2 - \mathcal{Z}_1\rho_{12}\right)^2}{\left(1-\rho_{12}^2\right)^2}. \tag{142}$$

(i) If $2\alpha^c - 1 < \rho \leq 0$, then (142) holds trivially, and thus

$$\sup_{z_2 \geq \mathcal{Z}_1} H(z_2) = \lim_{z_2 \to \infty} H(z_2) < \Phi\left(\mathcal{Z}_1\right),$$

and (141) holds.

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

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

Thus

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

(iii) if $\rho \in \left(0, \rho^*(\alpha, \alpha^c) - \frac{\xi}{2}\right)$, straightforward calculation shows that $H(z_2)$ increases in $[\max\left\{\mathcal{Z}_1, z_2^*\right\}, z_2^*]$ and decreases in $[z_2^*, +\infty)$, thus

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

where the last inequality is by Lemma 10. □

Lemma 12. *Suppose Assumptions 1, 2, 3, and 4 hold. And $\hat{\Lambda}_\eta$ is defined as in (68). Let $\alpha \in (0, \frac{1}{2})$, $\alpha^c \in (\frac{\alpha}{2}, \alpha)$, $\eta \in [0, \frac{\alpha - \alpha^c}{2})$. In addition, assume (36). It holds that*

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

Proof. If $\eta = 0$, trivial, following from the discussion around (17). If $\eta \in (0, \frac{\alpha - \alpha^c}{2})$, trivial too, as $\hat{c}^{\mathrm{t}} \xrightarrow{p} \Phi^{-1}(1 - \alpha + \eta)$. □

D Proof for Section 4

This proof follows closely from Kaido et al. (2019). The key difference is that (i) I adjust the covariance matrix Σ for the estimation uncertainty in $\hat{\varphi}$; (ii) I linearize the moment conditions and $\hat{\lambda}_\ell$, $\hat{\lambda}_u$ with a slightly different Taylor expansion, thus the definition of the Jacobian D is different from theirs; (iii) because of different objective function, the linear program here has a different structure from theirs.

D.1 Notation

The ε expansion of set A is defined as

$$A^{\epsilon} = \left\{ a \in \mathbb{R}^{d_a} : d_H(a, A) \leq \epsilon \right\},$$

where d_H is the Hausdorff distance

$$d_H(a, A) = \inf_{\tilde{a} \in A} \|a - \tilde{a}\| .$$

Let $m(X, \beta)$ be a K dimensional vector, with $K = J + d_\ell + d_u + d_\varphi$, and

$$m(X, \beta) = \begin{pmatrix} m_{\mathcal{J}}(X; \beta, \varphi) \\ m_\ell(X; \beta, \varphi) \\ m_u(X; \beta, \varphi) \\ m_\varphi(X) \end{pmatrix}.$$

That is, with slight abuse of notation, I denote m_{J+1} as the first element of m_ℓ, etc. For each $(\beta, \varphi), (\tilde{\beta}, \varphi) \in \overline{\mathcal{B}} \times \Phi$ and P, let $\Omega_P\Big((\beta, \varphi), (\tilde{\beta}, \tilde{\varphi})\Big) \in \mathbb{R}^{K \times K}$ denote

$$\Omega_P\Big((\beta, \varphi), (\tilde{\beta}, \tilde{\varphi})\Big) = \operatorname{cov}\Big(\ m(X_i; \beta, \varphi)\ ,\ m'(X_i; \tilde{\beta}, \tilde{\varphi})\ \Big),$$

and I use $\Omega_P(\beta)$ for $\Omega_P\left((\beta, \varphi_P), (\beta, \varphi_P)\right)$. And let $\omega_j(\beta, \varphi) = \sqrt{\Omega_{P,jj}\left((\beta, \varphi), (\beta, \varphi)\right)}$. Let

$$\widetilde{\mathbb{G}}_n(\beta, \varphi) = \begin{pmatrix} \widetilde{\mathbb{G}}_{n,\mathcal{J}}(\beta, \varphi) \\ \widetilde{\mathbb{G}}_{n,\ell}(\beta, \varphi) \\ \widetilde{\mathbb{G}}_{n,u}(\beta, \varphi) \\ \widetilde{\mathbb{G}}_{n,\varphi} \end{pmatrix} = \begin{pmatrix} \sqrt{n}\left(\bar{m}_{\mathcal{J}}(\beta, \varphi) - E\left[m_{\mathcal{J}}(X_i; \beta, \varphi)\right]\right) \\ \sqrt{n}\left(\bar{m}_\ell(\beta, \varphi) - E\left[m_\ell(X_i; \beta, \varphi)\right]\right) \\ \sqrt{n}\left(\bar{m}_u(\beta, \varphi) - E\left[m_u(X_i; \beta, \varphi)\right]\right) \\ \sqrt{n}\left(\bar{m}_\varphi - E\left[m_\varphi(X_i)\right]\right) \end{pmatrix}. \tag{143}$$

Table 5 summarizes other notations used in this proof.

D.2 Additional Assumptions

Assumption 10. *All distributions $P \in \mathcal{P}$ satisfy the following:*

1. *$E_P\left[m_j\left(X_i, \beta\right)\right] \leq 0, j = 1, \ldots, J_1$ and $E_P\left[m_j\left(X_i, \beta\right)\right] = 0, j = J_1 + 1, \ldots, J_1 + J_2$ for some $\beta \in \overline{\mathcal{B}}$;*

Table 5: Notation

Notation	Defined in Equation	Notation	Defined in Equation
$\widetilde{\mathbb{G}}_n$	(143)	$\zeta_{n,j}$, $\zeta_{n,k}$	(155), (163)
$\gamma_{1,P,j}(\beta)$, $\gamma_{1,P,k}(\beta)$	(146), (164)	c_{π^*}	(200)
$\gamma_{0,P,j}(\beta)$	(147)	$c_n^I(\beta_n)$	(201)
π_{1j}	(150)	$c_{n,\rho}^I(\beta)$	(205)
π_{1j}^*	(151)	$\hat{c}_{n,\rho}$	(186)
$\bar{u}_{n,j,\beta_n}(\Delta)$	(157), (160)	$\hat{\mu}_{n,j}$, $j=1,...,2R_1$	(145)
$u_{n,j,\beta_n}(\Delta) u_{n,k,\beta_n}(\Delta)$	(182) (183)	$\hat{\sigma}_j^M$, $j=1,...,2R_1$	(144)
$U_{n,\ell}$, $U_{n,u}$	(184), (185)	$\boldsymbol{\Delta}_{n,\rho}$	(165)
$v_{n,j,\beta_n'}^I(\Delta)$	(203)	ϕ_j^*	(204)
$V_n^I(\beta_n',c)$	(202)	$\mathbb{Z}_{n,j,\beta_n}(\Delta)$, $\mathbb{Z}_{n,k,\beta_n}$	(158), (162)
$v_{n,j,\beta_n'}^s(\Delta)$	(253)	$\hat{\eta}_{n,j,\beta_n}(\Delta)$	(159), (161)
$V_n^s(\beta_n',c)$	(252)	$\hat{\tau}_{n,j,\beta}$, $\hat{\tau}_{n,k,\beta}$	(170), (176)
$\mathfrak{w}_j(\Delta)$	(195)	$\tilde{\mu}_j$, $\tilde{\mu}_k$	(168), (169), (175)
$\mathfrak{W}$, $\mathfrak{W}^*$	(196), (209)	U_n^*	(208)
$\mathfrak{W}^\delta(c)$	(197)		

2. *$\{X_i, i \geq 1\}$ are i.i.d.;*

3. *There is $\epsilon > 0$ such that $\sigma_{P,j}^2(\beta) \in [\epsilon, 1/\epsilon]$ for $j = 1, \ldots, J_1 + J_2, \ell, u$ for all $\beta \in \overline{\mathcal{B}}$;*

4. *For some $\delta > 0$ and $M \in (0, \infty)$ and*

$$E_P\left[\sup_{\beta\in\mathcal{B}} |m_j(X_i;\beta,\varphi_P)|^{2+\delta}\right] \leq M, \ \forall j = 1, ..., K.$$

Assumption 11. *All distributions $P \in \mathcal{P}$ satisfy one of the following two conditions for some constants $\omega > 0, \varepsilon > 0, M < \infty$. The functions $m_j(X_i;\beta,\varphi)$, $j = 1, ..., K$, are defined on $\mathcal{X} \times \overline{\mathcal{B}}^\varepsilon \times \Phi^\varepsilon$. There exists $R_1 \in \mathbb{N}$, $1 \leq R_1 \leq J_1/2$ and measurable functions $t_j : \mathcal{X} \times \overline{\mathcal{B}}^\varepsilon \times \Phi^\varepsilon \to [0, M]$, $j \in \mathcal{R}_1 = \{1, ..., R_1\}$ such that for each $j \in \mathcal{R}_1$,*

$$m_{j+R_1}(X;\beta,\varphi) = -m_j(X;\beta,\varphi) - t_j(X;\beta,\varphi)$$

For each $j \in \mathcal{R}_1 \cap \mathcal{J}_1(P,\beta,\varepsilon)$, and any choice $\ddot{j} \in \{j, j+R_1\}$, one of the following holds:

1. *One has*

$$\inf_{\beta\in\mathcal{B}(P)} \text{eig}\left(\Sigma_{\ddot{\mathcal{J}}_k}\right) \geq \omega,$$

where

$$\begin{aligned}\ddot{\mathcal{J}}_k = &\{\ddot{j} : j \in \mathcal{R}_1 \cap \mathcal{J}_1(P,\beta,\varepsilon)\}\\ &\cup \mathcal{J}_1(P,\beta,\varepsilon) \backslash \{1, ..., 2R_1\}\\ &\cup \mathcal{J}_k(P,\beta;\varepsilon).\end{aligned}$$

and $\mathcal{J}_1$ and $\mathcal{J}_k$ are defined in (65) and (66).

2. *There is* $t_0 : \overline{\mathcal{B}}^{\varepsilon} \times \Phi^{\varepsilon} \to [0, M]$ *such that*

$$\lambda_\ell^\dagger \left(E_P\left[m_\ell(X_i;\beta,\varphi)\right]\right) - \lambda_u^\dagger \left(E_P\left[m_u(X_i;\beta,\varphi)\right]\right) = -t_0(\beta,\varphi).$$

Let $k = \ell, u$ *and*

$$\mathring{\mathcal{J}}_k = \ddot{\mathcal{J}}_k \backslash [k].$$

One has

$$\inf_{\beta\in\mathcal{B}(P)} \operatorname{eig}\left(\Sigma_{\mathring{\mathcal{J}}_\ell}\right) \geq \omega, \text{ and } \inf_{\beta\in\mathcal{B}(P)} \operatorname{eig}\left(\Sigma_{\mathring{\mathcal{J}}_u}\right) \geq \omega.$$

Assumption 11 is an analog of Kaido et al. (2019) Assumption E3.2. Under the condition that the sum of two moments is non-positive, we only need the rank condition to hold for each of the moments, but not jointly.

Assumption 12. *All distributions* $P \in \mathcal{P}$ *satisfy the following conditions:*

1. *The class of functions* $\left\{\omega_{P,j}^{-1}(\beta,\varphi)m_j(\cdot,\beta,\varphi) : \mathcal{X} \to \mathbb{R}, \beta \in \overline{\mathcal{B}}, \varphi \in \Phi\right\}$ *is measurable for each* $j = 1,\ldots,K$.

2. *The empirical process* $\widetilde{\mathbb{G}}_n$ *is uniformly asymptotically* ϱ_P*-equicontinuous. That is, for any* $\epsilon > 0$,

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \sup_{P\in\mathcal{P}} P\left(\sup_{\varrho_P((\beta,\varphi),(\tilde\beta,\tilde\varphi))<\delta} \left\|\widetilde{\mathbb{G}}_n(\beta,\varphi) - \widetilde{\mathbb{G}}_n(\tilde\beta,\tilde\varphi)\right\| > \epsilon\right) = 0$$

3. Ω_P *satisfies*

$$\lim_{\delta\downarrow 0} \sup_{\left\|\left((\beta_1,\varphi_1),(\tilde\beta_1,\tilde\varphi_1)\right)-\left((\beta_2,\varphi_2),(\tilde\beta_2,\tilde\varphi_2)\right)\right\|<\delta} \sup_{P\in\mathcal{P}} \left\|\Omega_P\left((\beta_1,\varphi_1),(\tilde\beta_1,\tilde\varphi_1)\right) - \Omega_P\left((\beta_2,\varphi_2),(\tilde\beta_2,\varphi_2)\right)\right\| = 0.$$

D.3 Details of the Inference Procedure

If Assumption 11 is invoked, we make the following adjustment to the inference procedure.

In (51), we replace the estimated standard deviation $\hat\sigma_1, ..., \hat\sigma_{2R_1}$ with $\hat\sigma_1^M, ..., \hat\sigma_{2R_1}^M$. For $j = 1, ..., R_1$, let $[j] = j + R_1$ and

$$\hat\sigma_j^M(\beta) = \hat\sigma_{[j]}^M(\beta) = \hat\mu_{n,j}(\beta)\hat\sigma_j(\beta) + \left(1 - \hat\mu_{n,j}(\beta)\right)\hat\sigma_{[j]}(\beta). \tag{144}$$

with

$$\hat\mu_{n,[j]}\left(\beta\right) = \min\left\{\max\left(0, \frac{\frac{\bar m_{n,j}(\beta,\hat\varphi)}{\hat\sigma_{n,j}(\beta,\hat\varphi)}}{\frac{\bar m_{n,[j]}(\beta,\hat\varphi)}{\hat\sigma_{n,[j]}(\beta,\hat\varphi)} + \frac{\bar m_{n,j}(\beta,\hat\varphi)}{\hat\sigma_{n,j}(\beta,\hat\varphi)}}\right), 1\right\},$$
$$\hat\mu_{n,j}\left(\beta\right) = 1 - \hat\mu_{n,[j]}\left(\beta,\hat\varphi\right). \tag{145}$$

In (55), if

$$\hat\xi_{n,j}(\beta) = 0 = \hat\xi_{n,[j]}(\beta),$$

we replace $\mathbb{Z}_{n,[j]}^b(\beta)$ with $-\mathbb{Z}_{n,j}^b(\beta)$ and $\hat D_{n,[j]}(\beta)$ with $-\hat D_{n,j}(\beta)$ for $j = 1, ..., R_1$.

And if Assumption 11.2 is invoked, we do similar adjustment for $\hat{\lambda}_\ell$ and $\hat{\lambda}_u$. That is, if $\hat{\xi}_{n,0}(\beta) = 0$, we replace $\hat{\sigma}_\ell(\beta)$ with $\hat{\sigma}_u(\beta)$ in (50), and replace $\mathbb{Z}^*_{n,\ell}(\beta)$ with $-\mathbb{Z}^*_u(\beta)$ and $\hat{D}_{n,\ell}(\beta)$ with $-\hat{D}_{n,u}(\beta)$ in (53), (54).

D.4 Proof of Theorem 1

Proof. Let $\gamma_P = (\gamma_{1,P}, \gamma_{2,P}, \gamma_{3,P})$, where $\gamma_{1,P} = (\gamma_{1,P,1}, \ldots, \gamma_{1,P,J}, \gamma_{1,P,0})$ with

$$\gamma_{1,P,j}(\beta) = \sigma^{-1}_{P,j}(\beta) E_P \left[m_j \left(X_i, \beta, \varphi_P \right) \right], j = 1, ..., J, \tag{146}$$

$$\gamma_{1,P,0}(\beta) = \frac{\lambda_{P,\ell}(\beta, \varphi_P) - \lambda_{P,u}(\beta, \varphi_P)}{\max \left\{ \sigma_{P,u}(\beta), \sigma_{P,\ell}(\beta), \sigma_{P,u\ell}(\beta) \right\}} \tag{147}$$

$\gamma_{2,P} = \left(\text{vech}\left(\Omega_P(\beta) \right), \text{vec}\left(D_P(\beta) \right), \text{vec}\left(G_P(\beta) \right) \right)$, and $\gamma_{3,P} = P$. We proceed in steps.

Step 1. Let

$$\{P_n, \beta_n, \theta_n\} \in \{(P, \beta, \theta) : P \in \mathcal{P}, \beta \in \mathcal{B}(P), \theta \in [\lambda_{P,\ell}(\beta, \varphi_P), \lambda_{P,u}(\beta, \varphi_P)]\}$$

be a sequence such that

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\beta\in\mathcal{B}(P)} \inf_{\theta\in[\lambda_\ell(\beta,\varphi_P),\lambda_u(\beta,\varphi_P)]} P\left(\theta_n \in CI_n\right) = \liminf_{n\to\infty} P_n\left(\theta_n \in CI_n\right). \tag{148}$$

Let $\{l_n\}$ be a subsequence of $\{n\}$ such that

$$\liminf_{n\to\infty} P_n\left(\theta_n \in CI_n\right) = \lim_{n\to\infty} P_{l_n}\left(\theta_{l_n} \in CI_{l_n}\right). \tag{149}$$

Then there is a further subsequence $\{a_n\}$ of $\{l_n\}$ such that

$$\lim_{a_n\to\infty} \kappa^{-1}_{a_n} \sqrt{a_n} \gamma_{1,P_{a_n},j}\left(\beta_{a_n}\right) = \pi_{1,j} \in \mathbb{R}_{[-\infty]}, \quad j = 0, 1, \ldots, J. \tag{150}$$

To simplify notation, I write (P_n, β_n, θ_n) to refer to $(P_{a_n}, \beta_{a_n}, \theta_{a_n})$ throughout this Appendix. For $j = 0, 1, ..., J$, let

$$\pi^*_{1,j} = \begin{cases} 0 & \text{if } \pi_{1,j} = 0 \\ -\infty & \text{if } \pi_{1,j} < 0 \end{cases}. \tag{151}$$

The true value θ_n is covered when

$$\left\{ \begin{array}{c} \inf_{\tilde{\beta}} \hat{\lambda}_\ell(\tilde{\beta}) - \frac{\hat{\sigma}_\ell(\tilde{\beta})}{\sqrt{n}} \hat{c}(\tilde{\beta}) \\ \text{s.t. } \tilde{\beta} \in \overline{\mathcal{B}}, \quad \frac{\sqrt{n}\bar{m}_{n,j}(\tilde{\beta},\hat{\varphi})}{\hat{\sigma}_j(\tilde{\beta})} \leq \hat{c}(\tilde{\beta}), \forall j = 1, ..., J \end{array} \right\} \leq \theta_n$$
$$\leq \left\{ \begin{array}{c} \sup_{\tilde{\beta}} \hat{\lambda}_u(\tilde{\beta}) + \frac{\hat{\sigma}_u(\tilde{\beta})}{\sqrt{n}} \hat{c}(\tilde{\beta}) \\ \text{s.t. } \tilde{\beta} \in \overline{\mathcal{B}}, \quad \frac{\sqrt{n}\bar{m}_{n,j}(\tilde{\beta},\hat{\varphi})}{\hat{\sigma}_j(\tilde{\beta})} \leq \hat{c}(\tilde{\beta}), \forall j = 1, ..., J \end{array} \right\}$$

$$\Leftrightarrow \left\{ \begin{array}{c} \inf_{\Delta} \frac{\sqrt{n}\left(\hat{\lambda}_{\ell}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)-\theta_n\right)}{\hat{\sigma}_{\ell}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)} - \hat{c}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right) \\ \text{s.t. } \Delta \in \frac{\sqrt{n}}{\rho}\left(\overline{\mathcal{B}}-\beta_n\right), \frac{\sqrt{n}\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)}{\hat{\sigma}_j\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)} \leq \hat{c}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right), \forall j=1,...,J \end{array} \right\} \leq 0 \text{ and}$$

$$\left\{ \begin{array}{c} \inf_{\Delta} \frac{\sqrt{n}\left(\theta_n-\hat{\lambda}_{u}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)\right)}{\hat{\sigma}_{u}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)} - \hat{c}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right) \\ \text{s.t. } \Delta \in \frac{\sqrt{n}}{\rho}\left(\overline{\mathcal{B}}-\beta_n\right), \quad \frac{\sqrt{n}\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)}{\hat{\sigma}_j\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)} \leq \hat{c}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right), \forall j=1,...,J \end{array} \right\} \leq 0 \tag{152}$$

where in (152), I simply replace $\tilde{\beta}$ with $\beta_n + \frac{\Delta\rho}{\sqrt{n}}$.

