Essays on Inference in Non-standard Situations

by

Xinyue Bei

Department of Economics
Duke University

Defense Date: March 15, 2024

Approved:

Adam Rosen, Supervisor

Federico Bugni

Matt Masten

Enrique Sentana

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2024

ABSTRACT

Essays on Inference in Non-standard Situations

by

Xinyue Bei

Department of Economics
Duke University

Defense Date: March 15, 2024

Approved:

Adam Rosen, Supervisor

Federico Bugni

Matt Masten

Enrique Sentana

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics in the Graduate School of Duke University
2024

Abstract

This dissertation consists of three essays on inference in non-standard scenarios. The first essay investigates inference on union bounds, prevalent in various empirical contexts such as relaxations of the difference-in-differences assumption and counterfactual analysis with partially identified structural parameters. I propose novel methods for both finite and infinite sets of bounds.

The second essay presents a bootstrap-based profiling inference technique for subvectors in moment inequality models. By refining critical value calculation and simplifying non-linear moment conditions, this method offers improved computational efficiency upon existing methods.

The third essay, conducted jointly with Dante Amengual and Enrique Sentana, extends score-type tests in likelihood contexts in which the nullity of the information matrix under the null is greater than one. Our approach, leveraging higher-order derivatives, is asymptotically equivalent to likelihood ratio tests while requiring estimation only under the null.

Contents

List of Tables

List of Figures

Acknowledgements

I am very fortunate to have Adam Rosen as my advisor who has been giving me invaluable guidance, encouragement, and support throughout the journey in the past six years. Adam opened the door to the world of partial identification for me and guided me along the road to explore the unknown realm. I started to work with Adam when I was writing my second-year paper. Adam instructed me to carve and polish my first independent piece of work and it wouldn't end up as a publication without his help. His insights, advice, and comments are also indispensable building blocks of my job market paper.

I am also very grateful to my committee members Federico Bugni, Matt Masten, and Enrique Sentana. Federico's sharp insights and expertise in partial identification always lightened the correct way to go amid the mist of uncertainty. His advice has saved me a gigantic amount of time for the right work.

I benefited a lot from the regular group meetings with Matt who widened my horizon in econometrics. I came up with many of the empirical examples in my job market paper from those meetings. The paper won't be as appreciated by economists from other fields without Matt's knowledge. Matt also shaped me into a more organized scholar and set up the milestones for me along this journey, ensuring that I always sailed at the right pace.

I won't be as productive had I not met Enrique Sentana and Dante Amengual. I am honored to be their student, research assistant, and coauthor on five research projects in the past eight years. They sent me to Duke and paved a smooth way for me to transition from a student to a junior scholar. This fantastic journey wouldn't happen without them and wouldn't be this fantastic without them.

I am also very thankful to other faculty members at Duke econ. I learned a lot from their classes and benefited a lot from their comments and discussions. It's my great honor to have studied at Duke econ.

The six years of PhD life have been full of joy with my amazing cohorts and friends. I cherish every minute we have spent together, studying, working, eating, and playing. I particularly want to thank my boyfriend Lin Zhao for his love and support in my life.

Finally, I want to thank my family for always supporting my decisions and being my strong backing.

1. Introduction

This dissertation consists of three essays on inference in non-standard situations. The first essay studies inference on union bounds. 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. 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 et al. (2022) and the counterfactual analysis in Dickstein and Morales (2018).

The second essay introduces a bootstrap-based profiling inference method for subvectors in moment inequality models following insights from Bugni et al. (2017). Compared to their paper, the new method calculates the critical value by searching over a local neighborhood of a pre-estimator, instead of the whole null parameter space, to profile out nuisance parameters. In this way, non-linear moment conditions are simplified by linear expansion and the bootstrap iterates over quadratic programming problems, which significantly simplifies and accelerates computation. This method controls asymptotic size uniformly over a large class of data generating processes. In the Monte Carlo simulations, the new procedure improves upon the computing time of Bugni et al. (2017) and Kaido et al. (2019) significantly. I provide an empirical illustration estimating an airline entry game. [1]

The third essay is a joint work with Dante Amengual and Enrique Sentana. We study

[1] This essay was originally published in the Journal of Econometrics, Volume 238, Issue 1, Pages 105594, [2024]

score-type tests in likelihood contexts in which the nullity of the information matrix under the null is greater than one, thereby generalizing existing results in the literature. Examples include multivariate regressions with sample selectivity, semi-nonparametric distributions, Hermite expansions of Gaussian copulas, and purely non-linear predictive regressions among others. Our proposal, which involves higher-order derivatives, is asymptotically equivalent to the likelihood ratio test but only requires estimation under the null, a substantial advantage for resampling-based inference. We conduct extensive Monte Carlo exercises to study the finite sample size and power properties of our proposal, comparing it to alternative approaches.

2. Inference on Union Bounds

In this essay, I provide the first general and systematic study of inference on union bounds.

2.1 Introduction

This essay 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 2.2.

In this essay, 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 J. L. 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 et al. (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 2.3, 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 2.5, 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 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.

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

In Section 2.4, 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 et al. (2019), where a subvector is a known function, usually a single element, of the structural parameter. Different from the subvector inference procedure, in this essay (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 2.5 confirm good size and power properties.

In Section 2.6, 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 et al. (2012), Kolesár and Rothe (2018), Hasegawa et al. (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 J. L. 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 et al. (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 essay 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 et al. (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 et al. (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 et al. (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 essay studies a specific non-convex null space, but the modified conditional procedure is potentially applicable to more general settings.

My essay widens the use of the conditional inference technique. There is a growing literature on conditional inference, see, e.g. M. J. Moreira (2003), Kleibergen (2005), I. Andrews and Mikusheva (2016), I. Andrews et al. (2019), I. Andrews et al. (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.