Step 2. Simplify

$$\frac{\sqrt{n}\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)}{\hat{\sigma}_j\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)}, \frac{\hat{\lambda}_{\ell}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)-\theta_n}{\hat{\sigma}_{\ell}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)/\sqrt{n}} \text{ and } \frac{\theta_n-\hat{\lambda}_{u}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)}{\hat{\sigma}_{u}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)/\sqrt{n}}.$$

Straightforward calculation gives

$$\frac{\sqrt{n}\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)}{\hat{\sigma}_j\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)} = \left(1+\frac{\sigma_{P_n,j}\left(\beta_n\right)}{\hat{\sigma}_j\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)}-1\right) \times \left(\frac{\widetilde{\mathbb{G}}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\varphi_{P_n}\right)}{\sigma_{P_n,j}\left(\beta_n\right)}\right.$$

$$+\frac{\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)-\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\varphi_{P_n}\right)}{\sigma_{P_n,j}\left(\beta_n\right)/\sqrt{n}} \tag{153}$$

$$\left.\frac{E_{P_n}\left[m_j\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\varphi_{P_n}\right)\right]-E_{P_n}\left[m_j\left(\beta_n,\varphi_{P_n}\right)\right]}{\sigma_{P_n,j}\left(\beta_n\right)/\sqrt{n}}+\sqrt{n}\gamma_{1,n,j}\left(\beta_n\right)\right). \tag{154}$$

Then I further simplify each element. For (153),

$$\begin{aligned} &\frac{\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)-\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\varphi_{P_n}\right)}{\sigma_{P_n,j}\left(\beta_n\right)/\sqrt{n}} \\ =&\frac{E_{P_n}\left[m_j\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\varphi\right)\right]_{\varphi=\hat{\varphi}}-E_{P_n}\left[m_j\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\varphi_{P_n}\right)\right]}{\sigma_{P_n,j}\left(\beta_n\right)/\sqrt{n}}+\zeta_{n,j} \\ =&\frac{\nabla_{\varphi'}E_{P_n}\left[m_j\left(\beta_n,\bar{\varphi}\right)\right]\nabla_{m'_{\varphi}}\bar{\varphi}^{\dagger}}{\sigma_{P_n,j}\left(\beta_n\right)}\widetilde{\mathbb{G}}_{n,\varphi}+\zeta_{n,j} \end{aligned}$$

where $\widetilde{\mathbb{G}}_n$ is defined in (143), $\bar{\varphi}$ is between φ_{P_n} and $\hat{\varphi}$, $\nabla_{m'_{\varphi}}\bar{\varphi}^{\dagger} = \nabla_{m'_{\varphi}}\varphi^{\dagger}\Big|_{\tilde{m}_{\varphi}}$ for some $\tilde{m}_{\varphi}$ between $\bar{m}_{\varphi}$ and $E[m_{\varphi}(W)]$,

$$\zeta_{n,j} = \frac{\widetilde{\mathbb{G}}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}_n\right)-\widetilde{\mathbb{G}}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\varphi_{P_n}\right)}{\sigma_{P_n,j}\left(\beta_n\right)}. \tag{155}$$

As for (154), there is some $\bar{\beta}_n$ between β_n and $\beta_n + \frac{\Delta\rho}{\sqrt{n}}$ such that

$$\frac{E_{P_n}\left[m_j\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \varphi_{P_n}\right)\right] - E_{P_n}\left[m_j\left(\beta_n, \varphi_{P_n}\right)\right]}{\sigma_{P_n,j}\left(\beta_n\right)/\sqrt{n}} = \frac{\sigma_{P_n,j}\left(\bar{\beta}_n\right)}{\sigma_{P_n,j}\left(\beta_n\right)} D_{P_n,j}\left(\bar{\beta}_n\right)\Delta\rho$$

In sum,

$$\frac{\sqrt{n}\bar{m}_{n,j}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \hat{\varphi}\right)}{\hat{\sigma}_{n,j}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)} = \bar{u}_{n,j,\beta_n}(\Delta) \tag{156}$$

where

$$\bar{u}_{n,j,\beta_n}(\Delta) = \left(1 + \hat{\eta}_{n,j,\beta_n}(\Delta)\right)\left(\mathbb{Z}_{n,j,\beta_n} + \frac{\sigma_{P_n,j}\left(\bar{\beta}_n\right)}{\sigma_{P_n,j}\left(\beta_n\right)} D_{P_n,j}\left(\bar{\beta}_n\right)\Delta\rho + \sqrt{n}\gamma_{1,P_n,j}\left(\beta_n, \varphi_{P_n}\right)\right), \tag{157}$$

$$\mathbb{Z}_{n,j,\beta_n}(\Delta) = \frac{\widetilde{\mathbb{G}}_{n,j}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \varphi_{P_n}\right) + \nabla_{\varphi'} E_{P_n}\left[m_j\left(\bar{\beta}_n, \varphi_{P_n}\right)\right]\nabla_{m'_\varphi}\bar{\varphi}^{\dagger}\widetilde{\mathbb{G}}_{n,\varphi}}{\sigma_{P_n,j}\left(\beta_n\right)} + \zeta_{n,j}, \tag{158}$$

$$\hat{\eta}_{n,j,\beta_n}(\Delta) = \frac{\sigma_{P_n,j}\left(\beta_n\right)}{\hat{\sigma}_j\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)} - 1. \tag{159}$$

Similarly, for $k = \ell, u$, it holds that

$$\frac{\sqrt{n}\left(\hat{\lambda}_\ell\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \hat{\varphi}\right) - \theta_n\right)}{\hat{\sigma}_\ell\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)} = \bar{u}_{n,\ell,\beta_n}(\Delta),$$

$$\frac{\sqrt{n}\left(\theta_n - \hat{\lambda}_u\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \hat{\varphi}\right)\right)}{\hat{\sigma}_u\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)} = \bar{u}_{n,u,\beta_n}(\Delta),$$

where

$$\bar{u}_{n,k,\beta_n}(\Delta) = \left(1 + \hat{\eta}_{n,k,\beta_n}(\Delta)\right)\left(\mathbb{Z}_{n,k,\beta_n} + \frac{\sigma_{P_n,k}\left(\bar{\beta}_n\right)}{\sigma_{P_n,k}\left(\beta_n\right)} D_{P_n,k}\left(\bar{\beta}_n\right)\Delta\rho + \sqrt{n}\gamma_{1,P_n,k}(\beta_n)\right) \tag{160}$$

with

$$\hat{\eta}_{n,k,\beta_n}(\Delta) = \frac{\sigma_{P_n,k}\left(\beta_n\right)}{\hat{\sigma}_k\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)} - 1, \tag{161}$$

$$\mathbb{Z}_{n,k,\beta_n} = \frac{\widetilde{\mathbb{G}}_{n,k}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \varphi_{P_n}\right) + \nabla_{m'_k}\lambda_{P_n,k}\nabla_{\varphi'} E_{P_n}\left[m_k\left(\bar{\beta}_n, \varphi_{P_n}\right)\right]\nabla_{m'_\varphi}\bar{\varphi}^{\dagger}\widetilde{\mathbb{G}}_{n,\varphi}}{\sigma_{P_n,j}\left(\beta_n\right)} + \zeta_{n,k}, \tag{162}$$

$$\zeta_{n,k} = \frac{\widetilde{\mathbb{G}}_{n,k}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \hat{\varphi}_n\right) - \widetilde{\mathbb{G}}_{n,k}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \varphi_{P_n}\right)}{\sigma_{P_n,k}\left(\beta_n\right)}, \tag{163}$$

$$\gamma_{1,P_n,\ell}(\beta_n) = \frac{\lambda_\ell\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \varphi_P\right) - \theta_n}{\sigma_{P_n,\ell}\left(\beta_n\right)}, \quad \gamma_{1,P_n,u}(\beta_n) = \frac{\theta_n - \lambda_u\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \varphi_P\right)}{\sigma_{P_n,u}\left(\beta_n\right)}. \tag{164}$$

Therefore,

$$(152) \Leftrightarrow \left\{ \begin{array}{c} \inf_\Delta \bar{u}_{n,\ell,\beta_n}(\Delta) - \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right) \\ \text{s.t. } \Delta \in \frac{\sqrt{n}}{\rho}\left(\overline{\mathcal{B}} - \beta_n\right), \\ \bar{u}_{n,j,\beta_n}(\Delta) \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j = 1, ..., J \end{array} \right\} \leq 0, \text{ and } \left\{ \begin{array}{c} \inf_\Delta \bar{u}_{n,u,\beta_n}(\Delta) - \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right) \\ \text{s.t. } \Delta \in \frac{\sqrt{n}}{\rho}\left(\overline{\mathcal{B}} - \beta_n\right), \\ \bar{u}_{n,j,\beta_n}(\Delta) \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j = 1, ..., J \end{array} \right\}$$

Denote

$$\boldsymbol{\Delta}_{n,\rho} = \frac{\sqrt{n}}{\rho}\left(\overline{\mathcal{B}} - \beta_n\right) \cap \boldsymbol{\Delta}, \tag{165}$$

with $\boldsymbol{\Delta} = \left\{x \in \mathbb{R}^d : |x_i| \leq 1, i = 1, \ldots, d\right\}$. Then the event in (152) is implied by

$$\left\{ \begin{array}{c} \inf_\Delta \bar{u}_{n,\ell,\beta_n}(\Delta) - \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right) \\ \text{s.t. } \Delta \in \boldsymbol{\Delta}_{n,\rho}, \\ \bar{u}_{n,j,\beta_n}(\Delta) \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j = 1, ..., J \end{array} \right\} \leq 0, \text{ and } \left\{ \begin{array}{c} \inf_\Delta \bar{u}_{n,u,\beta_n}(\Delta) - \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right) \\ \text{s.t. } \Delta \in \boldsymbol{\Delta}_{n,\rho}, \\ \bar{u}_{n,j,\beta_n}(\Delta) \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j = 1, ..., J \end{array} \right\} \leq 0. \tag{166}$$

Step 3. This step is used only when Assumption 11 is invoked. When this assumption is invoked, recall we use modification in Section D.3. For each $j = 1, ..., R_1$ such that

$$\pi^*_{1,j} = \pi^*_{1,j+R_1} = 0, \tag{167}$$

let

$$\tilde{\mu}_j = 1 - \tilde{\mu}_{j+R_1}, \tag{168}$$

$$\tilde{\mu}_{j+R_1} = \left\{ \begin{array}{cc} 0 & \text{if } \gamma_{1,P_n,j}\left(\beta\right) = \gamma_{1,P_n,j+R_1}\left(\beta\right) = 0 \\ \frac{\hat{\tau}_{n,j,\beta_n}(\Delta)}{\hat{\tau}_{n,j,\beta_n}(\Delta) + \hat{\tau}_{n,j+R_1,\beta_n}(\Delta)} & \text{otherwise,} \end{array} \right. \tag{169}$$

with

$$\hat{\tau}_{n,j,\beta}(\Delta) = \gamma_{1,P_n,j}\left(\beta_n\right)\left(1 + \hat{\eta}_{n,j,\beta_n}\left(\Delta\right)\right). \tag{170}$$

For each $j = 1, ..., R_1$, replace the constraint indexed by j, that is

$$\frac{\sqrt{n}\bar{m}_{n,j}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \hat{\varphi}\right)}{\hat{\sigma}^M_j\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)} \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \tag{171}$$

with the following weighted sum of the paired inequalities

$$\tilde{\mu}_j \frac{\sqrt{n}\bar{m}_{n,j}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \hat{\varphi}\right)}{\hat{\sigma}^M_j\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)} - \tilde{\mu}_{j+R_1} \frac{\sqrt{n}\bar{m}_{j+R_1,n}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \hat{\varphi}\right)}{\hat{\sigma}^M_{j+R_1}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)} \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \tag{172}$$

and replace the constraint indexed by $j+R_1$, that is

$$\frac{\sqrt{n}\bar{m}_{n,j+R_1}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)}{\hat{\sigma}^M_{j+R_1}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)}\leq\hat{c}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right), \tag{173}$$

with

$$-\tilde{\mu}_j\frac{\sqrt{n}\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)}{\hat{\sigma}^M_{j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)}+\tilde{\mu}_{j+R_1}\frac{\sqrt{n}\bar{m}_{j+R_1,n}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)}{\hat{\sigma}^M_{j+R_1}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)}\leq\hat{c}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right). \tag{174}$$

It then follows from Assumption 11 that these replacements are conservative because

$$\frac{\bar{m}_{n,j+R_1}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)}{\hat{\sigma}^M_{j+R_1}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)}\leq-\frac{\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)}{\hat{\sigma}^M_{j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)},$$

and therefore (172) implies (171) and (174) implies (173). Similarly, if Assumption 11.2 is invoked, for $k=\ell,u$, replace

$$\bar{u}_{n,k,\beta_n}(\Delta)\leq\hat{c}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)$$

with

$$\tilde{\mu}_k\bar{u}_{n,k,\beta_n}(\Delta)-\tilde{\mu}_{[k]}\bar{u}_{n,[k],\beta_n}(\Delta)\leq\hat{c}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right),$$

where $\tilde{\mu}_u=1-\tilde{\mu}_\ell$ and

$$\tilde{\mu}_\ell=\begin{cases}0 & \text{if } \gamma_{1,P_n,\ell}(\beta)=\gamma_{1,P_n,u}(\beta)=0\\ \frac{\hat{\tau}_{n,u,\beta_n}(\Delta)}{\hat{\tau}_{n,\ell,\beta_n}(\Delta)+\hat{\tau}_{n,u,\beta_n}(\Delta)} & \text{otherwise,}\end{cases} \tag{175}$$

with

$$\hat{\tau}_{n,k,\beta}(\Delta)=\gamma_{1,P_n,k}(\beta_n)\left(1+\hat{\eta}_{n,k,\beta_n}(\Delta)\right). \tag{176}$$

Step 4. Next, I show that we can replace the term $\sqrt{n}\gamma_{1,P_n,j}(\beta_n)$ with $\pi^*_{1,j}$. For $j=1,...,J_1$,

$$\pi^*_{1,j}=0\Rightarrow\pi^*_{1,j}\geq\sqrt{n}\gamma_{1,P_n,j}(\beta_n), \tag{177}$$

$$\pi^*_{1,j}=-\infty\Rightarrow\sqrt{n}\gamma_{1,P_n,j}(\beta_n)\to-\infty. \tag{178}$$

For any constraint j for which $\pi^*_{1,j}=0$, (177) yields that replacing $\sqrt{n}\gamma_{1,P_n,j}(\beta_n)$ in (166) with $\pi^*_{1,j}$ introduces a conservative distortion. Under Assumption 11, for any j such that (167) holds, the substitutions in (172) and (174) yield

$$\tilde{\mu}_j\gamma_{1,P_n,j}(\beta_n,\varphi_{P_n})\left(1+\hat{\eta}_{n,j,\beta_n}(\Delta)\right)-\tilde{\mu}_{j+R_1}\gamma_{1,P_n,j+R_1}(\beta_n,\varphi_{P_n})\left(1+\hat{\eta}_{n,j+R_1,\beta_n}(\Delta)\right)=0$$

and therefore replacing this term with $\pi^*_{1,j}=0=\pi^*_{1,j+R_1}$ is inconsequential. Same applies to the constraints on $\tilde{u}_{n,u,\beta_n}$ and $\tilde{u}_{n,\ell,\beta_n}$.

For any j for which $\pi^*_{1,j}=-\infty$, (178) yields that for n large enough, $\sqrt{n}\gamma_{1,P_n,j}(\beta_n)$ can be replaced with $\pi_{1,j}^*$. To see this, note that under Assumption 7, Assumption 10.3, Assumption 12.3 and $\Delta \in \mathbf{\Delta}_{n,\rho}^d$, it follows that

$$\frac{\sigma_{F_n,j}\left(\bar{\beta}_n\right)}{\sigma_{F_n,j}\left(\beta_n\right)} D_{F_n,j}\left(\bar{\beta}_n\right)\Delta\rho = O(1).$$

Together with Lemma 21.1 and Lemma 22, it holds that

$$P_n\left(\max_{j:\pi_{1,j}^*=-\infty} \bar{u}_{n,j,\beta_n}(\Delta) \leq 0, \quad \forall \Delta \in \mathbf{\Delta}_{n,\rho}^d\right) \to 1. \tag{179}$$

Since $\hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right) \geq 0$ by construction, $\bar{u}_{n,j,\beta_n}(\Delta) \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)$ with $\pi_{1,j}^* = -\infty$ is asmptotically not bingding and thus negligible. We therefore have that for $n \geq N$,

$$\begin{aligned}
&P_n\left(\left\{\begin{array}{c} \inf_\Delta \bar{u}_{n,\ell,\beta_n}(\Delta) \\ \text{s.t. } \Delta \in \mathbf{\Delta}_{n,\rho}^d, \\ \bar{u}_{n,j,\beta_n}(\Delta) \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j \end{array}\right\} \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \text{ and} \right. &(180)\\
&\qquad \left.\left\{\begin{array}{c} \inf_\Delta \bar{u}_{n,u,\beta_n}(\Delta) \\ \text{s.t. } \Delta \in \mathbf{\Delta}_{n,\rho}^d, \\ \bar{u}_{n,j,\beta_n}(\Delta) \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j \end{array}\right\} \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)\right) + o(1) \\
\geq &P_n\left(\left\{\begin{array}{c} \inf_\Delta \bar{u}_{n,\ell,\beta_n}(\Delta) \\ \text{s.t. } \Delta \in \mathbf{\Delta}_{n,\rho}^d, \\ u_{n,j,\beta_n}(\Delta) \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j \end{array}\right\} \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \text{ and} \right. &(181)\\
&\qquad \left.\left\{\begin{array}{c} \inf_\Delta \bar{u}_{n,u,\beta_n}(\Delta) \\ \text{s.t. } \Delta \in \mathbf{\Delta}_{n,\rho}^d, \\ u_{n,j,\beta_n}(\Delta) \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j \end{array}\right\} \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)\right).
\end{aligned}$$

where

$$u_{n,j,\beta_n}(\Delta) = (1 + \hat{\eta}_{n,j,\beta_n}(\Delta))\left(\mathbb{Z}_{n,j,\beta_n}(\Delta) + \frac{\sigma_{P_n,j}\left(\bar{\beta}_n\right)}{\sigma_{P_n,j}\left(\beta_n\right)} D_{P_n,j}\left(\bar{\beta}_n\right)\Delta\rho + \pi_{1,j}^*\right). \tag{182}$$

Hence, I focus on the event in (181) here and after.

Step 5. I replace $\bar{u}_{n,\ell,\beta_n}$ and $\bar{u}_{n,u,\beta_n}$ with u_{n,ℓ,β_n} and u_{n,u,β_n}. First, note that if $\pi_{1,0}^* = 0$,

$$\begin{aligned}
(181) \geq &P_n\left(\left\{\begin{array}{c} \inf_\Delta u_{n,\ell,\beta_n}(\Delta) \\ \text{s.t. } \Delta \in \mathbf{\Delta}_{n,\rho}^d, \\ u_{n,j,\beta_n}(\Delta) \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j \end{array}\right\} \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \text{ and} \right. \\
&\qquad \left.\left\{\begin{array}{c} \inf_\Delta u_{n,u,\beta_n}(\Delta) \\ \text{s.t. } \Delta \in \mathbf{\Delta}_{n,\rho}^d, \\ u_{n,j,\beta_n}(\Delta) \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j \end{array}\right\} \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)\right)
\end{aligned}$$

where

$$u_{n,k,\beta_n}(\Delta) = (1+\hat{\eta}_{n,k,\beta_n}(\Delta))\left(\mathbb{Z}_{n,k,\beta_n}(\Delta) + \frac{\sigma_{P_n,k}\left(\bar{\beta}_n\right)}{\sigma_{P_n,k}\left(\beta_n\right)} D_{P_n,k}\left(\bar{\beta}_n\right)\Delta\rho\right). \tag{183}$$

If $\pi^*_{1,0} = -\infty$, we have either $\sqrt{n}\gamma_{1,P_n,\ell}(\beta_n) \to -\infty$ or $\sqrt{n}\gamma_{1,P_n,u}(\beta_n) \to -\infty$ or both, thus

$$\begin{aligned}(181) \geq \min\Bigg\{ & P_n\left(\left\{\begin{array}{c}\inf_\Delta u_{n,\ell,\beta_n}(\Delta)\\ \text{s.t. } \Delta \in \mathbf{\Delta}^d_{n,\rho},\\ u_{n,j,\beta_n}(\Delta) \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j\end{array}\right\} \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)\right),\\ & P_n\left(\left\{\begin{array}{c}\inf_\Delta u_{n,u,\beta_n}(\Delta)\\ \text{s.t. } \Delta \in \mathbf{\Delta}^d_{n,\rho},\\ u_{n,j,\beta_n}(\Delta) \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j\end{array}\right\} \leq \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)\right)\Bigg\} + o(1).\end{aligned}$$

In sum, let

$$U_{n,\ell}(\beta_n, c) = \left\{\Delta \in \mathbf{\Delta}^d_{n,\rho} : u_{n,\ell,\beta_n}(\Delta) \leq c, u_{n,u,\beta_n}(\Delta) + \pi^*_{1,0} \leq c, u_{n,j,\beta_n}(\Delta) \leq c, \forall j = 1,\dots,J\right\}, \tag{184}$$

$$U_{n,u}(\beta_n, c) = \left\{\Delta \in \mathbf{\Delta}^d_{n,\rho} : u_{n,u,\beta_n}(\Delta) \leq c, u_{n,\ell,\beta_n}(\Delta) + \pi^*_{1,0} \leq c, u_{n,j,\beta_n}(\Delta) \leq c, \forall j = 1,\dots,J\right\}. \tag{185}$$

It holds that

$$(181) \geq \min\left\{P_n\left(U_{n,\ell}\left(\beta_n, \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)\right) \neq \emptyset\right), P_n\left(U_{n,u}\left(\beta_n, \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)\right) \neq \emptyset\right)\right\}.$$

Step 6. Simplify $\hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)$. By definition $\hat{c}(\cdot) \geq 0$ and thus $\hat{c}_{n,\rho}$ defined by

$$\hat{c}_{n,\rho} = \inf_{\Delta \in \mathbf{\Delta}^d_{n,\rho}} \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right) \tag{186}$$

exists. Therefore, the event whose probability is evaluated in (181) is implied by the event

$$\left\{\begin{array}{c}\inf_\Delta u_{n,\ell,\beta_n}(\Delta)\\ \text{s.t. } \Delta \in \mathbf{\Delta}^d_{n,\rho},\\ u_{n,j,\beta_n}(\Delta) \leq \hat{c}_{n,\rho}, \forall j = 1,...,J\end{array}\right\} \leq \hat{c}_{n,\rho}, \text{ and } \left\{\begin{array}{c}\inf_\Delta u_{n,u,\beta_n}(\Delta)\\ \text{s.t. } \Delta \in \mathbf{\Delta}^d_{n,\rho},\\ u_{n,j,\beta_n}(\Delta) \leq \hat{c}_{n,\rho}, \forall j = 1,...,J\end{array}\right\} \leq \hat{c}_{n,\rho}. \tag{187}$$

Then by (187) and the definition of $U_{n,\ell}$ and $U_{n,u}$, we obtain

$$P_n(\theta_n \in CI_n) \geq \min\left\{P_n\left(U_{n,\ell}(\beta_n, \hat{c}_{n,\rho}) \neq \emptyset\right), P_n\left(U_{n,u}(\beta_n, \hat{c}_{n,\rho}) \neq \emptyset\right)\right\}. \tag{188}$$

By passing to a further subsequence, we may assume that

$$D_{P_n}(\beta_n) \to D, G_{P_n}(\beta_n) \to G,$$

for some $(J+2)\times d$ matrix D such that $\|D\| \le M$, and some $(J+2)\times K$ matrix G such that $\|G\| \le M$. In addition, we may assume that $\Omega_{P_n} \xrightarrow{u} \overline{\Omega}$ for some covariance kernel Ω. And $\Omega_{P_n}(\beta_n) \to \Omega$. By Lemma 13,

$$\liminf_{n\to\infty} \min\left\{P_n\left(U_{n,\ell}\left(\beta_n, \hat{c}_{n,\rho}\right) \neq \emptyset\right), P_n\left(U_{n,u}\left(\beta_n, \hat{c}_{n,\rho}\right) \neq \emptyset\right)\right\} \ge 1-\alpha. \tag{189}$$

The conclusion of the theorem then follows from (148), (149), (188), and (189). □

D.5 Lemmas

In the proof, I focus on

$$\liminf_{n\to\infty} P_n\left(U_{n,\ell}\left(\beta_n, \hat{c}_{n,\rho}\right) \neq \emptyset\right) \ge 1-\alpha,$$

and the proof for $P_n\left(U_{n,u}\left(\beta_n, \hat{c}_{n,\rho}\right) \neq \emptyset\right)$ is similar.

Throughout this Appendix, let (P_n, β_n, θ_n) be a subsequence as defined in Step 1 in the proof of Theorem 4. That is, along

$$\{P_n, \beta_n, \theta_n\} \in \{(P, \beta, \theta) : P \in \mathcal{P}, \beta \in \mathcal{B}(P), \theta \in [\lambda_{P,\ell}(\beta, \varphi_P), \lambda_{P,u}(\beta, \varphi_P)]\}$$

one has

$$\kappa_n^{-1}\sqrt{n}\gamma_{1,P_n,j}\left(\beta_n\right) \to \pi_{1j} \in \mathbb{R}_{[-\infty]}, j = 0, \ldots, J_1 \tag{190}$$

$$\Omega_{P_n} \xrightarrow{u} \overline{\Omega},$$

$$D_{P_n}\left(\beta_n\right) \to D, \tag{191}$$

$$G_{P_n}\left(\beta_n\right) \to G. \tag{192}$$

When Assumption 11 is invoked, I use modification in Section D.3. And if

$$\pi^*_{1,j} = 0 = \pi^*_{1,j+R_1},$$

replace the constraints

$$\mathbb{Z}_j + \rho D_j \Delta \le c,$$

$$\mathbb{Z}_{j+R_1} + \rho D_{j+R_1}\Delta \le c,$$

with

$$\mu_j(\beta)\left\{\mathbb{Z}_j + \rho D_j\Delta\right\} - \mu_{j+R_1}(\beta)\left\{\mathbb{Z}_{j+R_1} + \rho D_{j+R_1}\Delta\right\} \le c$$

$$-\mu_j(\beta)\left\{\mathbb{Z}_j + \rho D_j\Delta\right\} + \mu_{j+R_1}(\beta)\left\{\mathbb{Z}_{j+R_1} + \rho D_{j+R_1}\Delta\right\} \le c$$

where

$$\mu_j(\beta) = \begin{cases} 1 & \text{if } \gamma_{1,P_n,j}(\beta, \varphi_{P_n}) = 0 = \gamma_{1,P_n,j+R_1}(\beta, \varphi_{P_n}), \\ \frac{\gamma_{1,P_n,j+R_1}(\beta,\varphi_{P_n})}{\gamma_{1,P_n,j+R_1}(\beta,\varphi_{P_n}) + \gamma_{1,P_n,j}(\beta,\varphi_{P_n})} & \text{otherwise,} \end{cases} \tag{193}$$

$$\mu_{j+R_1}(\beta) = \begin{cases} 0 & \text{if } \gamma_{1,P_n,j}(\beta) = 0 = \gamma_{1,P_n,j+R_1}(\beta), \\ \frac{\gamma_{1,P_n,j}(\beta,\varphi_{P_n})}{\gamma_{1,P_n,j+R_1}(\beta,\varphi_{P_n})+\gamma_{1,P_n,j}(\beta,\varphi_{P_n})} & \text{otherwise,} \end{cases} . \tag{194}$$

The same applies to the constraint on $\mathbb{Z}_\ell$ and $\mathbb{Z}_u$.

To simplify notation, in the following proof we do not differentiate ℓ, u, and j. Moreover, due to the substitutions in equations (172) and (174), the paired inequalities are now genuine equalities. With some absue of notation, we index them among the $j = J_1 + 1, ..., J$. That is, under Assumption 9,

$$\begin{aligned} \mathbb{J}_1 &= \{1, ..., J_1, \ell, u\}, \\ \mathbb{J}_2 &= \{J_1 + 1, ..., J_1 + 2J_2\}. \end{aligned}$$

Under Assumption 11.1,

$$\begin{aligned} \mathbb{J}_1 &= \{2R_1 + 1, ..., J_1, \ell, u\}, \\ \mathbb{J}_2 &= \{1, ..., R_1, J_1 + 1, ..., J_1 + J_2\}. \end{aligned}$$

Under Assumption 11.2,

$$\begin{aligned} \mathbb{J}_1 &= \{2R_1 + 1, ..., J_1\}, \\ \mathbb{J}_2 &= \{1, ..., R_1, J_1 + 1, ..., J_1 + J_2, \ell\}. \end{aligned}$$

In all three cases,

$$\begin{aligned} [\mathbb{J}_2] &= \{[j] : j \in \mathbb{J}_2\}, \\ \mathbb{J} &= \mathbb{J}_1 \cup \mathbb{J}_2 \cup [\mathbb{J}_2]. \end{aligned}$$

Fix $c \geq 0$. For each $\Delta \in \mathbb{R}^d$ and $\beta \in (\beta_n + \rho/\sqrt{n}\boldsymbol{\Delta}) \cap \mathcal{B}$, let

$$\begin{aligned} \mathfrak{w}_j(\Delta) &= \mathbb{Z}_j + \rho D_j \Delta + \pi^*_{1,j}, \\ \mathfrak{w}_\ell(\Delta) &= \mathbb{Z}_\ell + \rho D_\ell \Delta, \\ \mathfrak{w}_u(\Delta) &= \mathbb{Z}_u + \rho D_u \Delta + \pi^*_{1,0}. \end{aligned} \tag{195}$$

Let $\boldsymbol{\Delta}^d_{\infty,\rho} = \lim_{n\to\infty} \boldsymbol{\Delta}^d_{n,\rho}$. Let

$$\mathfrak{W}(c) = \left\{ \Delta \in \boldsymbol{\Delta}^d_{\infty,\rho} : \mathfrak{w}_j(\Delta) \leq c, \forall j \in \mathbb{J} \right\}. \tag{196}$$

With that convention, for given $\delta \in \mathbb{R}$, define

$$\mathfrak{W}^\delta(c) = \left\{ \Delta \in \boldsymbol{\Delta}^d_{\infty,\rho} : \mathfrak{w}_j(\Delta) \leq c + \delta, \forall j \in \mathbb{J}_1, \mathfrak{w}_j(\Delta) \leq c, \forall j \in \mathbb{J}_2 \cup [\bar{\mathbb{J}}_2] \right\}. \tag{197}$$

Define the $|\mathbb{J}| + 2d$ matrix

$$K_P(\beta,\rho) = \begin{bmatrix} [\rho D_{P,j}(\beta)]_{j\in\mathbb{J}_1\cup\mathbb{J}_2} \\ [-\rho D_{P,j}(\beta)]_{j\in\bar{\mathbb{J}}_2} \\ I_d \\ -I_d \end{bmatrix}.$$

Given a square matrix A, we let $\text{eig}(A)$ denote its smallest eigenvalue. In all lemmas below, we assume $\alpha < 1/2$.

Lemma 13. *Suppose Assumptions 7, 8, 6, 10 hold. In addition, suppose Assumption 9 or 11 hold. Then,*

$$\liminf_{n\to\infty} P_n\left(U_{\ell,n}\left(\beta_n, \hat{c}_{n,\rho}\right) \neq \emptyset\right) \geq 1-\alpha. \tag{198}$$

Proof. Consider a subsequence along which

$$\liminf_{n\to\infty} P_n\left(U_{\ell,n}\left(\beta_n, \hat{c}_{n,\rho}\right) \neq \emptyset\right)$$

is achieved as a limit. For notational simplicity, we use $\{n\}$ for this subsequence below. Below, we construct a sequence of critical values $c_n^I\left(\beta_n'\right)$ such that

$$\hat{c}_n\left(\beta_n'\right) \geq c_n^I\left(\beta_n'\right) + o_p(1) \tag{199}$$

and $c_n^I\left(\beta_n'\right) \xrightarrow{p} c_{\pi^*}$ for any $\beta_n' \in \left(\beta_n + \rho/\sqrt{n}\boldsymbol{\Delta}\right) \cap \overline{\mathcal{B}}$, where

$$c_{\pi^*} = \inf\left\{c \in \mathbb{R}_+ : P\left(\mathfrak{W}(c) \neq \emptyset\right) \geq 1-\alpha\right\}. \tag{200}$$

The construction is as follows. When $c_{\pi^*} = 0$, let $c_n^I\left(\beta_n'\right) = 0$ for all $\beta_n' \in \left(\beta_n + \rho/\sqrt{n}\boldsymbol{\Delta}\right) \cap \overline{\mathcal{B}}$, and hence $c_n^I\left(\beta_n'\right) \xrightarrow{p} c_{\pi^*}$. If $c_{\pi^*} > 0$, let

$$c_n^I\left(\beta_n\right) = \inf\left\{c \in \mathbb{R}_+ : P^s\left(V_n^I\left(\beta_n, c\right) \neq \emptyset\right) \geq 1-\alpha\right\}, \tag{201}$$

where

$$V_n^I\left(\beta_n', c\right) = \left\{\Delta \in \boldsymbol{\Delta}_{n,\rho}^d : v_{n,j,\beta_n'}^I(\Delta) \leq c, j \in \mathbb{J}\right\}, \tag{202}$$

$$v_{n,j,\beta_n'}^I(\Delta) = \mathbb{Z}_{n,j}^s\left(\beta_n'\right) + \rho\hat{D}_{n,j}\left(\beta_n'\right)\Delta + \phi_j^*\left(\hat{\xi}_{n,j}\left(\beta_n'\right)\right), \tag{203}$$

$$v_{n,\ell,\beta_n'}^I(\Delta) = \mathbb{Z}_{n,\ell}^s\left(\beta_n\right) + \rho\hat{D}_{n,\ell}\left(\beta_n'\right)\Delta$$

$$v_{n,u,\beta_n'}^I(\Delta) = \mathbb{Z}_{n,u}^s\left(\beta_n\right) + \rho\hat{D}_{n,u}\left(\beta_n'\right)\Delta + \phi_0^*\left(\hat{\xi}_{n,0}\left(\beta_n'\right)\right)$$

$$\phi_j^*(\xi) = \begin{cases} \xi & \pi_{1,j} = 0 \\ -\infty & \pi_{1,j} < 0 \end{cases}, \qquad \phi_0^*(\xi) = \begin{cases} \xi & \pi_{1,0} = 0 \\ -\infty & \pi_{1,0} < 0 \end{cases}. \tag{204}$$

and P^s is from the conditional distribution of $\mathbb{Z}_n^s$ conditional on the estimators. By Lemma 15.3, this critical value sequence satisfies (199). Further, by Lemma 15.2, $c_n^I\left(\beta_n'\right) \xrightarrow{p} c_{\pi^*}$ for any

$\beta_n' \in (\beta_n + \rho/\sqrt{n}\mathbf{\Delta}) \cap \overline{\mathcal{B}}$. For each $\beta \in \overline{\mathcal{B}}$, let

$$c_{n,\rho}^I(\beta) = \inf_{\Delta \in \mathbf{\Delta}_{n,\rho}^d} c_n^I\left(\beta + \frac{\Delta\rho}{\sqrt{n}}\right). \tag{205}$$

Since the $o_p(1)$ term in (199) does not affect the argument below, I redefine $c_{n,\rho}^I(\beta_n)$ as $c_{n,\rho}^I(\beta_n) + o_p(1)$. By (199) and simple addition and subtraction,

$$\begin{aligned}
&P_n\left(U_{n,\ell}\left(\beta_n, \hat{c}_{n,\rho}\left(\beta_n\right)\right) \neq \emptyset\right) \\
\geq &P_n\left(U_{n,\ell}\left(\beta_n, c_{n,\rho}^I\left(\beta_n\right)\right) \neq \emptyset\right) \\
= &P\left(\mathfrak{W}\left(c_{\pi^*}\right) \neq \emptyset\right) + \left[P_n\left(U_{n,\ell}\left(\beta_n, c_{n,\rho}^I\left(\beta_n\right)\right) \neq \emptyset\right) - P\left(\mathfrak{W}\left(c_{\pi^*}\right) \neq \emptyset\right)\right].
\end{aligned} \tag{206}$$

By Lemma 22,

$$\mathbb{Z}_{n,j,\beta_n}(\Delta) \xrightarrow{d} \mathbb{Z} \sim \mathcal{N}(0, \Sigma).$$

And by Assumption 9 and Assumption 12.3,

$$\sup_{\beta \in \mathcal{B}} \sup_{\Delta \in \mathbf{\Delta}} \frac{\sigma_{P_n,k}\left(\bar{\beta}_n\right)}{\sigma_{P_n,k}\left(\beta_n\right)} \to 1$$

Moreover, by Lemma 21,

$$\sup_{\beta \in \mathcal{B}} \sup_{\Delta \in \mathbf{\Delta}} \|\hat{\eta}_{n,j,\beta}(\Delta)\| \xrightarrow{p} 0$$

uniformly in $\mathcal{P}$, and by Lemma 15, $c_{n,\rho}^I(\beta_n) \xrightarrow{p} c_{\pi^*}$. Therefore, uniformly in $\Delta \in \mathbf{\Delta}$, it holds that

$$\left(\mathbb{Z}_{n,\beta_n}(\Delta),\ \left\{\hat{\eta}_{n,\beta}(\Delta),\ c_{n,\rho}^I\left(\beta_n\right)\right)\right. \xrightarrow{d} (\mathbb{Z},\ 0,\ c_{\pi^*}). \tag{207}$$

In what follows, using Lemma 1.10.4 in Van Der Vaart and Wellner (1996) I take

$$\left(\mathbb{Z}_n^*(\Delta), \eta_n^*(\Delta), \left\{D_{n,j}^*\right\}_{j \leq J}, c_n^*\right)$$

to be the almost sure representation of the left hand side of (207), defined on some probability space $(\Omega, \mathcal{F}, P)$ such that $(\mathbb{Z}_n^*(\Delta), \eta_n^*(\Delta), c_n^*) \xrightarrow{a.s.} (\mathbb{Z}^*, 0, c_{\pi^*})$, where $\mathbb{Z}^* \stackrel{d}{=} \mathbb{Z}$. For each $\Delta \in \mathbb{R}^d$, we define

$$\begin{aligned}
u_{n,j,\beta_n}^*(\Delta) &= (1 + \eta_n^*(\Delta))\left\{\mathbb{Z}_{n,j}^*(\Delta) + \rho\frac{\sigma_{P_n,j}\left(\bar{\beta}_n\right)}{\sigma_{P_n,j}\left(\beta_n\right)} D_{P_n,j}\left(\bar{\beta}_n\right)\Delta + \pi_{1,j}^*\right\}, \\
\mathfrak{w}_j^*(\Delta) &= \mathbb{Z}_j^* + \rho D_j \Delta + \pi_{1,j}^*,
\end{aligned}$$

where we used that by Lemma 23.1, $\kappa_n^{-1}\sqrt{n}\gamma_{1,P,j}(\beta_n) - \kappa_n^{-1}\sqrt{n}\gamma_{1,P,j}(\beta_n') = o(1)$ uniformly over $\beta_n' \in (\beta_n + \rho/\sqrt{n}\mathbf{\Delta}) \cap \mathcal{B}$ and therefore $\pi_{1,j}^*$ is constant over this neighborhood. Similarly, let

$$U_n^*\left(\beta_n, c_n^*\right) = \left\{\Delta \in \mathbf{\Delta}_{n,\rho}^d : u_{n,j,\beta_n}^*(\Delta) \leq c_n^*,\ \forall j \in \mathbb{J}\right\}, \tag{208}$$

$$\mathfrak{W}^*\left(c_{\pi^*}\right) = \left\{\Delta \in \mathbf{\Delta}_{\infty,\rho}^d : \mathfrak{w}_j^*(\Delta) \leq c_{\pi^*},\ \forall j \in \mathbb{J}\right\}. \tag{209}$$

It then follows that equation (206) can be rewritten as

$$
\begin{aligned}
&P_n\left(\left\{U_n\left(\beta_n,\hat{c}_{n,\rho}\left(\beta_n\right)\right)\neq\emptyset\right\}\right)\\
\geq &P\left(\mathfrak{W}^*\left(c_{\pi^*}\right)\neq\emptyset\right)+\left[P_n\left(U_n^*\left(\beta_n,c_n^*\right)\neq\emptyset\right)-P\left(\mathfrak{W}^*\left(c_{\pi^*}\right)\neq\emptyset\right)\right].
\end{aligned}
\tag{210}
$$

By the definition of c_{π^*}, we have

$$
P\left(\mathfrak{W}^*\left(c_{\pi^*}\right)\neq\emptyset\right)\geq 1-\alpha.
$$

Therefore, we are left to show that the second term on the right hand side of (210) tends to 0 as $n\to\infty$. Note that

$$
\begin{aligned}
&\left|P\left(U_n^*\left(\beta_n,c_n^*\right)\neq\emptyset\right)-P\left(\mathfrak{W}^*\left(c_{\pi^*}\right)\neq\emptyset\right)\right|\\
\leq &P_n\left(\left\{U_n^*(\beta_n,c_n^*)=\emptyset\right\}\cap\left\{\mathfrak{W}^*(c_{\pi^*})\neq\emptyset\right\}\right)+P_n\left(\left\{U_n^*(\beta_n,c_n^*)\neq\emptyset\right\}\cap\left\{\mathfrak{W}^*(c_{\pi^*})=\emptyset\right\}\right).
\end{aligned}
\tag{211}
$$

The conclusion holds because (211) converges to zero by Lemma 14. □

Lemma 14. *Suppose Assumptions 7, 8, 6, 10 hold. In addition, suppose Assumption 9 or 11 hold. Let (P_n,β_n,θ_n) have the almost sure representations given in Lemma 13. For any $\eta>0$, there exists $N\in\mathbb{N}$ such that*

$$
P_n\left(\left\{U_n^*\left(\beta_n,c_n^*\right)\neq\emptyset\right\}\cap\left\{\mathfrak{W}^*\left(c_{\pi^*}\right)=\emptyset\right\}\right)\leq\eta/2 \tag{212}
$$
$$
P_n\left(\left\{U_n^*\left(\beta_n,c_n^*\right)=\emptyset\right\}\cap\left\{\mathfrak{W}^*\left(c_{\pi^*}\right)\neq\emptyset\right\}\right)\leq\eta/2, \tag{213}
$$

for all $n\geq N$, where the sets in the above expressions are defined in equations (208) and (209).

Proof. Let $\mathbb{J}^*=\left\{j\in\mathbb{J}:\pi_{1,j}^*=0\right\}$. Observe that for $j=1,...,\mathbb{J}$, if $\pi_{1,j}^*=-\infty$, the corresponding inequalities

$$
\begin{aligned}
u_{n,j,\beta_n}^*(\Delta)&=(1+\eta_n^*(\Delta))\left\{\mathbb{Z}_{n,j}^*(\Delta)+\rho\frac{\sigma_{P_n,j}\left(\bar{\beta}_n\right)}{\sigma_{P_n,j}\left(\beta_n\right)}D_{P_n,j}\left(\bar{\beta}_n\right)\Delta+\pi_{1,j}^*\right\}\leq c_n^*\\
\mathfrak{w}_j^*(\Delta)&=\mathbb{Z}_j^*+\rho D_j\Delta+\pi_{1,j}^*\leq c_{\pi^*}
\end{aligned}
$$

are satisfied with probability approaching one by similar arguments as in (179). Hence, we can redefine the sets of interest as

$$
U_n^*\left(\beta_n,c_n^*\right)=\left\{\Delta\in\mathbf{\Delta}_{n,\rho}^d:u_{n,j,\beta_n}^*(\Delta)\leq c_n^*,\forall j\in\mathbb{J}^*\right\} \tag{214}
$$
$$
\mathfrak{W}^*\left(c_{\pi^*}\right)=\left\{\Delta\in\mathbf{\Delta}_{\infty,\rho}^d:\mathfrak{w}_j^*(\Delta)\leq c_{\pi^*},\forall j\in\mathbb{J}^*\right\}. \tag{215}
$$

I first show (212). I bound the left hand side of (212) as

$$
\begin{aligned}
&P\left(\left\{U_n^*\left(\beta_n,c_n^*\right)\neq\emptyset\right\}\cap\left\{\mathfrak{W}^*\left(c_{\pi^*}\right)=\emptyset\right\}\right)\\
\leq &P\left(\left\{U_n^*\left(\beta_n,c_n^*\right)\neq\emptyset\right\}\cap\left\{\mathfrak{W}^{*,+\delta}\left(c_{\pi^*}\right)=\emptyset\right\}\right) \qquad (216)\\
&+P\left(\left\{\mathfrak{W}^{*,+\delta}\left(c_{\pi^*}\right)\neq\emptyset\right\}\cap\left\{\mathfrak{W}^*\left(c_{\pi^*}\right)=\emptyset\right\}\right), \qquad (217)
\end{aligned}
$$

following from $P(A\cap B)\leq P(A\cap C)+P(B\cap C^c)$ for any events A,B, and C. I will then specify δ such that (216)$\leq\frac{\eta}{2}$ and (217)$\leq\frac{\eta}{2}$. Lemma 17 (259) implies that there is some $\delta>0$ such that

$$P\left(\left\{\mathfrak{W}^{*,+\delta}\left(c_{\pi^*}\right)\neq\emptyset\right\}\cap\left\{\mathfrak{W}^{*}\left(c_{\pi^*}\right)=\emptyset\right\}\right)$$
$$\leq\sup_{c\geq 0}P\left(\left\{\mathfrak{W}^{*}\left(c\right)\neq\emptyset\right\}\cap\left\{\mathfrak{W}^{*,-\delta}\left(c\right)=\emptyset\right\}\right)\leq\eta/2. \tag{218}$$

Next, define the events

$$A_n(\delta)=\left\{\sup_{\Delta\in\boldsymbol{\Delta}}\max_{j\in\mathbb{J}^*}\left|\left(u^*_{n,j,\beta_n}(\Delta)-c^*_n\right)-\left(\mathfrak{w}^*_j(\Delta)-c_{\pi^*}\right)\right|\geq\delta\right\},$$

with δ specified in (218). By Lemma 16 (257), for any $\eta>0$ there exists $N\in\mathbb{N}$ such that

$$P\left(A_n\right)<\eta/2,\forall n\geq N. \tag{219}$$

Define the event $L_n=\{U^*_n\left(\beta_n,c^*_n\right)\subseteq\mathfrak{W}^*\left(c_{\pi^*}+\delta\right)\}$ and note that $A^c_n\subseteq L_n$. The right hand side of (216) can further be bounded as

$$P\left(\left\{U^*_n\left(\beta_n,c^*_n\right)\neq\emptyset\right\}\cap\left\{\mathfrak{W}^*\left(c_{\pi^*}+\delta\right)=\emptyset\right\}\right)\leq P\left(U^*_n\left(\beta_n,c^*_n\right)\nsubseteq\mathfrak{W}^*\left(c_{\pi^*}+\delta\right)\right)$$
$$=P\left(L^c_n\right)\leq P\left(A_n\right)<\eta/2\qquad\forall n\geq N, \tag{220}$$

where the penultimate inequality follows from $A^c_n\subseteq L_n$ as argued above, and the last inequality follows from (219). Hence, (212) follows from (216), (217), (218), and (220).

To establish (213), I distinguish three cases.

Case 1. Suppose first that $\mathbb{J}_2=\emptyset$, and hence one has only moment inequalities. Define the event

$$\widetilde{R}_{2n}=\left\{\mathfrak{W}^*\left(c_{\pi^*}-\delta\right)\subseteq U^*_n\left(\beta_n,c^*_n\right)\right\}. \tag{221}$$

and note that $A^c_n\subseteq\widetilde{R}_{2n}$. The result in equation (213) then follows by Lemma 17 (259) using again similar steps to (216)-(220).

Case 2. Next suppose that $|\mathbb{J}_2|>d$. Note that, by Lemma 19, there exists $N\in\mathbb{N}$ such that for all $n\geq N$, $c^I_n(\theta)$ is bounded from below by some $\underline{c}>0$ with probability approaching one uniformly in $P\in\mathcal{P}$ and $\beta\in\mathcal{B}(P)$. This ensures c_{π^*} is bounded from below by $\underline{c}>0$. Note that $A^c_n\subseteq\widetilde{R}_{2n}$, where $\widetilde{R}_{2n}$ is defined as in (221), and therefore the same argument as in the previous case applies using Lemma 17(261).

Case 3. Finally, suppose that $1\leq|\mathbb{J}_2|\leq d$. Recall that, with probability 1,

$$c_{\pi^*}=\lim_{n\to\infty}c^*_n, \tag{222}$$

and note that by construction $c_{\pi^*}\geq 0$. Consider first the case that $c_{\pi^*}>0$. Then, by taking $\delta<c_{\pi^*}$, the argument in Case 2 applies.

Next, consider the case that $c_{\pi^*} = 0$. Observe that

$$
\begin{aligned}
&P\left(\left\{U_n^*\left(\beta_n, c_n^*\right)=\emptyset\right\} \cap\left\{\mathfrak{W}^*\left(c_{\pi^*}\right) \neq \emptyset\right\}\right), \\
\leq & P\left(\left\{U_n^*\left(\beta_n, c_n^*\right)=\emptyset\right\} \cap\left\{\mathfrak{W}^{*,-\delta}(0) \neq \emptyset\right\}\right)+P\left(\left\{\mathfrak{W}^{*,-\delta}(0)=\emptyset\right\} \cap\left\{\mathfrak{W}^*(0) \neq \emptyset\right\}\right) .
\end{aligned} \tag{223}
$$

By Lemma 17, for any $\eta > 0$ there exists $\delta > 0$ and $N \in \mathbb{N}$ such that

$$
P\left(\left\{\mathfrak{W}^{*,-\delta}(0)=\emptyset\right\} \cap\left\{\mathfrak{W}^*(0) \neq \emptyset\right\}\right)<\eta / 3
$$

for all $n \geq N$. Therefore, the second term of (223) can be made arbitrarily small.

I now consider the first term of (223), and it contains the following four steps.

Step 1. More notation. Let g be a $|\mathbb{J}| + 2d$ vector with

$$
g_j=\begin{cases}-\mathbb{Z}_j, & j \in \mathbb{J}, \\ 1, & j=|\mathbb{J}|+1, \ldots,|\mathbb{J}|+2 d,\end{cases} \tag{224}
$$

where I use that $\pi_{1,j}^* = 0$ and the last assignment is without loss of generality because of the considerations leading to the sets in (214), (215). For a given set $C \subset \{1, \ldots, |\mathbb{J}| + 2d\}$, let the vector g^C collect the entries of g^C corresponding to indices in C. Let

$$
K=\left[\begin{array}{c}{\left[\rho D_j\right]_{j \in \mathbb{J}_1 \cup \mathbb{J}_2}} \\ {\left[-\rho D_j\right]_{j \in \mathbb{J}_2}} \\ I_d \\ -I_d\end{array}\right] . \tag{225}
$$

Let the matrix K^C collect the rows of K corresponding to indices in C. Let $\widetilde{\mathcal{C}}$ collect all size d subsets C of $\{1, \ldots, |\mathbb{J}| + 2d\}$ ordered lexicographically by their smallest, then second smallest, etc. elements. Let the random variable $\mathcal{C}$ equal the first element of $\widetilde{\mathcal{C}}$ s.t. $\det K^C \neq 0$ and $\Delta^C = \left(K^C\right)^{-1} g^C \in \mathfrak{W}^{*,-\delta}(0)$ if such an element exists; else, let $\mathcal{C} = \{|\mathbb{J}| + 1, \ldots, |\mathbb{J}| + d\}$ and $\Delta^C = 1_d$, where 1_d denotes a d vector with each entry equal to 1. Recall that $\mathfrak{W}^{*,-\delta}(0)$ is a (possibly empty) measurable random polyhedron in a compact subset of $\mathbb{R}^d$, see, e.g., Molchanov and Molchanov (2005) (Definition 1.1.1). Thus, if $\mathfrak{W}^{*,-\delta}(0) \neq \emptyset$, then $\mathfrak{W}^{*,-\delta}(0)$ has extreme points, each of which is characterized as the intersection of d (not necessarily unique) linearly independent constraints interpreted as equalities. Therefore, $\mathfrak{W}^{*,-\delta}(0) \neq \emptyset$ implies that $\Delta^{\mathcal{C}} \in \mathfrak{W}^{*,-\delta}(0)$. Note that the associated random vector $\Delta^{\mathcal{C}}$ is a measurable selection of a random closed set that equals $\mathfrak{W}^{*,-\delta}(0)$ if $\mathfrak{W}^{*,-\delta}(0) \neq \emptyset$ and equals $\boldsymbol{\Delta}$ otherwise, see, e.g., Molchanov and Molchanov (2005) (Definition 1.2.2).

Let g_n be a $|\mathbb{J}| + 2d$ vector with

$$
g_{n,j}(\Delta)=\begin{cases}c_n^* /\left(1+\eta_{n,j}^*(\Delta)\right)-\mathbb{Z}_{n,j}^*(\Delta) & \text{if } j \in \mathbb{J} \\ 1, & \text{if } j=|\mathbb{J}|+1, \ldots,|\mathbb{J}|+2 d\end{cases} \tag{226}
$$

using again that $\pi^*_{1,j} = 0$ for $j \in \mathbb{J}^*$. For each $P \in \mathcal{P}$, let

$$K_P(\beta, \bar{\beta}, \rho) = \begin{bmatrix} \left[\rho \frac{\sigma_{P_n,j}(\bar{\beta})}{\sigma_{P_n,j}(\beta)} D_{P,j}(\bar{\beta})\right]_{j \in \mathbb{J}_1 \cup \mathbb{J}_2} \\ \left[-\rho \frac{\sigma_{P_n,j}(\bar{\beta})}{\sigma_{P_n,j}(\beta)} D_{P,j}(\bar{\beta})\right]_{j \in \mathbb{J}_2} \\ I_d \\ -I_d \end{bmatrix}. \tag{227}$$

For each n and $\Delta \in \boldsymbol{\Delta}$, define the mapping $\phi_n : \boldsymbol{\Delta} \to \mathbb{R}^d_{[\pm\infty]}$ by

$$\phi_n(\Delta) = \left(K^{\mathcal{C}}_{P_n}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta\right), \rho\right)\right)^{-1} g^{\mathcal{C}}_n\left(\Delta\right), \tag{228}$$

where the notation $\bar{\beta}\left(\beta_n, \Delta\right)$ emphasizes that $\bar{\beta}$ depends on β_n and Δ because it lies component-wise between β_n and $\beta_n + \frac{\Delta\rho}{\sqrt{n}}$.

Step 2. I show that ϕ_n is a contraction mapping and hence has a fixed point. For any $\Delta, \Delta' \in \boldsymbol{\Delta}$ write

$$\begin{aligned} &\left\|\phi_n(\Delta) - \phi_n\left(\Delta'\right)\right\| \\ =&\left\|\left(K^{\mathcal{C}}_{P_n}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta\right), \rho\right)\right)^{-1} g^{\mathcal{C}}_n\left(\Delta\right) - \left(K^{\mathcal{C}}_{P_n}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta'\right), \rho\right)\right)^{-1} g^{\mathcal{C}}_n\left(\Delta'\right)\right\| \\ \leq&\left\|\left(K^{\mathcal{C}}_{P_n}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta\right), \rho\right)\right)^{-1}\right\|_2 \left\|g^{\mathcal{C}}_n\left(\Delta\right) - g^{\mathcal{C}}_n\left(\Delta'\right)\right\| \\ &+ \left\|\left(K^{\mathcal{C}}_{P_n}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta\right), \rho\right)\right)^{-1} - \left(K^{\mathcal{C}}_{P_n}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta'\right), \rho\right)\right)^{-1}\right\|_2 \left\|g^{\mathcal{C}}_n\left(\Delta'\right)\right\|, \end{aligned} \tag{229}$$

where $\|\cdot\|_2$ denotes the spectral norm (induced by the Euclidean norm). By Assumption 12.2, for any $\eta > 0, k > 0$, there is $N \in \mathbb{N}$ such that

$$\begin{aligned} &P_n\left(\left\|g^{\mathcal{C}}_n(\Delta) - g^{\mathcal{C}}_n(\Delta')\right\| \leq k\left\|\Delta - \Delta'\right\|\right) \\ =&P_n\left(\left\|\mathbb{Z}^{*;\mathcal{C}}_n(\Delta) - \mathbb{Z}^{*;\mathcal{C}}_n(\Delta')\right\| \leq k\left\|\Delta - \Delta'\right\|\right) \geq 1 - \eta, \forall n \geq N. \end{aligned}$$

Moreover, for any η there exist $0 < L < \infty$ and $N \in \mathbb{N}$ such that $\forall n \geq N$

$$\mathrm{P}\left(\sup_{\Delta' \in \boldsymbol{\Delta}} \left\|g^{\mathcal{C}}_n\left(\Delta'\right)\right\| \leq L\right) \geq 1 - \eta. \tag{230}$$

For any invertible matrix K, $\left\|K^{-1}\right\|_2 = \left(\min\left\{\sqrt{\alpha} : \alpha \text{ is an eigenvalue of } KK'\right\}\right)^{-1}$. Hence, by the proof of Lemma 18 and the definition of $\mathcal{C}$, for any $\eta > 0$, there exists $0 < L < \infty$ and $N \in \mathbb{N}$ such that

$$\mathrm{P}\left(\left\|\left(K^{\mathcal{C}}\right)^{-1}\right\|_2 \leq L\right) \geq 1 - \eta, \forall n \geq N.$$

By Johnson and Horn (1985) (ch. 5.8), for any invertible matrices $K, \tilde{K}$ such that $\left\|\tilde{K}^{-1}(K - \tilde{K})\right\|_2 < 1$,

$$\left\|K^{-1} - \tilde{K}^{-1}\right\|_2 \leq \frac{\left\|\tilde{K}^{-1}(K - \tilde{K})\right\|_2}{1 - \left\|\tilde{K}^{-1}(K - \tilde{K})\right\|_2} \left\|\tilde{K}^{-1}\right\|_2. \tag{231}$$

By the assumption that $D_{P_n}(\beta_n) \to D$ and $\frac{\sigma_{P_n,j}(\bar{\beta})}{\sigma_{P_n,j}(\beta)} \to 1$ and Assumption 7, for any $\eta > 0$, there exists $N \in \mathbb{N}$ such that

$$\sup_{\Delta \in \mathbf{\Delta}} \left\| K^{\mathcal{C}}_{P_n}\left(\bar{\beta}\left(\beta_n, \Delta\right), \rho\right) - K^{\mathcal{C}} \right\|_2 \leq \eta, \forall n \geq N.$$

By (231), the definition of the spectral norm, and the triangle inequality, for any $\eta > 0$, there exist $0 < L_1, L_2 < \infty$ and $N \in \mathbb{N}$ such that

$$\begin{aligned}
&P_n\left(\sup_{\Delta \in \mathbf{\Delta}} \left\| \left(K^{\mathcal{C}}_{P_n}\left(\bar{\beta}\left(\beta_n, \Delta\right), \rho\right)\right)^{-1} \right\|_2 \leq 2L_1\right) \\
\geq &P_n\left(\left\| \left(K^{\mathcal{C}}\right)^{-1} \right\|_2 + \sup_{\Delta \in \mathbf{\Delta}} \left\| K^{\mathcal{C}}_{P_n}\left(\bar{\beta}\left(\beta_n, \Delta\right), \rho\right)^{-1} - \left(K^{\mathcal{C}}\right)^{-1} \right\|_2 \leq 2L_1\right) \\
\geq &P_n\left(\left\| \left(K^{\mathcal{C}}\right)^{-1} \right\|_2 \leq L_1, \frac{\left\| \left(K^{\mathcal{C}}\right)^{-1} \right\|_2^2}{1 - \left\| \left(K^{\mathcal{C}}\right)^{-1}\left(K^{\mathcal{C}}_{P_n}\left(\bar{\beta}\left(\beta_n, \Delta\right), \rho\right) - K^{\mathcal{C}}\right) \right\|_2} \leq L_2, \right. \\
&\qquad \left. \sup_{\Delta \in \mathbf{\Delta}} \left\| K^{\mathcal{C}}_{P_n}\left(\bar{\beta}\left(\beta_n, \Delta\right), \rho\right) - K^{\mathcal{C}} \right\|_2 \leq \frac{L_1}{L_2}\right) \\
\geq &1 - 2\eta, \forall n \geq N,
\end{aligned} \tag{232}$$

Again by applying (231), for any $k > 0$, there exists $N \in \mathbb{N}$ such that

$$\begin{aligned}
&P_n\left(\left\| \left(K^{\mathcal{C}}_{P_n}\left(\bar{\beta}\left(\beta_n, \Delta\right)\right)\right)^{-1} - \left(K^{\mathcal{C}}_{P_n}\left(\bar{\beta}\left(\beta_n, \Delta'\right)\right)\right)^{-1} \right\|_2 \leq k \left\| \Delta - \Delta' \right\|\right) \\
\geq &P_n\left(\sup_{\Delta \in \mathbf{\Delta}} \left\| \left(K^{\mathcal{C}}_{P_n}\left(\bar{\beta}\left(\beta_n, \Delta\right)\right)\right)^{-1} \right\|_2^2 \frac{M\rho^2}{\sqrt{n}} \left\| \Delta - \Delta' \right\| \leq k \left\| \Delta - \Delta' \right\|\right) \geq 1 - \eta, \quad \forall n \geq N,
\end{aligned} \tag{233}$$

where the first inequality follows from

$$\left\| K^{\mathcal{C}}_{P_n}\left(\bar{\beta}\left(\beta_n, \Delta\right)\right) - K^{\mathcal{C}}_{P_n}\left(\bar{\beta}\left(\beta_n, \Delta'\right)\right) \right\|_2 \leq M\rho \left\| \bar{\beta}\left(\beta_n, \Delta\right) - \bar{\beta}\left(\beta_n, \Delta'\right) \right\| \leq M\rho^2/\sqrt{n} \left\| \Delta - \Delta' \right\|$$

by Assumption 7, and the last inequality follows from (232).

By (229)-(230) and (232)-(233), it then follows that there exists $a \in [0, 1)$ such that for any $\eta > 0$, there exists $N \in \mathbb{N}$ such that

$$\mathrm{P}\left(\left|\phi_n(\Delta) - \phi_n\left(\Delta'\right)\right| \leq a \left\| \Delta - \Delta' \right\|, \quad \forall \Delta, \Delta' \in \mathbf{\Delta}\right) \geq 1 - \eta, \quad \forall n \geq N. \tag{234}$$

This implies that with probability approaching 1, each $\phi_n(\cdot)$ is a contraction, and therefore by the Contraction Mapping Theorem it has a fixed point (e.g., Pata and others (2019) (Theorem 1.1)). This in turn implies that for any $\eta > 0$ there exists a $N \in \mathbb{N}$ such that

$$P\left(\exists \Delta^f_n : \Delta^f_n = \phi_n\left(\Delta^f_n\right)\right) \geq 1 - \eta, \forall n \geq N. \tag{235}$$

Step 3. I show that $\Delta^{\mathcal{C}}_n$ and Δ^f_n are close enough. Next, define the mapping

$$\psi_n(\Delta) = \left(K^{\mathcal{C}}\right)^{-1} g^{\mathcal{C}} \tag{236}$$

This map is constant in Δ and hence is uniformly continuous and a contraction with Lipschitz constant equal to zero. It therefore has $\Delta_n^{\mathcal{C}}$ as its fixed point. Moreover, by (228) and (236) arguing as in (229), it follows that for any $\Delta \in \mathbf{\Delta}$,

$$\|\psi_n(\Delta) - \phi_n(\Delta)\| \leq \left\| \left(K_{P_n}^{\mathcal{C}}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta\right), \rho\right)\right)^{-1} \right\|_2 \left\| g^{\mathcal{C}} - g_n^{\mathcal{C}}\left(\Delta\right) \right\| + \left\| \left(K^{\mathcal{C}}\right)^{-1} - \left(K_{P_n}^{\mathcal{C}}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta\right), \rho\right)\right)^{-1} \right\|_2 \left\| g^{\mathcal{C}} \right\|.$$

By (224) and (226)

$$\begin{aligned} \left\| g^{\mathcal{C}} - g_n^{\mathcal{C}}(\Delta) \right\| &\leq \max_{j \in \mathbb{J}^*} \left| -\mathbb{Z}_j^* - c_n^* / \left(1 + \eta_{n,j}^*(\Delta)\right) + \mathbb{Z}_{n,j}^*(\Delta) \right| \\ &\leq \max_{j \in \mathbb{J}^*} \left| \mathbb{Z}_j^* - \mathbb{Z}_{n,j}^*\left(\Delta\right) \right| + \max_{j \in \mathbb{J}^*} \left| c_n^* / \left(1 + \eta_{n,j}^*(\Delta)\right) \right|. \end{aligned} \tag{237}$$

Moreover, $\mathbb{Z}_n^*\left(\Delta\right) \overset{a.s.}{\to} \mathbb{Z}^*$ and (222) implies $c_n^* \to 0$ so that we have

$$\sup_{\Delta \in \mathbf{\Delta}} \left\| g^{\mathcal{C}} - g_n^{\mathcal{C}}\left(\Delta\right) \right\| \overset{a.s.}{\to} 0.$$

Further, by (226), $D_{P_n} \to D$ and, Assumption 7, for any $\eta > 0$, there exists $N \in \mathbb{N}$ such that

$$\sup_{\Delta \in \mathbf{\Delta}} \left\| \left(K^{\mathcal{C}}\right)^{-1} - \left(K_{P_n}^{\mathcal{C}}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta\right), \rho\right)\right)^{-1} \right\|_2 \leq \eta, \forall n \geq N. \tag{238}$$

In sum, by (230), (232), and (237), (238), for any $\eta, \nu > 0$, there exists $N \geq \mathbb{N}$ such that

$$P\left(\sup_{\Delta \in \mathbf{\Delta}} \|\psi_n(\Delta) - \phi_n(\Delta)\| < \nu\right) \geq 1 - \eta, \forall n \geq \mathbb{N}. \tag{239}$$

Hence, for a specific choice of $\nu = \kappa(1-a)$, where a is defined in equation (234), we have that $\sup_{\Delta \in \mathbf{\Delta}} \|\psi_n(\Delta) - \phi_n(\Delta)\| < \kappa(1-a)$ implies

$$\begin{aligned} \left\| \Delta_n^{\mathcal{C}} - \Delta_n^f \right\| &= \left\| \psi_n\left(\Delta_n^{\mathcal{C}}\right) - \phi_n\left(\Delta_n^f\right) \right\| \\ &\leq \left\| \psi_n\left(\Delta_n^{\mathcal{C}}\right) - \phi_n\left(\Delta_n^{\mathcal{C}}\right) \right\| + \left\| \phi_n\left(\Delta_n^{\mathcal{C}}\right) - \phi_n\left(\Delta_n^f\right) \right\| \\ &\leq \kappa(1-a) + a \left\| \Delta_n^{\mathcal{C}} - \Delta_n^f \right\| \end{aligned}$$

Rearranging terms, we obtain $\left\| \Delta_n^{\mathcal{C}} - \Delta_n^f \right\| \leq \kappa$.

Step 4. I complete the proof. Note that by Assumptions 7 and 12.1, for any $\delta > 0$, there exists $\kappa_\delta > 0$ and $N \in \mathbb{N}$ such that

$$P\left(\sup_{\|\Delta - \Delta'\| \leq \kappa_\delta} \left| u_{n,j,\beta_n}^*(\Delta) - u_{n,j,\beta_n}^*\left(\Delta'\right) \right| < \delta\right) \geq 1 - \eta, \forall n \geq N. \tag{240}$$

For $\Delta_n^{\mathcal{C}} \in \mathfrak{W}^{*,-\delta}(0)$, one has

$$\mathfrak{w}_j^*\left(\Delta_n^{\mathcal{C}}\right) + \delta \leq 0, j \in \mathbb{J}_1. \tag{241}$$

Hence, by (219), (222), and (240)-(241), $\left\|\Delta_n^{\mathcal{C}} - \Delta_n^f\right\| \leq \kappa_{\delta/4}$, for each $j \in \mathbb{J}_1$ we have

$$u_{n,j,\beta_n}^*\left(\Delta_n^f\right) - c_n^*(\beta_n) \leq u_{n,j,\beta_n}^*\left(\Delta_n^{\mathcal{C}}\right) - c_n^*(\beta_n) + \delta/4 \leq \mathfrak{w}_j^*\left(\Delta_n^{\mathcal{C}}\right) + \delta/2 \leq 0.$$

For $j \in \{|\mathbb{J}_1| + 1, \ldots, |\mathbb{J}_1| + 2|\mathbb{J}_2|\}$, the inequalities hold by construction given the definition of $\mathcal{C}$. In sum, for any $\eta > 0$ there exists $\delta > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$\begin{aligned}
&P\left(\{U_n^*(\beta_n, c_n^*) = \emptyset\} \cap \left\{\mathfrak{W}^{*,-\delta}(0) \neq \emptyset\right\}\right) \leq P\left(\nexists \Delta_n^f \in U_n^*(\beta_n, c_n^*), \exists \Delta_n^{\mathcal{C}} \in \mathfrak{W}^{*,-\delta}(0)\right)\\
&\leq P\left(\left\{\sup_{\Delta \in \mathbf{\Delta}} \|\psi_n(\Delta) - \phi_n(\Delta)\| < \kappa_{\delta/4}(1-a) \cap A_n(\frac{\delta}{2})\right\}^c\right) \leq \eta/3,
\end{aligned}$$

where A^c denotes the complement of the set A, and the last inequality follows from (219) and (239). □

Lemma 15. *Suppose Assumptions 7, 8, 6, 10 hold. In addition, suppose Assumption 9 or 11 hold. Let $\{P_n, \beta_n, \theta_n\}$ be a sequence satisfying (190)-(191). Then, for any $\{\beta_n'\}$ such that $\beta_n' \in (\beta_n + \rho/\sqrt{n}\mathbf{\Delta}) \cap \overline{\mathcal{B}}$ for all n, it holds that*

1. *For any $c > 0$,*
$$P_n^b\left(V_n^I(\beta_n', c) \neq \emptyset\right) - P(\mathfrak{W}(c) \neq \emptyset) \to 0,$$
with probability approaching 1 ;
2. *If $c_{\pi^*} > 0, c_n^I(\beta_n') \xrightarrow{P} c_{\pi^*}$;*
3. $$\hat{c}_n(\beta_n') \geq c_n^I(\beta_n') + o_P(1).$$

Proof. (i) Throughout, let $c > 0$ and let $\{\beta_n'\}$ be a sequence such that $\beta_n' \in (\beta_n + \rho/\sqrt{n}\mathbf{\Delta}) \cap \overline{\mathcal{B}}$ for all n. By Assumption 7, $\left\|\hat{D}(\beta_n') - D_{P_n}(\beta_n)\right\| \xrightarrow{p} 0$. Further, by Lemma 23, $\hat{\xi}_{n,j}(\beta_n') \xrightarrow{P} \pi_{1,j}$. Therefore,

$$\left(\mathbb{Z}_n^b(\beta_n'), \hat{D}(\beta_n'), \hat{\xi}_n(\beta_n')\right) \mid \{X_i\}_{i=1}^\infty \xrightarrow{d} (\mathbb{Z}, D, \pi_1) \tag{242}$$

for almost all sample paths $\{X_i\}_{i=1}^\infty$. By Lemma H.17 in Kaido et al. (2019), conditional on the sample path, there exists an almost sure representation $\left(\tilde{\mathbb{Z}}_n^b, \tilde{D}_n, \tilde{\xi}_n\right)$ of $\left(\mathbb{Z}_n^b(\beta_n'), \hat{D}(\beta_n'), \hat{\xi}_n(\beta_n')\right)$ defined on another probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ such that $\left(\tilde{\mathbb{Z}}_n^b, \tilde{D}_n, \tilde{\xi}_n\right) \stackrel{d}{=} \left(\mathbb{Z}_n^b(\beta_n'), \hat{D}(\beta_n'), \hat{\xi}_n(\beta_n')\right)$ conditional on the sample path. In particular, conditional on the sample, $\left(\hat{D}(\beta_n'), \hat{\xi}_n(\beta_n')\right)$ are non-stochastic. Therefore, we set $\left(\tilde{D}_n, \tilde{\xi}_n\right) = \left(\hat{D}(\beta_n'), \hat{\xi}_n(\beta_n')\right)$, $\tilde{P} - a.s.$ The almost sure representation satisfies $\left(\tilde{\mathbb{Z}}_n^b, \tilde{D}_n, \tilde{\xi}_{n,j}\right) \xrightarrow{\text{a.s.}} \left(\tilde{\mathbb{Z}}, D, \pi_1\right)$ for almost all sample paths, where $\tilde{\mathbb{Z}} \stackrel{d}{=} \mathbb{Z}$. The almost sure representation $\left(\tilde{\mathbb{Z}}_n^b, \tilde{D}_n, \tilde{\xi}_n\right)$ is defined for each sample path $x^\infty = \{x_i\}_{i=1}^\infty$, but we suppress its dependence on x^∞ for notational simplicity. Using this representation, define

$$\begin{aligned}
\tilde{v}_{n,j,\beta_n'}^I(\Delta) &= \tilde{\mathbb{Z}}_{n,j}^s + \rho\tilde{D}_{n,j}\Delta + \varphi_j^*\left(\tilde{\xi}_{n,j}\right), && (243)\\
\tilde{v}_{n,\ell,\beta_n'}^I(\Delta) &= \tilde{\mathbb{Z}}_{n,\ell}^s + \rho\tilde{D}_{n,\ell}\Delta,
\end{aligned}$$

$$
\begin{aligned}
\tilde{v}^I_{n,u,\beta'_n}(\Delta) &= \tilde{\mathbb{Z}}^s_{n,u} + \rho\tilde{D}_{n,u}\Delta + \varphi^*_0\left(\tilde{\xi}_{n,0}\right), \\
\tilde{\mathfrak{w}}_j(\Delta) &= \tilde{\mathbb{Z}}_j + \rho D_j\Delta + \pi^*_{1,j} \\
\tilde{\mathfrak{w}}_\ell(\Delta) &= \tilde{\mathbb{Z}}_\ell + \rho D_\ell\Delta \\
\tilde{\mathfrak{w}}_u(\Delta) &= \tilde{\mathbb{Z}}_u + \rho D_u\Delta + \pi^*_{1,0}
\end{aligned}
$$

where $\tilde{\mathbb{Z}} \overset{d}{=} \mathbb{Z}$, and $\tilde{\mathbb{Z}}^b_n \to \tilde{\mathbb{Z}}, \tilde{P} - a.s.$ conditional on $\{X_i\}_{i=1}^\infty$. With this construction, one may write

$$
\begin{aligned}
&\left|P^s_n\left(V^I_n\left(\beta'_n, c\right) \neq \emptyset\right) - P(\mathfrak{W}(c) \neq \emptyset)\right| \\
=&\left|\tilde{P}\left(\tilde{V}^I_n\left(\beta'_n, c\right) \neq \emptyset\right) - \tilde{P}(\tilde{\mathfrak{W}}(c) \neq \emptyset)\right| \\
\leq&\tilde{P}\left(\tilde{V}^I_n\left(\beta'_n, c\right) = \emptyset \cap \tilde{\mathfrak{W}}(c) \neq \emptyset\right) + \tilde{P}\left(\tilde{V}^I_n\left(\beta'_n, c\right) \neq \emptyset \cap \tilde{\mathfrak{W}}(c) = \emptyset\right).
\end{aligned} \tag{244}
$$

First, we bound the first term of (244). Note that

$$
\begin{aligned}
&\tilde{P}\left(\tilde{V}^I_n\left(\beta'_n, c\right) = \emptyset \cap \tilde{\mathfrak{W}}(c) \neq \emptyset\right) \\
\leq&\tilde{P}\left(\tilde{V}^I_n\left(\beta'_n, c+\delta\right) = \emptyset \cap \tilde{\mathfrak{W}}(c) \neq \emptyset\right) + \tilde{P}\left(\tilde{V}^I_n\left(\beta'_n, c+\delta\right) \neq \emptyset \cap \tilde{V}^I_n\left(\beta'_n, c\right) = \emptyset\right).
\end{aligned} \tag{245}
$$

Let

$$
A_n = \left\{\tilde{\omega} \in \tilde{\Omega} : \sup_{\Delta\in\boldsymbol{\Delta}} \max_{j\in\mathbb{J}^*} \left|\tilde{v}^I_{n,j,\beta'_n}(\Delta) - \tilde{\mathfrak{w}}_j(\Delta)\right| \geq \delta\right\}.
$$

Let

$$
E = \left\{\{x_i\}_{i=1}^\infty : \left\|\hat{D}\left(\beta'_n\right) - D\right\| < \eta,\ \max_{j\in\mathbb{J}^*}\left|\phi^*_j\left(\hat{\xi}_{n,j}\left(\beta'_n\right)\right) - \pi^*_{1,j}\right| < \eta\right\}.
$$

Note that, $P_n(E) \geq 1-\eta$ for all n sufficiently large by Assumption 7 and Lemma 23. On E, we therefore have $\left\|\tilde{D}_n - D\right\| < \eta$ and $\max_{j\in\mathbb{J}^*}\left|\phi^*_j\left(\tilde{\xi}_{n,j}\right) - \pi^*_{1,j}\right| < \eta, \tilde{P} - a.s.$ Below, we condition on $\{X_i\}_{i=1}^\infty \in E$. For any $j \in \mathbb{J}^*$,

$$
\left|\tilde{v}^I_{n,j,\beta'_n}(\Delta) - \tilde{\mathfrak{w}}_j(\Delta)\right| \leq \left|\tilde{\mathbb{Z}}^s_{n,j} - \tilde{\mathbb{Z}}_j\right| + \rho\left\|\tilde{D}_{j,n} - D_j\right\|\|\Delta\| + \left|\phi^*_j\left(\tilde{\xi}_{n,j}\right) - \pi^*_{1,j}\right| \leq (2+\rho)\eta,
$$

uniformly in $\Delta \in \boldsymbol{\Delta}$, where we used $\tilde{\mathbb{Z}}^s_n \to \tilde{\mathbb{Z}}, \tilde{P}$ - a.s. Since η can be chosen arbitrarily small, this in turn implies

$$
\tilde{P}\left(A_n\right) < \eta/2,
$$

for all n sufficiently large. Note also that $\sup_{\Delta\in\boldsymbol{\Delta}}\max_{j\in\mathbb{J}}\left|\tilde{v}^I_{n,j,\beta'_n}(\Delta) - \tilde{\mathfrak{w}}_j(\Delta)\right| < \delta$ implies $\tilde{\mathfrak{W}}(c) \subseteq \tilde{V}^I_n\left(\beta'_n, c+\delta\right)$, and hence A^c_n is a subset of

$$
L_n = \left\{\tilde{\omega} \in \tilde{\Omega} : \tilde{\mathfrak{W}}(c) \subseteq \tilde{V}^I_n\left(\beta'_n, c+\delta\right)\right\}.
$$

Using this,

$$
\begin{aligned}
\tilde{P}\left(\tilde{V}^I_n\left(\beta'_n, c+\delta\right) = \emptyset \cap \tilde{\mathfrak{W}}(c) \neq \emptyset\right) &\leq \tilde{P}\left(\tilde{\mathfrak{W}}(c) \nsubseteq \tilde{V}^I_n\left(\beta'_n, c+\delta\right)\right) \\
&= \tilde{P}\left(L^c_n\right) \leq \tilde{P}\left(A_n\right) < \eta/2,
\end{aligned} \tag{246}
$$

for all n sufficiently large. Also, by Lemma 17,

$$\tilde{P}\left(\tilde{V}_n^I\left(\beta_n', c+\delta\right) \neq \emptyset \cap \tilde{V}_n^I\left(\beta_n', c\right) = \emptyset\right) < \eta/2, \tag{247}$$

for all n sufficiently large. Combining (245), (246), (247), and using $P_n(E) \geq 1-\eta$ for all n, we have

$$\begin{aligned}&\int_E \tilde{P}\left(\tilde{V}_n^I\left(\beta_n', c\right) = \emptyset \cap \tilde{\mathfrak{W}}(c) \neq \emptyset\right) dP_n + \int_{E^c} \tilde{P}\left(\tilde{V}_n^I\left(\beta_n', c\right) = \emptyset \cap \tilde{\mathfrak{W}}(c) \neq \emptyset\right) dP_n \\ \leq &\eta(1-\eta) + \eta \leq 2\eta.\end{aligned}$$

The second term of the right hand side of (244) can be bounded similarly. Therefore,

$$\left|P^s\left(V_n^I\left(\beta_n', c\right) \neq \emptyset\right) - P\left(\mathfrak{W}(c) \neq \emptyset\right)\right| \to 0$$

with probability (under P_n) approaching 1. This establishes the first claim.

(ii) By Part (i), for $c > 0$, we have

$$P_n^s\left(V_n^I\left(\beta_n', c\right) \neq \emptyset\right) - P\left(\mathfrak{W}(c) \neq \emptyset\right) \to 0$$

Fix $c > 0$, and set

$$g_j = \begin{cases} c - \mathbb{Z}_j, & j \in \mathbb{J}, \\ 1, & j = |\mathbb{J}|+1, \ldots, |\mathbb{J}|+2d, \end{cases}$$

Mimic the argument following (247) and apply Lemma 17. Then, there is $\delta > 0$ such that

$$\begin{aligned}|P(\mathfrak{W}(c) \neq \emptyset) - P(\mathfrak{W}(c-\delta) \neq \emptyset)| &= P\left(\{\mathfrak{W}(c) \neq \emptyset\} \cap \{\mathfrak{W}(c-\delta) = \emptyset\}\right) \leq \eta \\ |P\left(\mathfrak{W}(c+\delta) \neq \emptyset\right) - P\left(\mathfrak{W}(c) \neq \emptyset\right)| &= P\left(\{\mathfrak{W}(c+\delta) \neq \emptyset\} \cap \{\mathfrak{W}(c) = \emptyset\}\right) \leq \eta\end{aligned}$$

which therefore ensures that $c \mapsto P(\mathfrak{W}(c) \neq \emptyset)$ is continuous at $c > 0$.

Next, we show $c \mapsto P(\mathfrak{W}(c) \neq \emptyset)$ is strictly increasing at any $c > 0$. For this, consider $c > 0$ and $c - \delta > 0$ for $\delta > 0$. Define the $|\mathbb{J}|$ vector e to have elements $e_j = c - \mathbb{Z}_j, j \in \mathbb{J}$. Suppose for simplicity that $\mathbb{J}^* = \left\{j \in \mathbb{J} : \pi_{1,j}^* = 0\right\}$ contains the first J^* inequality constraints. Let $e^{\mathbb{J}^*}$ denote the subvector of e that only contains elements corresponding to $j \in \mathbb{J}^*$, define $D^{\mathbb{J}^*}$ correspondingly, and write

$$K = \begin{bmatrix} D^{\mathbb{J}^*:} \\ I_d \\ -I_d \end{bmatrix}, g = \begin{bmatrix} e^{\mathbb{J}^*} \\ \rho \cdot 1_d \\ \rho \cdot 1_d \end{bmatrix}, \tau = \begin{bmatrix} 1_{J*} \\ 0_d \\ 0_d \end{bmatrix}. \tag{248}$$

By Farkas' lemma (R. Rockafellar (1970) Theorem 22.1) and arguing as in (268),

$$\begin{aligned}&P\left(\{\mathfrak{W}(c) \neq \emptyset\} \cap \{\mathfrak{W}(c-\delta) = \emptyset\}\right) \\ =&P\left(\left\{\mu' g \geq 0, \forall \mu \in \mathcal{M}\right\} \cap \left\{\mu'(g - \delta\tau) < 0, \exists \mu \in \mathcal{M}\right\}\right),\end{aligned} \tag{249}$$

where $\mathcal{M} = \left\{\mu \in \mathbb{R}_+^{J^*+2d} : \mu' K = 0\right\}$. By Minkowski-Weyl's theorem (R. T. Rockafellar and Wets (2009) Theorem 3.52), there exists $\left\{\nu^t \in \mathcal{M}, t = 1, \ldots, T\right\}$, for which one may write

$$\mathcal{M} = \left\{\mu : \mu = b \sum_{t=1}^{T} a_t \nu^t, b > 0, a_t \geq 0, \sum_{t=1}^{T} a_t = 1\right\}.$$

This implies

$$\mu' g \geq 0, \forall \mu \in \mathcal{M} \Leftrightarrow \nu^{t\prime} g \geq 0, \forall t \in \{1, \ldots, T\},$$

$$\mu'(g - \delta\tau) < 0, \exists \mu \in \mathcal{M} \Leftrightarrow \nu^{t\prime} g < \delta \nu^{t\prime}\tau, \exists t \in \{1, \ldots, T\}.$$

Hence,

$$(249) = P\left(0 \leq \nu^{s\prime} g, 0 \leq \nu^{t\prime} g < \delta \nu^{t\prime}\tau, \forall s, \exists t\right) \tag{250}$$

Note that by (248), for each $s \in \{1, \ldots, T\}$,

$$\nu^{s\prime} g = (\nu^{s,\mathbb{J}^*})' \left(c 1_{\mathbb{J}^*} - \mathbb{Z}_{\mathbb{J}^*}\right) + \rho \sum_{j=J^*+1}^{J^*+2d} \nu^{s,j},$$

$$\nu^{s\prime}\tau = \sum_{j=1}^{J^*} \nu^{s,j}.$$

Hence

$$h_s^U = c \sum_{j=1}^{J^*} \nu^{s,j} + \rho \sum_{j=J^*+1}^{J^*+2d} \nu^{s,j}$$

$$h_s^L = (c - \delta) \sum_{j=1}^{J^*} \nu^{s,j}$$

where $0 \leq h_s^L < h_s^U$ for all $s \in \{1, \ldots, T\}$ due to $0 < c - \delta < c$ and $\nu^s \in \mathbb{R}_+^{J^*+2d}$. One may therefore rewrite the probability on the right hand side of (250) as

$$\begin{aligned} &P\left(0 \leq \nu^{s\prime} g, 0 \leq \nu^{t\prime} g < \delta \nu^{t\prime}\tau, \forall s, \exists t\right) \\ =&P\left(\nu^{s,\mathbb{J}^*\prime} \mathbb{Z}_{\mathbb{J}^*} \leq h_s^U, h_t^L < \nu^{t,\mathbb{J}^*\prime} \mathbb{Z}_{\mathbb{J}^*} \leq h_t^U,\ \forall s,\ \exists t\right) > 0, \end{aligned} \tag{251}$$

where the last inequality follows because $\mathbb{Z}_{\mathbb{J}^*}$'s correlation matrix Ω has an eigenvalue bounded away from 0 by Assumption 9 and 11. By (249), (250), and (251), $c \mapsto P(\mathfrak{W}(c) \neq \emptyset)$ is strictly increasing at any $c > 0$.

Suppose that $c_{\pi^*} > 0$, then arguing as in Lemma 5.(i) of D. W. K. Andrews and Guggenberger (2010), we obtain $c_n^I(\beta_n') \stackrel{P_n}{\to} c_{\pi^*}$

(iii) Begin with observing that one can equivalently express $\hat{c}_\ell$ (originally defined in (53)) as $\hat{c}_\ell(\beta) = \inf\left\{c \in \mathbb{R}_+ : P_n^s\left(V_n^s(\beta, c) \neq \emptyset\right) \geq 1 - \alpha\right\}$ where

$$V_n^s\left(\beta_n', c\right) = \left\{\Delta \in \mathbf{\Delta}_{n,\rho}^d : v_{n,j,\beta_n'}^s(\Delta) \leq c, j \in \mathbb{J}\right\}, \tag{252}$$

$$v_{n,j,\beta'_n}^s(\Delta) = \mathbb{Z}_{n,j}^s\left(\beta'_n\right) + \rho\hat{D}_{n,j}\left(\beta'_n\right)\Delta + \hat{\xi}_{n,j}\left(\beta'_n\right). \tag{253}$$

Suppose first that Assumption 9 holds. In this case, there are no paired inequalities, and V_n^I differs from V_n^s only in terms of the function ϕ_j^* used in place of the GMS function ξ. In particular, $\phi_j^*(\xi) \leq \xi$ for any j and ξ, and therefore $\hat{c}_\ell(\beta_n) \geq c_n^I(\beta_n)$ by construction.

Next, suppose that Assumption 11 holds. The only case that might create concern is one in which

$$\pi_{1,j} \in [-1, 0) \text{ and } \pi_{1,j+R_1} = 0.$$

In this case, only the $j+R_1$-th inequality binds in the limit, but with probability approaching 1, GMS selects both in the pair. Therefore, we have

$$\pi_{1,j}^* = -\infty, \text{ and } \pi_{1,j+R_1}^* = 0,$$

$$\hat{\xi}_{n,j}\left(\beta'_n\right) = 0, \text{ and } \hat{\xi}_{n,j+R_1}\left(\beta'_n\right) = 0,$$

so that in $V_n^I(\beta'_n, c)$, inequality $j+R_1$, which is

$$\mathbb{Z}_{n,j+R_1}^s\left(\beta'_n\right) + \rho\hat{D}_{n,j+R_1}\left(\beta'_n\right)\Delta \leq c$$

is replaced with inequality

$$-\mathbb{Z}_{n,j}^s\left(\beta'_n\right) - \rho\hat{D}_{n,j}\left(\beta'_n\right)\Delta \leq c,$$

as explained in Section D.3. In this case, $\hat{c}_n(\beta_n) \geq c_n^I(\beta_n)$ is not guaranteed in finite sample. However, let v_n^{IP} be as in (203) but for $j \in \mathbb{J}_2$, replacing $[j]$-th component $v_{n,[j]}^I$ with $-v_{n,j}^I$. Define V_n^{IP} as in (202) but replacing v_n^I with v_n^{IP}. Define

$$c_n^{IP}(\beta_n) = \inf\left\{c \in \mathbb{R}_+ : P^*\left(V_n^{IP}(\beta_n, c)\right) \geq 1-\alpha\right\}.$$

By construction, $\hat{c}_n(\beta'_n) \geq c_n^{IP}(\beta'_n)$ for any $\beta'_n \in \left(\beta_n + \rho/\sqrt{n}B^d\right) \cap \overline{\mathcal{B}}$. Therefore, it suffices to show that $c_n^{IP}(\beta'_n) - c_n^I(\beta'_n) \xrightarrow{p} 0$. For this, note that Lemma 20.3 and 20.4 establishes

$$\sup_{\Delta \in \boldsymbol{\Delta}_{n,\rho}} \left\| \mathbb{Z}_{n,j+R_1}^b\left(\beta'_n\right) + \rho\hat{D}_{n,j+R_1}\left(\beta'_n\right)\Delta + \mathbb{Z}_{n,j}^b\left(\beta'_n\right) + \rho\hat{D}_{n,j}\left(\beta'_n\right)\Delta \right\| = o_P(1),$$

for almost all sample paths $\{X_i\}_{i=1}^\infty$. Therefore, replacing the $j+R_1$-th inequality with the j-th inequality in V_n^{IP} is asymptotically negligible. Mimicking the arguments in Parts (i) and (ii) then yields

$$c_n^{IP}\left(\beta'_n\right) \xrightarrow{p} c_{\pi^*}.$$

This therefore ensures $c_n^{IP}(\beta'_n) - c_n^I(\beta'_n) \xrightarrow{p} 0$. □

Lemma 16. *Suppose Assumptions 7, 8, 6, 10 hold. In addition, suppose Assumption 9 or 11 hold. For any $\varepsilon, \eta > 0$ and $\beta'_n \in \left(\beta_n + \rho/\sqrt{n}[-1,1]^d\right) \cap \overline{\mathcal{B}}$, there exists $N' \in \mathbb{N}$ such that for all $n \geq N'$,*

$$P\left(\sup_{\Delta \in \boldsymbol{\Delta}} \left| \max_{j \in \mathbb{J}}\left(u_{n,j,\beta_n}^*(\Delta) - c_n^*\right) - \max_{j \in \mathbb{J}}\left(\mathfrak{w}_j^*(\Delta) - c_{\pi^*}\right) \right| \geq \varepsilon\right) < \eta, \tag{254}$$

$$\tilde{P}\left(\sup_{\Delta\in\boldsymbol{\Delta}}\left|\max_{j\in\mathbb{J}}\tilde{\mathfrak{w}}_j(\Delta)-\max_{j\in\mathbb{J}}\tilde{v}^I_{n,j,\beta'_n}(\Delta)\right|\geq\varepsilon\right)<\eta,\ \ w.p.\ 1. \tag{255}$$

Proof. We first establish (254). By definition, $\pi^*_{1,j}=-\infty$ for all $j\notin\mathbb{J}^*$ and therefore

$$\begin{aligned}
&P\left(\sup_{\Delta\in\boldsymbol{\Delta}}\left|\max_{j\in\mathbb{J}}\left(u^*_{n,j,\beta_n}(\Delta)-c^*_n\right)-\max_{j\in\mathbb{J}}\left(\mathfrak{w}^*_j(\Delta)-c_{\pi^*}\right)\right|\geq\varepsilon\right)\\
=&P\left(\sup_{\Delta\in\boldsymbol{\Delta}}\left|\max_{j\in\mathbb{J}^*}\left(u^*_{n,j,\beta_n}(\Delta)-c^*_n\right)-\max_{j\in\mathbb{J}^*}\left(\mathfrak{w}^*_j(\Delta)-c_{\pi^*}\right)\right|\geq\varepsilon\right).
\end{aligned}\tag{256}$$

Hence, for the conclusion of the lemma, it suffices to show, for any $\varepsilon>0$,

$$\lim_{n\to\infty}P\left(\sup_{\Delta\in\boldsymbol{\Delta}}\left|\max_{j\in\mathbb{J}^*}\left(u^*_{n,j,\beta_n}(\Delta)-c^*_n\right)-\max_{j\in\mathbb{J}^*}\left(\mathfrak{w}^*_j(\Delta)-c_{\pi^*}\right)\right|\geq\varepsilon\right)=0.$$

For each $\Delta\in\mathbb{R}^d$, define $r_{n,j,\beta_n}(\Delta)\equiv\left(u^*_{n,j,\beta_n}(\Delta)-c^*_n\right)-\left(\mathfrak{w}^*_j(\Delta)-c_{\pi^*}\right)$. Using the fact that $\pi^*_{1,j}=0$ for $j\in\mathbb{J}^*$, and the triangle and Cauchy-Schwarz inequalities, for any $\Delta\in\boldsymbol{\Delta}\cap\frac{\sqrt{n}}{\rho}\left(\overline{\mathcal{B}}-\beta_n\right)$ and $j\in\mathbb{J}^*$, we have

$$\begin{aligned}
|r_{n,j,\beta_n}(\Delta)|\leq&\left|\mathbb{Z}^*_{n,j,\beta_n}-\mathbb{Z}^*_j\right|+\rho\left\|\frac{\sigma_{F,j}\left(\bar{\beta}_n\right)}{\sigma_{F,j}\left(\beta_n\right)}D_{P_n,j}\left(\bar{\beta}_n\right)-D_j\right\|\|\Delta\|+|c^*_n-c_{\pi^*}|\\
&+|\hat{\eta}_{n,j,\beta_n}(\Delta)|\left\|\mathbb{Z}_{n,j,\beta_n}(\Delta)+\frac{\sigma_{P_n,j}\left(\bar{\beta}_n\right)}{\sigma_{P_n,j}\left(\beta_n\right)}D_{P_n,j}\left(\bar{\beta}_n\right)\Delta\rho\right\|\\
=&o_P(1)
\end{aligned}\tag{257}$$

where the first equality follows from $\|\Delta\|\leq\sqrt{d}$, $D_{P_n}\left(\bar{\beta}_n\right)\to D$ due to $D_{P_n}\left(\beta_n\right)\to D$, Assumption 7, Assumption 8, and $\bar{\beta}_n$ being a mean value between β_n and $\beta_n+\Delta\rho/\sqrt{n}$. We also note that $\left\|\mathbb{Z}^*_{n,j,\beta_n}\right\|=O_P(1)$, $\|D_{P,j}(\beta)\|$ being uniformly bounded for $\beta\in\mathcal{B}(P)$ (Assumption 8.1), and Lemma 21.

Note that when paired inequalities are merged, for each $j=1,\ldots,R_1$ such that $\pi^*_{1,j}=0=\pi^*_{1,j+R_1}$ we have that $|\tilde{\mu}_j-\mu_j|=o_P(1)$, where $\tilde{\mu}_j$ and μ_j were defined in (168)-(169) and (193)-(194) respectively. By (257) and the fact that $j\in\mathbb{J}^*$, we have

$$\sup_{\Delta\in\boldsymbol{\Delta}}\left|\max_{j\in\mathbb{J}^*}\left(u^*_{n,j,\beta_n}(\Delta)-c^*_n\right)-\max_{j\in\mathbb{J}^*}\left(\mathfrak{w}^*_j(\Delta)-c_{\pi}\right)\right|\leq\sup_{\Delta\in\boldsymbol{\Delta}}\max_{j\in\mathbb{J}^*}|r_{n,j,\beta_n}(\Delta)|=o_P(1). \tag{258}$$

The conclusion of the lemma then follows from (256) and (258). The result in (255) follows from similar arguments. □

Lemma 17. *Let Assumptions 7, 8, 6, 10 hold. And suppose Assumption 9 or 11 hold. For any $\beta'_n\in(\beta_n+\rho/\sqrt{n}\boldsymbol{\Delta})\cap\overline{\mathcal{B}}$,*

1. *For any $\eta>0$, there exist $\delta>0$ such that*

$$\sup_{c\geq0}P\left(\{\mathfrak{W}(c)\neq\emptyset\}\cap\left\{\mathfrak{W}^{-\delta}(c)=\emptyset\right\}\right)<\eta. \tag{259}$$

Moreover, for any $\eta > 0$, there exist $\delta > 0$ and $N \in \mathbb{N}$ such that

$$\sup_{c \geq 0} P_n^s \left(\left\{ V_n^I \left(\beta_n', c \right) \neq \emptyset \right\} \cap \left\{ V_n^{I,-\delta} \left(\beta_n', c \right) = \emptyset \right\} \right) < \eta, \quad \forall n \geq N. \tag{260}$$

2. *Fix $\underline{c} > 0$. Then for any $\eta > 0$, there exists $\delta > 0$ such that*

$$\sup_{c \geq \underline{c}} P \left(\{ \mathfrak{W}(c) \neq \emptyset \} \cap \{ \mathfrak{W}(c - \delta) = \emptyset \} \right) < \eta. \tag{261}$$

Moreover, for any $\eta > 0$, there exist $\delta > 0$ and $N \in \mathbb{N}$ such that

$$\sup_{c \geq \underline{c}} P_n^s \left(\left\{ V_n^I \left(\beta_n', c \right) \neq \emptyset \right\} \cap \left\{ V_n^I \left(\beta_n', c - \delta \right) = \emptyset \right\} \right) < \eta, \quad \forall n \geq N. \tag{262}$$

Proof. We first show (259). Any inequality indexed by $j \notin \mathbb{J}^*$ is satisfied with probability approaching one by similar arguments as in (179) (both with c and with $c - \delta$). Hence, one could argue for sets $\mathfrak{W}(c), \mathfrak{W}^{-\delta}(c)$ but with $j \in \mathbb{J}^*$. To keep the notation simple, below I argue as if $\mathbb{J} = \mathbb{J}^*$. Let $c \geq 0$ be given. Let g be a $|\mathbb{J}| + 2d$ vector with entries

$$g_j = \begin{cases} c - \mathbb{Z}_j, & j \in \mathbb{J}, \\ 1, & j = |\mathbb{J}| + 1, \ldots, |\mathbb{J}| + 2d, \end{cases} \tag{263}$$

recalling that $\pi_{1,j}^* = 0$ for $j = |\mathbb{J}_1| + 1, \ldots, |\mathbb{J}|$. Let τ be a $(|\mathbb{J}| + 2d)$ vector with entries

$$\tau_j = \begin{cases} 1, & j = 1, \ldots, |\mathbb{J}_1| \\ 0, & j = |\mathbb{J}_1| + 1, \ldots, |\mathbb{J}| + 2d. \end{cases} \tag{264}$$

Then we can express the sets of interest as

$$\mathfrak{W}(c) = \{ \Delta : K\Delta \leq g \}, \tag{265}$$

$$\mathfrak{W}^{-\delta}(c) = \{ \Delta : K\Delta \leq g - \delta\tau \} \tag{266}$$

By Farkas' Lemma, e.g. R. Rockafellar (1970) (Theorem 22.1), a solution to the system of linear inequalities in (265) exists if and only if for all $\mu \in \mathbb{R}_+^{J+2d}$ such that $\mu' K = 0$, one has $\mu' g \geq 0$. Similarly, a solution to the system of linear inequalities in (266) exists if and only if for all $\mu \in \mathbb{R}^{|\mathbb{J}|+2d}$ such that $\mu' K = 0$, one has $\mu'(g - \delta\tau) \geq 0$. Define

$$\mathcal{M} \equiv \left\{ \mu \in \mathbb{R}_+^{|\mathbb{J}|+2d} : \mu' K = 0 \right\}. \tag{267}$$

Then, one may write

$$\begin{aligned} & P \left(\{ \mathfrak{W}(c) \neq \emptyset \} \cap \left\{ \mathfrak{W}^{-\delta}(c) = \emptyset \right\} \right) \\ = & P \left(\left\{ \mu' g \geq 0, \forall \mu \in \mathcal{M} \right\} \cap \left\{ \mu'(g - \delta\tau) < 0, \exists \mu \in \mathcal{M} \right\} \right) \\ = & P \left(\left\{ \mu' g \geq 0, \forall \mu \in \mathcal{M} \right\} \cap \left\{ \mu' g < \delta \mu' \tau, \exists \mu \in \mathcal{M} \right\} \right). \end{aligned} \tag{268}$$

Note that the set $\mathcal{M}$ is a non-stochastic polyhedral cone. By Minkowski Weyl's theorem (see, e.g. R. T. Rockafellar and Wets (2009) (Theorem 3.52)), there exist $\{\nu^t \in \mathcal{M}, t = 1, \ldots, T\}$, with $T < \infty$ a constant that depends only on $|\mathbb{J}|$ and d, such that any $\mu \in \mathcal{M}$ can be represented as

$$\mu = b \sum_{t=1}^{T} a_t \nu^t,$$

where $b > 0$ and $a_t \geq 0$, $t = 1, \ldots, T$, $\sum_{t=1}^{T} a_t = 1$. Hence, if $\mu \in \mathcal{M}$ satisfies $\mu' g < \delta \mu' \tau$, denoting $\nu^{t\prime}$ the transpose of vector ν^t, we have

$$\sum_{t=1}^{T} a_t \nu^{t\prime} g < \delta \sum_{t=1}^{T} a_t \nu^{t\prime} \tau$$

However, due to $a_t \geq 0, \forall t$ and $\nu^t \in \mathcal{M}$, this means $\nu^{t\prime} g < \delta \nu^{t\prime} \tau$ for some $t \in \{1, \ldots, T\}$. Furthermore, since $\nu^t \in \mathcal{M}$, we have $0 \leq \nu^{t\prime} g$. Therefore,

$$\begin{aligned} & P\left(\left\{\mu' g \geq 0, \forall \mu \in \mathcal{M}\right\} \cap \left\{\mu' g < \delta \mu' \tau, \exists \mu \in \mathcal{M}\right\}\right) \\ & \leq P\left(0 \leq \nu^{t\prime} g < \delta \nu^{t\prime} \tau, \exists t \in \{1, \ldots, T\}\right) \leq \sum_{t=1}^{T} P\left(0 \leq \nu^{t\prime} g < \delta \nu^{t\prime} \tau\right). \end{aligned} \tag{269}$$

Case 1. Consider first any $t = 1, \ldots, T$ such that ν^t assigns positive weight only to constraints in $\{|\mathbb{J}| + 1, \ldots, |\mathbb{J}| + 2d\}$. Then

$$\nu^{t\prime} g = \sum_{j=|\mathbb{J}|+1}^{|\mathbb{J}|+2d} \nu_j^t,$$

$$\delta \nu^{t\prime} \tau = \delta \sum_{j=|\mathbb{J}|+1}^{|\mathbb{J}|+2d} \nu_j^t \tau_j = 0,$$

where the last equality follows by (264). Therefore $P\left(0 \leq \nu^{t\prime} g < \delta \nu^{t\prime} \tau\right) = 0$.

Case 2. Consider now any $t = 1, \ldots, T$ such that ν^t assigns positive weight also to constraints in $\mathbb{J}$. Recall that indices $j = \mathbb{J}_2 \cup [\mathbb{J}_2]$ correspond to moment equalities, each of which is written as two moment inequalities, therefore yielding a total of $2|\mathbb{J}_2|$ inequalities with $D_{[j]} = -D_j$ for $j \in \mathbb{J}_2$, and:

$$g = \begin{cases} c - \mathbb{Z}_j & j \in \mathbb{J}_2, \\ c + \mathbb{Z}_{[j]} & j \in [\mathbb{J}_2]. \end{cases} \tag{270}$$

For each ν^t, (270) implies

$$\sum_{j \in \mathbb{J}_2 \cup [\mathbb{J}_2]} \nu_j^t g_j = c \sum_{j \in \mathbb{J}_2 \cup [\mathbb{J}_2]} \nu_j^t + \sum_{j \in \mathbb{J}_2} \left(\nu_j^t - \nu_{j+J_2}^t\right) \mathbb{Z}_j.$$

For each $j \in \mathbb{J}_1 \cup \mathbb{J}_2$, define

$$\tilde{\nu}_j^t \equiv \begin{cases} -\nu_j^t & j \in \mathbb{J}_1 \\ -\nu_j^t + \nu_{[j]}^t & j \in \mathbb{J}_2. \end{cases} \tag{271}$$

We then let $\tilde{\nu}^t = \left(\tilde{\nu}^t_{n,1}, \ldots, \tilde{\nu}^t_{n,|\mathbb{J}_1|+|\mathbb{J}_2|}\right)'$ and have

$$\nu^{t\prime} g = \sum_{j \in \mathbb{J}_1 \cup \mathbb{J}_2} \tilde{\nu}^t_j \mathbb{Z}_j + c \sum_{j \in \mathbb{J}} \nu^t_j + \sum_{j=|\mathbb{J}|+1}^{|\mathbb{J}|+2d} \nu^t_j \tag{272}$$

Case 2-a. Suppose $\tilde{\nu}^t \neq 0$. Then, by (272), $\frac{\nu^{t\prime} g}{\nu^{t\prime}\tau}$ is a normal random variable with variance $\left(\tilde{\nu}^{t\prime}\tau\right)^{-2} \tilde{\nu}^{t\prime}\Omega\tilde{\nu}^t$. By Assumption 9 and Assumption 11, there exists a constant $\omega > 0$ such that the smallest eigenvalue of Ω is bounded from below by ω for all β'_n. Hence, letting $\|\cdot\|_p$ denote the p-norm in $\mathbb{R}^{|\mathbb{J}|+2d+2}$, we have

$$\frac{\tilde{\nu}'^t \Omega \tilde{\nu}^t}{(\tilde{\nu}^{t\prime}\tau)^2} \geq \frac{\omega \left\|\tilde{\nu}^t\right\|_2^2}{(|\mathbb{J}|+2d)^2 \left\|\tilde{\nu}^t\right\|_2^2} \geq \frac{\omega}{(|\mathbb{J}|+2d)^2}.$$

Therefore, the variance of the normal random variable in (269) is uniformly bounded away from 0, which in turn allows one to find $\delta > 0$ such that $P\left(0 \leq \frac{\nu^t g}{\nu^t \tau} < \delta\right) \leq \eta/T$.

Case 2-b. Next, consider the case $\tilde{\nu}^t = 0$. Because we are in the case that ν^t assigns positive weight also to constraints in $\mathbb{J}$, this must be because $\nu^t_j = 0$ for all $j \in \mathbb{J}_1$ and $\nu^t_j = \nu^t_{[j]}$ for all $j \in \mathbb{J}_2$, while $\nu^t_j \neq 0$ for some $j \in \mathbb{J}_2$. Then we have $\sum_{j \in \mathbb{J}} \nu^t_j g \geq 0$, and $\sum_{j \in \mathbb{J}} \nu^t_j \tau_j = 0$ because $\tau_j = 0$ for each $j \in \mathbb{J}_2 \cup [\mathbb{J}_2]$. Hence, the argument for the case that ν^t assigns positive weight only to constraints in $\{|\mathbb{J}|+1, \ldots, |\mathbb{J}|+2d\}$ applies and again $P\left(0 \leq \nu^{t\prime} g < \delta \nu^{t\prime}\tau\right) = 0$. This establishes equation (259).

As for (260), observe that the bootstrap distribution is conditional on $X_1, \ldots, X_n$. Therefore, the matrix $\hat{K}_n$, defined as the matrix in equation (227) but with $\hat{D}_n$ replacing D_P, can be treated as nonstochastic. This implies that the set $\hat{\mathcal{M}}_n$, defined as the set in equation (267) but with $\hat{K}_n$ replacing K, can be treated as nonstochastic as well.

By an application of Lemma D.2.8 in Bugni et al. (2015) together with Lemma H.17 (through an argument similar to that following equation (242), $\mathbb{Z}^s_n \xrightarrow{d} \mathbb{Z}$ in $l^\infty(\Theta)$ uniformly in $\mathcal{P}$ conditional on $\{X_1, \ldots, X_n\}$, and by Assumption 7 $\hat{D}_n\left(\beta'_n\right) \xrightarrow{P_n} D$, for almost all sample paths. Set

$$g_{P_n,j}\left(\beta'_n\right) = \begin{cases} c - \phi^*_j\left(\xi_{n,j}\left(\beta'_n\right)\right) - \mathbb{Z}^s_{n,\beta'_n,j}, & j \in \mathbb{J}, \\ 1, & j = |\mathbb{J}|+1, \ldots, |\mathbb{J}|+2d, \end{cases}$$

and note that $\left|\phi^*_j\left(\xi_{n,j}\left(\beta'_n\right)\right)\right| < \eta$ for all $j \in \mathbb{J}^*$, and $\mathbb{Z}^s_{n,\beta'_n,j} \mid \{X_i\}_{i=1}^\infty \xrightarrow{d} N(0,\Omega)$. Then one can mimic the argument following (263) to conclude (260).

The results in (261)-(262) follow by similar arguments, with proper redefinition of τ in equation (264). $\square$

Lemma 18. *Let Assumptions 7, 8, 6, 10 hold. And suppose Assumption 9 or 11 hold. Let $\widetilde{\mathcal{C}}$ collect all size d subsets C of $\{1, \ldots, |\mathbb{J}|+2d\}$ ordered lexicographically by their smallest, then second smallest, etc. elements. Let the random variable $\mathcal{C}$ equal the first element of $\widetilde{\mathcal{C}}$ s.t. $\det K^C \neq 0$ and $\Delta^C = \left(K^C\right)^{-1} g^C \in \mathfrak{W}^{*,-\delta}(0)$ if such an element exists; else, let $\mathcal{C} = \{|\mathbb{J}|+1, \ldots, |\mathbb{J}|+d\}$ and $\Delta^C = \mathbf{1}_d$, and K, g and $\mathfrak{W}^{*,-\delta}$ are as defined in Lemma 14. Then,* *for any $\eta > 0$, there exist $0 < \varepsilon_\eta < \infty$ and $N \in \mathbb{N}$ s.t. $n \geq N$ implies*

$$P\left(\mathfrak{W}^*(-\delta) \neq \emptyset, \left|\det K^{\mathcal{C}}\right| \leq \varepsilon_\eta\right) \leq \eta \tag{273}$$

Proof. (273) can be bounded as follows:

$$\begin{aligned}
P\left(\mathfrak{W}^{*,-\delta}(0) \neq \emptyset, \left|\det K^{\mathcal{C}}\right| \leq \varepsilon_\eta\right) &\leq P\left(\exists C \in \widetilde{\mathcal{C}} : \Delta^C \in \mathbf{\Delta}, \left|\det K^C\right| \leq \varepsilon_\eta\right) \\
&\leq \sum_{C \in \widetilde{\mathcal{C}}: \left|\det K^C\right| \leq \varepsilon_\eta} \mathrm{P}\left(\Delta^C \in \mathbf{\Delta}\right) \\
&\leq \sum_{C \in \widetilde{\mathcal{C}}: \left|\alpha^C\right| \leq \varepsilon_\eta^{2/d}} P\left(\Delta^C \in \mathbf{\Delta}\right)
\end{aligned}$$

where α^C denote the smallest eigenvalue of $K^C K^{C\prime}$. Here, the first inequality holds because $\mathfrak{W}^{*,-\delta} \subseteq \mathbf{\Delta}$ and so the event in the first probability implies the event in the next one; the second inequality is Boolean algebra; the last inequality follows because $\left|\det K^C\right| \geq \left|\alpha^C\right|^{d/2}$. Noting that $\widetilde{\mathcal{C}}$ has $\begin{pmatrix} |\mathbb{J}| + 2d \\ d \end{pmatrix}$ elements, it suffices to show that

$$\left|\alpha^C\right| \leq \varepsilon_\eta^{2/d} \Longrightarrow P\left(\Delta^C \in \mathbf{\Delta}\right) \leq \bar{\eta} \equiv \frac{\eta}{\binom{|\mathbb{J}|+2d}{d}}.$$

Thus, fix $C \in \widetilde{\mathcal{C}}$. Let q^C denote the eigenvector associated with α^C and recall that because $K^C K^{C\prime}$ is symmetric, $\left\|q^C\right\| = 1$. Thus the claim is equivalent to:

$$\left|q^{C\prime} K^C K^{C\prime} q^C\right| \leq \varepsilon_\eta^{2/d} \Longrightarrow P\left(\left(K^C\right)^{-1} g^C \in \mathbf{\Delta}\right) \leq \bar{\eta}.$$

Now, if $\left|q^{C\prime} K^C K^{C\prime} q^C\right| \leq \varepsilon_\eta^{2/d}$ and $\left(K^C\right)^{-1} g^C \in \mathbf{\Delta}$, then the Cauchy-Schwarz inequality yields

$$\left|q^{C\prime} g^C\right| = \left|q^{C\prime} K^C \left(K^C\right)^{-1} g^C\right| < \sqrt{d}\varepsilon_\eta^{1/d},$$

hence

$$P\left(\left(K^C\right)^{-1} g^C \in \mathbf{\Delta}\right) \leq P\left(\left|q^{C\prime} g^C\right| < \sqrt{d}\varepsilon_\eta^{1/d}\right).$$

If q^C assigns non-zero weight only to non-stochastic constraints, the result follows immediately. If q^C assigns non-zero weight also to stochastic constraints, Assumption 9 (or 11) and Assumption 12 yield

$$\begin{aligned}
&\mathrm{eig}(\tilde{\Omega}) \geq \omega \\
\Longrightarrow\ &\mathrm{Var}_P\left(q^{C\prime} g^C\right) \geq \omega \\
\Longrightarrow\ P\left(\left|q^{C\prime} g^C\right| < \sqrt{d}\varepsilon_\eta^{1/d}\right) &= P\left(-\sqrt{d}\varepsilon_\eta^{1/d} < q^{C\prime} g^C < \sqrt{d}\varepsilon_\eta^{1/d}\right) \\
&< \frac{2\sqrt{d}\varepsilon_\eta^{1/d}}{\sqrt{2\omega\pi}},
\end{aligned} \tag{274}$$

where the result in (274) uses that the density of a normal r.v. is maximized at the expected value. The result follows by choosing $\varepsilon_\eta = \left(\frac{\bar{\eta}\sqrt{2\omega\pi}}{2\sqrt{d}}\right)^d$. □

Lemma 19. *Assumptions 7, 8, 6, 10 hold. In addition, suppose Assumption 9 or 11 hold. If* $|\mathbb{J}_1| \geq 1, |\mathbb{J}_2| \geq d$, *or if* $|\mathbb{J}_2| > d$, *then* $\exists \underline{c} > 0$ *s.t.*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\beta\in\mathcal{B}(P)} P\left(c_n^I(\beta) \geq \underline{c}\right) = 1.$$

Proof. I first consider the case where $\mathbb{J}_1 \neq \emptyset$ and $|\mathbb{J}_2| \geq d$. Fix any $c \geq 0$ and restrict attention to constraints $\{|\mathbb{J}_1|+1,\ldots,|\mathbb{J}_1|+d, |\mathbb{J}_1|+|\mathbb{J}_2|+1,\ldots,|\mathbb{J}_1|+|\mathbb{J}_2|+d\}$, i.e. the inequalities that jointly correspond to the first d equalities. We separately analyze the case when (i) the corresponding estimated gradients $\left\{\hat{D}_{n,j}(\beta) : j = |\mathbb{J}_1|+1,\ldots,|\mathbb{J}_1|+d\right\}$ are linearly independent and (ii) they are not. If $\left\{\hat{D}_{n,j}(\beta) : j = |\mathbb{J}_1|+1,\ldots,|\mathbb{J}_1|+d\right\}$ converge to linearly independent limits, then only the former case occurs infinitely often; else, both may occur infinitely often, and we conduct the argument along two separate subsequences if necessary.

For the remainder of this proof, because the sequence $\{\beta_n\}$ is fixed and plays no direct role in the proof, we suppress dependence of $\hat{D}_{n,j}(\beta)$ and $\mathbb{Z}^s_{n,j}(\beta)$ on β. Also, if C is an index set picking certain constraints, then $\hat{D}_n^C$ is the matrix collecting the corresponding estimated gradients, and similarly for $\mathbb{Z}_n^{s,C}$.

Suppose now case (i), then there exists an index set

$$\bar{C} \subset \{|\mathbb{J}_1|+1,\ldots,|\mathbb{J}_1|+d, |\mathbb{J}_1|+|\mathbb{J}_2|+1,\ldots,|\mathbb{J}_1|+|\mathbb{J}_2|+d\}$$

picking one direction of each constraint s.t. $\hat{D}_{n,1}$ is a positive linear combination of the rows of $\hat{D}_n^C$. (This choice ensures that a Karush-Kuhn-Tucker condition holds, justifying the step from (275) to (277) below.) Then the coverage probability $P^s\left(V_n^I(\beta,c) \neq \emptyset\right)$ is asymptotically bounded above by

$$\begin{aligned}
&P^s\left(\left\{\Delta \in \rho\boldsymbol{\Delta}_n : \hat{D}_{n,j}\Delta \leq c - \mathbb{Z}^s_{n,j}, j \in \mathbb{J}^*\right\} \neq \emptyset\right) \\
\leq &P^s\left(\left\{\Delta \in \mathbb{R}^d : \hat{D}_{n,j}\Delta \leq c - \mathbb{Z}^s_{n,j}, j \in \bar{C}\right\} \neq \emptyset\right) && (275)\\
= &P^s\left(\hat{D}'_{n,1}\left(\hat{D}_n^{\bar{C}}\right)^{-1}\left(c\mathbf{1}_d - \mathbb{Z}_n^{s,\bar{C}}\right) \leq c - \mathbb{Z}^b_{n,1}\right) && (276)\\
= &P^s\left(\left(1 - \hat{D}'_{n,1}\left(\hat{D}_n^{\bar{C}}\right)^{-1}\mathbf{1}_d\right)c \geq \mathbb{Z}^s_{n,1} - c\hat{D}'_{n,1}\left(\hat{D}_n^{\bar{C}}\right)^{-1}\mathbb{Z}_n^{s,\bar{C}}\right) \\
= &P^s\left(\frac{\left(1 - \hat{D}'_{n,1}\left(\hat{D}_n^{\bar{C}}\right)^{-1}\mathbf{1}_d\right)c}{\sqrt{\Omega_P^C}} \geq \frac{\mathbb{Z}^s_{n,1} - c\hat{D}'_{n,1}\left(\hat{D}_n^{\bar{C}}\right)^{-1}\mathbb{Z}_n^{s,\bar{C}}}{\sqrt{\Omega_P^C}}\right) && (277)\\
\leq &\Phi\left(\frac{\left(1 + \left|\hat{D}'_{n,1}\left(\hat{D}_n^{\bar{C}}\right)^{-1}\mathbf{1}_d\right|\right)c}{\sqrt{\Omega_P^C}}\right) + o_p(1) && (278)
\end{aligned}$$

Here, (275) removes constraints and hence enlarges the feasible set; (276) solves in closed form;

and (278) uses that by Assumption 9 or 11.

In case (ii), there exists an index set

$$\bar{C} \subset \{|\mathbb{J}_1|+2, \ldots, |\mathbb{J}_1|+d+1, |\mathbb{J}_1|+|\mathbb{J}_2|+2, \ldots, |\mathbb{J}_1|+|\mathbb{J}_2|+d+1\}$$

collecting d or fewer linearly independent constraints s.t. $\hat{D}_{n,|\mathbb{J}_1|+1}$ is a positive linear combination of the rows of $\hat{D}_P^{\bar{C}}$. (Note that $\bar{C}$ cannot contain $|\mathbb{J}_1|+1$ or $|\mathbb{J}_1|+|\mathbb{J}_2|+1$.) One can then write

$$\begin{aligned}
&P^s\left(\left\{\Delta \in \rho\boldsymbol{\Delta}_n : \hat{D}_{n,j}\Delta \leq c - \mathbb{Z}^s_{n,j}, j \in \mathbb{J}^*\right\} \neq \emptyset\right) && (279)\\
\leq & P^s\left(\exists \Delta : \hat{D}_{n,j}\Delta \leq c - \mathbb{Z}^s_{n,j}, j \in \bar{C} \cup \{|\mathbb{J}_1|+|\mathbb{J}_2|+1\}\right)\\
\leq & P^s\left(\sup_{\Delta}\left\{\hat{D}_{n,|\mathbb{J}_1|+1}\Delta : \hat{D}_{n,j}\Delta \leq c - \mathbb{Z}^s_{n,j}, j \in \bar{C}\right\}\right.\\
&\qquad \left.\geq \inf_{\Delta}\left\{\hat{D}_{n,|\mathbb{J}_1|+1}\Delta : \hat{D}_{n,|\mathbb{J}_1|+|\mathbb{J}_2|+1}\Delta \leq c - \mathbb{Z}^s_{n,|\mathbb{J}_1|+|\mathbb{J}_2|+1}\right\}\right) && (280)\\
= & P^s\left(\hat{D}_{n,|\mathbb{J}_1|+1}\hat{D}_n^{\bar{C}\prime}\left(\hat{D}_n^{\bar{C}}\hat{D}_n^{\bar{C}\prime}\right)^{-1}\left(c\mathbf{1}_{\bar{d}} - \mathbb{Z}_n^{s,\bar{C}}\right) \geq -c + \mathbb{Z}^s_{n,|\mathbb{J}_1|+|\mathbb{J}_2|+1}\right) && (281)
\end{aligned}$$

Here, the reasoning from (279) to (280) holds because we evaluate the probability of increasingly larger events; in particular, if the event in (280) fails, then the constraint sets corresponding to the sup and inf can be separated by a hyperplane with gradient $\hat{D}_{n,|\mathbb{J}_1|+1}$ and so cannot intersect. The last step solves the optimization problems in closed form, using (for the sup) that a KarushKuhn-Tucker condition again holds by construction and (for the inf) that $\hat{D}_{n,|\mathbb{J}_1|+|\mathbb{J}_2|+1} = -\hat{D}_{n,|\mathbb{J}_1|+1}$. Expression (281) resembles (277), and the argument can be concluded in analogy to Case (ii). □

Lemma 20. *Assumptions 7, 8, 6, 10 hold. In addition, suppose Assumption 9 or 11 hold. Suppose that both $\pi_{1,j}$ and $\pi_{1,j+R_1}$ are finite. Let (P_n, β_n) be the sequence satisfying the conditions of Lemma 15. Then for any $\beta_n' \in (\beta_n + \rho/\sqrt{n}\boldsymbol{\Delta}) \cap \overline{\mathcal{B}}$,*

1. $\sigma^2_{P_n,j}(\beta_n')/\sigma^2_{P_n,[j]}(\beta_n') \to 1$ *for* $j \in \mathbb{J}_2$.
2. $\frac{\Sigma_{P_n,j,[j]}}{\sqrt{\Sigma_{P_n,j}\Sigma_{P_n,[j]}}} \to -1$ *for* $j \in \mathbb{J}_2$.
3. $\left|\mathbb{Z}_{n,j}(\beta_n') + \mathbb{Z}_{n,[j]}(\beta_n')\right| \overset{P_n}{\to} 0$, *and* $\left|\mathbb{Z}^b_{n,j}(\beta_n') + \mathbb{Z}^b_{n,[j]}(\beta_n')\right| \overset{P^*}{\to} 0$ *for almost all* $\{X_i\}_{i=1}^\infty$.
4. $\left\|D_{P_n,[j]}(\beta_n') + D_{P_n,j}(\beta_n')\right\| \to 0$.

Proof. By Lemma 23, for each j, $\lim_{n\to\infty}\kappa_n^{-1}\frac{\sqrt{n}E_{P_n}[m_j(X_i;\beta_n',\varphi_{P_n})]}{\sigma_{P_n,j}(\beta_n')} = \pi_{1,j}$, and hence the condition that $\pi_{1,j}, \pi_{1,[j]}$ are finite is inherited by the limit of the corresponding sequences $\frac{\sqrt{n}E_{P_n}[m_j(X_i;\beta_n',\varphi_{P_n})]}{\kappa_n\sigma_{P_n,j}(\beta_n')}$ and $\frac{\sqrt{n}E_{P_n}[m_{[j]}(X_i;\beta_n',\varphi_{P_n})]}{\kappa_n\sigma_{P_n,[j]}(\beta_n')}$.

(1) $\pi_{1,j}$ being finite implies that $E_{P_n}m_j(X_i;\beta_n',\varphi_{P_n}) \to 0$. Thus by Assumption 11, $E_{P_n}(t_j(X_i,\beta_n')) \to 0$. We then have, using Assumption 11 again,

$$\text{Var}\left(t_j(X_i,\beta_n')\right) = \int t_j(x,\beta_n')^2\,dP_n(x) - E_{P_n}\left[t_j(X_i,\beta_n')\right]^2$$

$$\leq M \int t_j\left(x, \beta_n'\right) dP_n(x) - E_{P_n}\left[t_j\left(X_i, \beta_n'\right)\right]^2 \to 0. \tag{282}$$

Hence,

$$\begin{aligned}
&\Omega_{P_n,[j]}\left(\beta_n', \varphi_{P_n}\right) - \Omega_{P_n,j}\left(\beta_n', \varphi_{P_n}\right) \\
=& \operatorname{var}_{P_n}\left(t_j\left(X_i; \beta_n', \varphi_{P_n}\right)\right) + 2\operatorname{cov}_{P_n}\left(m_j\left(X_i; \beta_n', \varphi_{P_n}\right), t_j\left(X_i; \beta_n', \varphi_{P_n}\right)\right) \\
\leq& \operatorname{var}_{P_n}\left(t_j\left(X_i, \beta_n'\right)\right) + 2\left(\operatorname{var}_{P_n}\left(t_j\left(X_i, \beta_n'\right)\right)\right)^{1/2}\left(\operatorname{var}_{P_n}\left(m_j\left(X_i; \beta_n', \varphi_{P_n}\right)\right)\right)^{1/2} \to 0,
\end{aligned} \tag{283}$$

And similarly

$$\begin{aligned}
\Omega_{P_n,[j],\varphi}\left(\beta_n', \varphi_{P_n}\right) - \Omega_{P_n,j,\varphi}\left(\beta_n', \varphi_{P_n}\right) &= \operatorname{cov}\left(t_j\left(X_i; \beta_n', \varphi_{P_n}\right), m_\varphi(X_i)\right) \\
&\leq \sqrt{\operatorname{var}_{P_n}\left(t_j\left(X_i, \beta_n'\right)\right)\operatorname{var}_{P_n}\left(m_\varphi(X_i)\right)} \to 0.
\end{aligned}$$

Lastly, we show that

$$\nabla_\varphi E\left[m_j\left(X_i; \beta_n', \varphi_{P_n}\right)\right] + \nabla_\varphi E\left[m_{[j]}\left(X_i; \beta_n', \varphi_{P_n}\right)\right] = -\nabla_\varphi E\left[t_j(X_i; \beta_n', \varphi_{P_n})\right] \to 0.$$

Let

$$q_n = \nabla_\varphi E\left[t_j(X_i; \beta_n', \varphi_{P_n})\right]$$

and by contradiction, assume that $q_n \to q \neq 0$. Let $r_n = \frac{-q_n}{\|q_n\|}\frac{\kappa_n^2}{\sqrt{n}}$, and then

$$\begin{aligned}
E_{P_n}\left[t_j\left(X_i; \beta_n', \varphi_{P_n} + r_n\right)\right] =& E_{P_n}\left[t_j\left(X_i; \beta_n', \varphi_{P_n}\right)\right] + \nabla_\varphi E_{P_n}\left[t_j\left(X_i; \beta_n', \varphi_{P_n}\right)\right] r_n \\
&+ \left(\nabla_\varphi E_{P_n}\left[t_j\left(X_i; \beta_n', \bar{\varphi}\right)\right] - \nabla_\varphi E_{P_n}\left[t_j\left(X_i; \beta_n', \varphi_{P_n}\right)\right]\right) r_n \\
=& O\left(\frac{\kappa_n}{\sqrt{n}}\right) - \|q_n\|\frac{\kappa_n^2}{\sqrt{n}} + O\left(\frac{\kappa_n^4}{n}\right)
\end{aligned}$$

Therefore, we will have

$$E_{P_n}\left[t_j\left(X_i; \beta_n', \varphi_{P_n} + r_n\right)\right] < 0$$

for n large enough, which contradicts Assumption 11 that $t_j\left(X_i; \beta_n', \varphi_{P_n} + r_n\right) \geq 0$.

Then by

$$\sigma_{P_n,j}^2\left(\beta_n'\right) = \left[\begin{array}{cc} 1 & \nabla_{\varphi'} E\left[m_j\left(X_i; \beta_n', \varphi_{P_n}\right)\right] \end{array}\right]\left[\begin{array}{cc} \Omega_{P_n,j}\left(\beta_n'\right) & \Omega_{P_n,j,\varphi}\left(\beta_n'\right) \\ \Omega_{P_n,j,\varphi}\left(\beta_n'\right) & \Omega_{P_n\varphi}\left(\beta_n'\right) \end{array}\right]\left[\begin{array}{cc} 1 & \nabla_{\varphi'} E\left[m_j\left(X_i; \beta_n', \varphi_{P_n}\right)\right] \end{array}\right]',$$

we have

$$\sigma_{P_n,j}^2\left(\beta_n', \varphi_{P_n}\right) / \sigma_{P_n,j+R_1}^2\left(\beta_n', \varphi_{P_n}\right) \to 1.$$

(2) Note that

$$\begin{aligned}
\frac{\Sigma_{P_n,j,j+R_1}}{\sqrt{\Sigma_{P_n,j}\Sigma_{P_n,j+R_1}}} =& \frac{G_{P_n,j}(\beta_n', \varphi_{P_n})\Omega_{P_n}(\beta_n', \varphi_{P_n})G_{P_n,j+R_1}'}{\sqrt{G_{P_n,j}(\beta_n', \varphi_{P_n})\Omega_{P_n}(\beta_n', \varphi_{P_n})G_{P_n,j}'}\sqrt{G_{P_n,j+R_1}(\beta_n', \varphi_{P_n})\Omega_{P_n}(\beta_n', \varphi_{P_n})G_{P_n,j+R_1}'}} \\
&\to -1,
\end{aligned}$$

where the result follows from (282) and (283).

(3) Note that, for $j \in \mathbb{J}_2$,

$$
\begin{aligned}
&\mathbb{Z}_{n,j}\left(\beta_n', \varphi_{P_n}\right) + \mathbb{Z}_{n,[j]}\left(\beta_n', \varphi_{P_n}\right) \\
=&\frac{G_j \mathbb{G}_n(\beta_n', \varphi_{P_n})}{\sigma_{P_n,j}(\beta_n')} - \frac{\sigma_{P_n,j}(\beta_n')}{\sigma_{P_n,[j]}(\beta_n')} \frac{G_{[j]} \mathbb{G}_n(\beta_n', \varphi_{P_n})}{\sigma_{P_n,j}(\beta_n')} + \frac{(\hat{G}_j - G_j)\mathbb{G}_n(\beta_n', \varphi_{P_n})}{\sigma_{P_n,j}(\beta_n')} - \frac{(\hat{G}_{[j]} - G_{[j]})\mathbb{G}_n(\beta_n', \varphi_{P_n})}{\sigma_{P_n,[j]}\left(\beta_n'\right)} \\
=&o_p(1)
\end{aligned}
$$

By Lemma H.15 in Kaido et al. (2019), $\left\{\mathbb{G}_n^b\right\}$ converges in law to the same limit as $\{\mathbb{G}_n\}$ for almost all sample paths $\{X_i\}_{i=1}^{\infty}$. This then implies the second half of Claim 3.

(4) This is similar to (2). □

Lemma 21. *Assumptions 7, 8, 6, 10 hold. And suppose Assumption 9 or 11 hold. Then,*

1. *for each $j = 1, ..., J_1 + J_2, \ell, u$,*

$$\inf_{P \in \mathcal{P}} P\left(\sup_{(\beta,\Delta) \in \overline{\mathcal{B}} \times \boldsymbol{\Delta}} |\hat{\eta}_{n,j,\beta}(\Delta)| \to 0\right).$$

2. *Let (P_n, β_n) be a sequence such that $P_n \in \mathcal{P}, \beta_n \in \overline{\mathcal{B}}$ for all n, and $\kappa_n^{-1}\sqrt{n}\gamma_{1,P_n,j}(\beta_n) \to \pi_{1j} \in \mathbb{R}_{[-\infty]}$. Then, for any $\eta > 0$, there exists $N \in \mathbb{N}$ such that*

$$P_n\left(\max_{j \in \mathbb{J}^*}\left|\frac{\sigma_{P_n,j}(\beta_n)}{\hat{\sigma}_{n,j}^M(\beta_n)} - 1\right| > \eta\right) < \eta$$

for all $n \geq N$.

Proof. (1) First, for any $\epsilon > 0$ and for any $j = 1, ..., J_1 + J_2$, by Assumption 10, 12, Lemma D.2.2 in Bugni et al. (2015) and the argument in Lemma H.10(i) in KMS19 , there is n_1 such that

$$\sup_{P \in \mathcal{P}} P\left(\sup_{m \geq n_1} \sup_{(\beta,\varphi) \in \overline{\mathcal{B}} \times \Phi}\left\|\hat{\Omega}(\beta, \varphi) - \Omega_P(\beta, \varphi)\right\| \leq \epsilon\right) \to 0.$$

Next, note that by Assumption 10,

$$\sup_{P \in \mathcal{P}} P\left(\sup_{m \geq n} \|\hat{\varphi} - \varphi_P\| \leq \epsilon\right) \to 0.$$

Therefore, we have

$$
\begin{aligned}
&\sup_{P \in \mathcal{P}} P\left(\sup_{m \geq n} \sup_{\beta \in \overline{\mathcal{B}}}\left\|\hat{\Omega}(\beta) - \Omega_P(\beta)\right\| > \epsilon\right) \\
\leq &\sup_{P \in \mathcal{P}} P\left(\sup_{m \geq n} \sup_{\beta \in \overline{\mathcal{B}}}\left\|\hat{\Omega}(\beta, \hat{\varphi}) - \Omega_P(\beta, \hat{\varphi})\right\| > \frac{\epsilon}{2}\right) + \sup_{P \in \mathcal{P}} P\left(\sup_{m \geq n} \sup_{\beta \in \mathcal{B}} \|\Omega_P(\beta, \hat{\varphi}) - \Omega_P(\beta, \varphi_P)\| > \frac{\epsilon}{2}\right) \\
\leq &\sup_{P \in \mathcal{P}} P\left(\sup_{m \geq n} \sup_{(\beta,\varphi) \in \overline{\mathcal{B}} \times \Phi}\left\|\hat{\Omega}(\beta, \varphi) - \Omega_P(\beta, \varphi)\right\| > \frac{\epsilon}{2}\right) + \sup_{P \in \mathcal{P}} P\left(\sup_{m \geq n} M\|\hat{\varphi} - \varphi_P\| > \frac{\epsilon}{2}\right) \to 0.
\end{aligned}
$$

Together with Assumption 2 that $\hat{G}(\beta) - G_P(\beta)$ converges almost sure to 0, and Assumption 10.3, Assumption 7 that Σ_P and G_P are bounded, we can get

$$\inf_{P\in\mathcal{P}} P\left(\sup_{m\geq n}\sup_{(\beta,\varphi)\in\overline{\mathcal{B}}\times\Phi}\left\|\hat{\Sigma}(\beta)-\Sigma_P(\beta)\right\|\leq\epsilon\right)\to 1.$$

And since

$$\begin{aligned}\hat{\eta}_{n,j,\beta}(\Delta) &= \frac{\sigma_{P,j}(\beta)}{\hat{\sigma}_{n,j}(\beta+\frac{\Delta\rho}{\sqrt{n}})}-1\\ &= \frac{\sigma_{P,j}(\beta+\frac{\Delta\rho}{\sqrt{n}},\varphi_P)}{\hat{\sigma}_{n,j}(\beta+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi})}\frac{\sigma_{P,j}(\beta,\varphi_P)}{\sigma_{P,j}(\beta+\frac{\Delta\rho}{\sqrt{n}},\varphi_P)}-1\\ &= \eta_{1,n,j,\beta}(\Delta)\eta_{2,n,j,\beta}(\Delta)+\eta_{1,n,j,\beta}(\Delta)+\eta_{2,n,j,\beta}(\Delta)\end{aligned}$$

where

$$\begin{aligned}\eta_{1,n,j,\beta}(\Delta) &= \frac{\sigma_{P,j}(\beta+\frac{\Delta\rho}{\sqrt{n}},\varphi_P)}{\hat{\sigma}_{n,j}(\beta+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi})}-1,\\ \eta_{2,n,j,\beta}(\Delta) &= \frac{\sigma_{P,j}(\beta,\varphi_P)}{\sigma_{P,j}(\beta+\frac{\Delta\rho}{\sqrt{n}},\varphi_P)}-1.\end{aligned}$$

We can conclude that

$$\inf_{P\in\mathcal{P}} P\left(\sup_{m\geq n}\sup_{(\beta,\varphi)\in\overline{\mathcal{B}}\times\Phi}|\hat{\eta}_{n,j,\beta}(\Delta)-1|\leq\epsilon\right)\to 1.$$

Finally, note that for any $\epsilon>0$,

$$\begin{aligned}1 &= \lim_{n\to\infty}\inf_{P\in\mathcal{P}} P\left(\sup_{m\geq n}\sup_{(\beta,\varphi)\in\overline{\mathcal{B}}\times\Phi}|\hat{\eta}_{n,j,\beta}(\Delta)|\leq\epsilon\right)\\ &\leq \inf_{P\in\mathcal{P}}\lim_{n\to\infty} P\left(\bigcap_{m\geq n}\left\{\sup_{(\beta,\varphi)\in\overline{\mathcal{B}}\times\Phi}|\hat{\eta}_{n,j,\beta}(\Delta)|\leq\epsilon\right\}\right)\\ &= \inf_{P\in\mathcal{P}} P\left(\lim_{n\to\infty}\bigcap_{m\geq n}\left\{\sup_{(\beta,\varphi)\in\overline{\mathcal{B}}\times\Phi}|\hat{\eta}_{n,j,\beta}(\Delta)|\leq\epsilon\right\}\right)\\ &= \inf_{P\in\mathcal{P}} P\left(\sup_{(\beta,\varphi)\in\overline{\mathcal{B}}\times\Phi}|\hat{\eta}_{n,j,\beta}(\Delta)|\leq\epsilon,\ \text{for all but finite } n\right),\end{aligned}$$

where the second equality is due to the continuity of probability with respect to monotone sequences. Therefore, the first conclusion of the lemma follows.

(2) I first give the limit of $\hat{\mu}_{n,j}$. Recall the definitions of $\hat{\mu}_{n,[j]}$ and $\hat{\mu}_{n,j}$:

$$\hat{\mu}_{n,[j]}(\beta) = \min\left\{\max\left(0, \frac{\frac{\bar{m}_{n,j}(\beta,\hat{\varphi})}{\hat{\sigma}_{n,j}(\beta)}}{\frac{\bar{m}_{n,[j]}(\beta,\hat{\varphi})}{\hat{\sigma}_{n,[j]}(\beta)} + \frac{\bar{m}_{n,j}(\beta,\hat{\varphi})}{\hat{\sigma}_{n,j}(\beta)}}\right), 1\right\},$$
$$\hat{\mu}_{n,j}(\beta) = 1 - \hat{\mu}_{n,[j]}(\beta).$$

Note that

$$\begin{aligned}
&\sup_{\beta_n' \in \beta_n + \rho/\sqrt{n}\boldsymbol{\Delta}} \left| \kappa_n^{-1} \frac{\sqrt{n}\bar{m}_{n,j}(\beta_n', \hat{\varphi})}{\hat{\sigma}_{n,j}(\beta_n')} - \kappa_n^{-1} \frac{\sqrt{n} E_{P_n}[m_j(X_i;\beta_n',\varphi_{P_n})]}{\sigma_{P_n,j}(\beta_n')} \right| \\
\leq &\sup_{\beta_n' \in \beta_n + \rho/\sqrt{n}\boldsymbol{\Delta}} \left| \frac{\mathbb{G}_n(\beta_n',\hat{\varphi})}{\kappa_n \hat{\sigma}_{n,j}(\beta_n')} + \frac{\sqrt{n}}{\kappa_n}\gamma_{1,P,j}(\beta_n')\hat{\eta}_{n,j,\beta_n'} + \frac{\nabla_\varphi E_{P_n}[m_j(X_i;\beta_n',\bar{\varphi})]\sqrt{n}(\hat{\varphi}-\varphi)}{\kappa_n \hat{\sigma}_{n,j}(\beta_n')} \right| \\
= &o_P(1),
\end{aligned}$$

where the last equality follows from $\sup_{\beta\in\overline{\mathcal{B}}} |\mathbb{G}_n(\beta_n', \hat{\varphi})| = O_{\mathcal{P}}(1)$ due to asymptotic tightness of $\{\mathbb{G}_n\}$ (uniformly in P) by Lemma D.1 in Bugni et al. (2015), Theorem 3.6.1 and Lemma 1.3.8 in Van Der Vaart and Wellner (1996), and $\sup_{\beta\in\overline{\mathcal{B}}} |\hat{\eta}_{n,j,\beta}(\Delta)| = o_P(1)$ by part (i) of this Lemma. Hence,

$$\hat{\mu}_{n,j}(\beta_n) \xrightarrow{p} 1 - \min\left\{\max\left(0, \frac{\pi_{1,j}}{\pi_{1,[j]} + \pi_{1,j}}\right), 1\right\},$$

unless $\pi_{1,[j]} + \pi_{1,j} = 0$ (this case is considered later). This implies that if $\pi_{1,j} \in (-\infty, 0]$ and $\pi_{1,[j]} = -\infty$, one has

$$\hat{\mu}_{n,j}(\beta_n) \xrightarrow{p} 1.$$

Now, one may write

$$\begin{aligned}
\frac{\sigma_{P_n,j}(\beta_n)}{\hat{\sigma}^M_{n,j}(\beta_n)} - 1 =& \frac{\sigma_{P_n,j}(\beta_n)}{\hat{\sigma}_{n,j}(\beta_n)}\left(\frac{\hat{\sigma}_{n,j}(\beta_n)}{\hat{\sigma}^M_{n,j}(\beta_n)} - 1\right) + \left(\frac{\sigma_{P_n,j}(\beta_n)}{\hat{\sigma}_{n,j}(\beta_n)} - 1\right) \\
=& O_{P_n}(1)\left(\frac{\hat{\sigma}_{n,j}(\beta_n,\hat{\varphi})}{\hat{\sigma}^M_{n,j}(\beta_n,\hat{\varphi})} - 1\right) + o_p(1),
\end{aligned} \tag{284}$$

where the second equality follows from the first conclusion of the lemma. Hence, for the second conclusion of the lemma, it suffices to show $\hat{\sigma}_{n,j}(\beta_n)/\hat{\sigma}^M_{n,j}(\beta_n) - 1 = o_p(1)$. For this, consider two cases.

Case 1. $j \in (\mathbb{J}_2 \cup [\mathbb{J}_2]) \cap \mathbb{J}^*$ and $[j] \notin \mathbb{J}^*$. Then, $\pi^*_{1,j} = 0$ and $\pi^*_{1,[j]} = -\infty$ and

$$\begin{aligned}
\frac{\hat{\sigma}_{n,j}(\beta_n)}{\hat{\sigma}^M_{n,j}(\beta_n)} - 1 =& \frac{\hat{\sigma}_{n,j}(\beta_n) - \hat{\sigma}^M_{n,j}(\beta_n)}{\hat{\sigma}^M_{n,j}(\beta_n)} \\
=& \frac{(1-\hat{\mu}_{n,j}(\beta_n))\, O_{P_n}(\hat{\sigma}_{n,j}(\beta_n))}{(1+o_{P_n}(1))\,\hat{\sigma}_{n,j}(\beta_n) + (1-\hat{\mu}_{n,j}(\beta_n))\, O_{P_n}(\hat{\sigma}_{n,j}(\beta_n))} = o_p(1),
\end{aligned} \tag{285}$$

where we used $\hat{\sigma}^{-1}_{n,j}(\beta_n) = O_p(1)$ by Assumption 9 or 11 and part (i) of the lemma. By (284) and (285), $\sigma_{P_n,j}(\beta_n)/\hat{\sigma}^M_{n,j}(\beta_n) - 1 = o_p(1)$.

Case 2. $j \in \mathbb{J}^*$ and $[j] \in \mathbb{J}^*$. Then, $\pi^*_{1,j} = 0$ and $\pi^*_{1,[j]} = 0$. In this case, $\hat{\mu}_{n,j}(\beta_n) \in [0, 1]$ for all n and by Lemma 20 (1),

$$\left|\frac{\sigma_{P_n,j}(\beta_n)}{\sigma_{P_n,[j]}(\beta_n)}-1\right|=o_p(1), \tag{286}$$

for $j=\mathbb{J}_2\cup[\mathbb{J}_2]$. Therefore,

$$\begin{aligned}
&\frac{\sigma_{P_n,j}(\beta_n)}{\hat{\sigma}^M_{n,j}(\beta_n)}-1=\frac{\sigma_{P_n,j}(\beta_n)-\hat{\sigma}^M_{n,j}(\beta_n)}{\hat{\sigma}^M_{n,j}(\beta_n)}\\
&=\frac{[\hat{\mu}_{n,j}(\beta_n)+(1-\hat{\mu}_{n,j}(\beta_n))]\,\sigma_{P_n,j}(\beta_n)-\left[\hat{\mu}_{n,j}(\beta_n)\hat{\sigma}_{n,j}(\beta_n)+(1-\hat{\mu}_{n,j}(\beta_n))\hat{\sigma}_{n,[j]}(\beta_n)\right]}{\hat{\sigma}^M_{n,j}(\beta_n)}\\
&=\frac{\hat{\mu}_{n,j}(\beta_n)\left[\sigma_{P_n,j}(\beta_n)-\hat{\sigma}_{n,j}(\beta_n)\right]}{\hat{\sigma}^M_{n,j}(\beta_n)}+\frac{(1-\hat{\mu}_{n,j}(\beta_n))\left[\sigma_{P_n,[j]}(\beta_n)-\hat{\sigma}_{n,[j]}(\beta_n)+o_{P_n}(1)\right]}{\hat{\sigma}^M_{n,j}(\beta_n)},
\end{aligned} \tag{287}$$

where the second equality follows from the definition of $\hat{\sigma}^M_{n,j}(\beta_n)$, and the third equality follows from (286) and $\sigma_{P_n,[j]}$ bounded away from 0 due to Assumption 9 or 11. Note that

$$\frac{\hat{\mu}_{n,j}(\beta_n)\left[\sigma_{P_n,j}(\beta_n)-\hat{\sigma}_{n,j}(\beta_n)\right]}{\hat{\sigma}^M_{n,j}(\beta_n)}=\hat{\mu}_{n,j}(\beta_n)\frac{\hat{\sigma}_{n,j}(\beta_n)}{\hat{\sigma}^M_{n,j}(\beta_n)}\left(\frac{\sigma_{P_n,j}(\beta_n)}{\hat{\sigma}_{n,j}(\beta_n)}-1\right)=o_{P_n}(1),$$

where the second equality follows from the first conclusion of the lemma. Similarly,

$$\begin{aligned}
&\frac{(1-\hat{\mu}_{n,j}(\beta_n))\left[\sigma_{P_n,[j]}(\beta_n)-\hat{\sigma}_{n,[j]}(\beta_n)+o_{P_n}(1)\right]}{\hat{\sigma}^M_{n,j}(\beta_n)}\\
&=(1-\hat{\mu}_{n,j}(\beta_n))\frac{\hat{\sigma}_{n,[j]}(\beta_n)}{\hat{\sigma}^M_{n,j}(\beta_n)}\left(\frac{\sigma_{P_n,[j]}(\beta_n)}{\hat{\sigma}_{n,[j]}(\beta_n)}-1+o_{P_n}(1)\right)=o_{P_n}(1).
\end{aligned} \tag{288}$$

By (287)-(288), it follows that $\sigma_{P_n,j}(\beta_n)/\hat{\sigma}^M_{n,j}(\beta_n)-1=o_{P_n}(1)$. □

Lemma 22. $\mathbb{Z}_{n\beta}(\Delta)\xrightarrow{d}\mathbb{Z}$.

Lemma 23. *Suppose Assumptions 7, 8, 6, 10 hold. In addition, suppose Assumption 9 or 11 hold. Given a sequence*

$$\{Q_n,\beta_n\}\in\{(P,\beta):P\in\mathcal{P},\beta\in\mathcal{B}(P)\}$$

such that $\lim_{n\to\infty}\kappa_n^{-1}\sqrt{n}\gamma_{1,Q_n,j}(\beta_n)$ *exists for each* $j=0,1,\dots J$, *let* $\chi_j(\{Q_n,\beta_n\})$ *be a function of the sequence* $\{Q_n,\beta_n\}$ *defined as*

$$\chi_j(\{Q_n,\beta_n\})=\begin{cases}0, & \textit{if } \lim_{n\to\infty}\kappa_n^{-1}\sqrt{n}\gamma_{1,Q_n,j}(\beta_n)=0\\ -\infty, & \textit{if } \lim_{n\to\infty}\kappa_n^{-1}\sqrt{n}\gamma_{1,Q_n,j}(\beta_n)<0\end{cases} \tag{289}$$

Then for any $\beta_n'\in\beta_n+\frac{\rho}{\sqrt{n}}[-1,1]^d$ *for all* n, *one has:*

1. $\kappa_n^{-1}\sqrt{n}\gamma_{1,Q_n,j}(\beta_n)-\kappa_n^{-1}\sqrt{n}\gamma_{1,Q_n,j}(\beta_n')=o(1)$;
2. $\chi(\{Q_n,\beta_n\})=\chi(\{Q_n,\beta_n'\})=\pi^*_{1,j}$;
3. $\frac{\sqrt{n}\bar{m}_{n,j}(\beta_n',\hat{\varphi})}{\kappa_n\hat{\sigma}_{n,j}(\beta_n')}-\frac{\sqrt{n}E_{Q_n}\left[m_j(X_i,\beta_n',\varphi_{Q_n})\right]}{\kappa_n\sigma_{P_n,j}(\beta_n')}=o_P(1)$.

Proof. For (i), the mean value theorem yields

$$
\begin{aligned}
&\sup_{P\in\mathcal{P}}\sup_{\beta\in\mathcal{B}(P),\tilde\beta\in\beta+\frac{\rho}{\sqrt{n}}\mathbf{\Delta}}\left|\frac{\sqrt{n}E_P\left[m_j(X;\beta,\varphi_P)\right]}{\kappa_n\sigma_{P,j}(\beta)}-\frac{\sqrt{n}E_P\left[m_j(X;\tilde\beta,\varphi_P)\right]}{\kappa_n\sigma_{P,j}(\tilde\beta)}\right|\\
\le&\sup_{P\in\mathcal{P}}\sup_{\beta\in\mathcal{B}(P),\tilde\beta\in\beta+\frac{\rho}{\sqrt{n}}\mathbf{\Delta}}\left|\frac{\sqrt{n}\left(E_P\left[m_j(X;\beta,\varphi_P)\right]-E_P\left[m_j(X;\tilde\beta,\varphi_P)\right]\right)}{\kappa_n\sigma_{P,j}(\beta)}\right|\\
&+\sup_{P\in\mathcal{P}}\sup_{\beta\in\mathcal{B}(P),\tilde\beta\in\beta+\frac{\rho}{\sqrt{n}}\mathbf{\Delta}}\left|E\left[m_j(X;\tilde\beta,\varphi_P)\right]\frac{\sqrt{n}\left(\sigma_{P,j}(\beta)-\sigma_{P,j}(\tilde\beta)\right)}{\kappa_n\sigma_{P,j}(\beta)\sigma_{P,j}(\tilde\beta)}\right|\\
\le&\sup_{P\in\mathcal{P}}\sup_{\beta\in\mathcal{B}(P),\tilde\beta\in\beta+\frac{\rho}{\sqrt{n}}\mathbf{\Delta}}\frac{\left\|D_{P,j}(\bar\beta)\right\|\sqrt{n}\left\|\tilde\beta-\beta\right\|}{\kappa_n}+\sup_{P\in\mathcal{P}}\sup_{\beta\in\mathcal{B}(P),\tilde\beta\in\beta+\frac{\rho}{\sqrt{n}}\mathbf{\Delta}}\frac{M^2\sqrt{n}\left\|\tilde\beta-\beta\right\|}{\kappa_n\sigma_{P,j}(\beta)\sigma_{P,j}(\tilde\beta)}\\
=&o(1),
\end{aligned}
$$

where $\bar\beta$ represents a mean value that lies componentwise between β and $\tilde\beta$ and where we used the fact that $\sup_{P\in\mathcal{P}}\sup_{\beta\in\mathcal{B}(P)}\|D_{P,j}(\beta)\|\le\bar M$, $\sqrt{n}\left\|\tilde\beta-\beta\right\|\le\rho$ and $\sigma_{P,j}(\beta)\in[\varepsilon,\ \frac{1}{\varepsilon}]$. And it is easy to show this for γ_0 using Chain rule and similar arguments.

Result (ii) then follows immediately from (289).

For (iii), note that by (156), we have

$$
\begin{aligned}
&\sup_{\tilde\beta\in\beta_n+\rho/\sqrt{n}\mathbf{\Delta}}\left|\kappa_n^{-1}\frac{\sqrt{n}\bar m_{n,j}\left(\tilde\beta,\hat\varphi\right)}{\hat\sigma_{n,j}\left(\tilde\beta\right)}-\kappa_n^{-1}\frac{\sqrt{n}E_{Q_n}\left[m_j\left(X_i,\tilde\beta,\varphi_{Q_n}\right)\right]}{\sigma_{Q_n,j}\left(\tilde\beta\right)}\right|\\
\le&\sup_{\tilde\beta\in\beta_n+\rho/\sqrt{n}\mathbf{\Delta}}\frac{\sqrt{n}\gamma_{1,P_n,j}\left(\beta_n\right)}{\kappa_n}\hat\eta_{n,j,\beta_n}(\Delta)\\
&\quad+\left(1+\hat\eta_{n,j,\beta_n}(\Delta)\right)\left(\frac{\mathbb{Z}_{n,j,\beta_n}}{\kappa_n}+\frac{\sigma_{Q_n,j}\left(\bar\beta_n\right)}{\sigma_{Q_n,j}\left(\beta_n\right)}\frac{D_{P_n,j}\left(\bar\beta_n\right)\Delta\rho}{\kappa_n}\right)\\
=&o_P(1),
\end{aligned}
$$

where the last equality follows from $\sup_{\beta\in\beta}|\mathbb{Z}_{n,\beta}|=O_P(1)$ due to asymptotic tightness of $\{\mathbb{Z}_n\}$ (uniformly in P) by Lemma D.1 in Bugni et al. (2015), Theorem 3.6.1 and Lemma 1.3.8 in Van Der Vaart and Wellner (1996), and $\sup_{\beta\in\beta}|\eta_{n,j}(\beta)|=o_P(1)$ by Lemma 21(i). □