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 et al. (2017), Bei (2024), Chernozhukov et al. (2023), Chen et al. (2018), G. 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.

[2] Specifically, the space of λ_ℓ and λ_u under (3.2) is not convex.

Hence, it is important to extend the previous work to construct the confidence intervals for counterfactual outcomes. In this essay, 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.

This essay is also related to the literature on counterfactual analysis and marginal treatment effect models, but my procedure applies to general moment inequality models. Kalouptsidi et al. (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 et al. (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 et al. (2018).

2.2 Setup and Examples

2.2.1 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 essay is to construct a uniformly valid confidence interval for the target object θ, whose identified set is characterized as a union bound

$$\theta \in \bigcup_{\beta \in \mathcal{B}} [\lambda_\ell(\beta),\ \lambda_u(\beta)]\,. \tag{2.1}$$

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

$$\theta \in \left[\inf_{\beta \in \mathcal{B}} \lambda_\ell(\beta),\ \sup_{\beta \in \mathcal{B}} \lambda_u(\beta)\right]. \tag{2.2}$$

In this essay, 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]

2.2.2 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*

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

panel data model $t = -\underline{T}, ..., 1$. *Let* $\gamma \in \mathbb{R}^{\underline{T}+1}$ *be a vector of "event study" coefficients, which can be decomposed as*

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

The target object θ *is the average treatment effect on the treated, and* ξ *is a bias from a difference in trend. Here* θ *and* ξ^{post} *are scalar,* $\xi^{pre} = \left(\xi^{pre}_{-\underline{T}}, ..., \xi^{pre}_{-1}\right)$ *and* $\gamma_0 = \xi^{pre}_0$ *is normalized to zero. Under parallel trends,* $\left(\xi^{pre}, \xi^{post}\right) = 0$ *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*

$$\left|\xi^{post} - 0\right| \leqslant M \max_{t=-1,...,-\underline{T}} \left|\xi^{pre}_{t+1} - \xi^{pre}_t\right|, \tag{2.3}$$

where $M \geqslant 0$ *is the degree of relaxation specified by the researcher. Manski and Pepper (2018) implement a similar concept with a natural benchmark* $M = 1$ *(see their Table 3). Under (2.3), the identified set of* θ *is a union bound in (2.2) with* $\mathcal{B} = \{1, ..., 2\underline{T}\}$,

$$\lambda_\ell(\beta) = \lambda_u(\beta) = \begin{cases} \gamma^{post} + M\left(\gamma^{pre}_{-\beta+1} - \gamma^{pre}_{-\beta}\right) & \text{if } \beta = 1, ..., \underline{T}, \\ \gamma^{post} - M\left(\gamma^{pre}_{\underline{T}-\beta+1} - \gamma^{pre}_{\underline{T}-\beta}\right) & \text{if } \beta = \underline{T}+1, ..., 2\underline{T}. \end{cases} \tag{2.4}$$

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 et al. (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 2.1a. 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 2.1b. 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

$$\sigma_\ell \min\{g(\xi_1), g(\xi_2)\} \leqslant g(\xi) \leqslant \sigma_u \max\{g(\xi_1), g(\xi_2)\} \text{ for } \xi \in [\xi_1, \xi_2] \tag{2.5}$$

for some $\sigma_u \geqslant 1 \geqslant \sigma_\ell > 0$, 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 $\mathcal{B} = \{1, 2\}$,

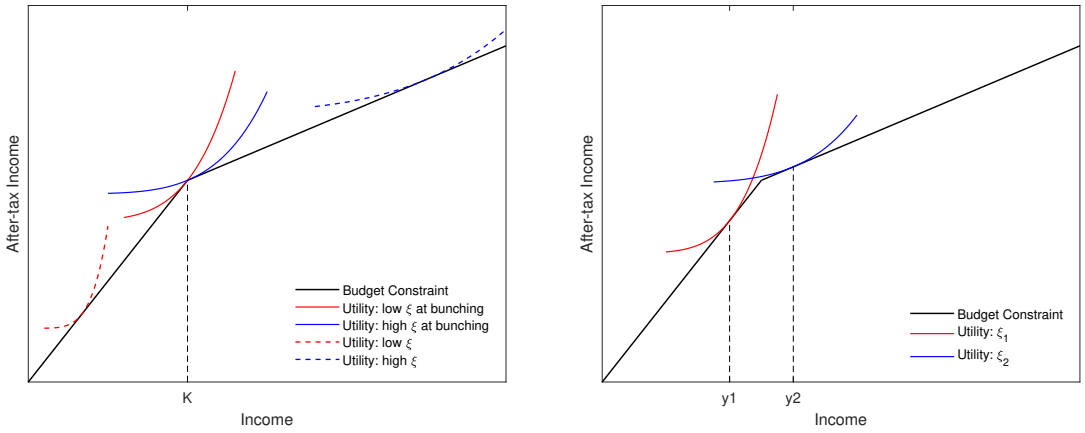
$$\lambda_\ell(1) = \frac{\log\left(\frac{y_1}{y_2} + \frac{P(y_1 \leqslant Y \leqslant y_2)}{f^-(y_1)\sigma_u y_2}\right)}{\log \rho_1 - \log \rho_2}, \qquad \lambda_\ell(2) = \frac{-\log\left(\frac{y_2}{y_1} - \frac{P(y_1 \leqslant Y \leqslant y_2)}{f^+(y_2)\sigma_u y_1}\right)}{\log \rho_1 - \log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

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

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

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. □

Example 3. ***(Regression Discontinuity Design).*** *Kolesár and Rothe (2018) study inference in regression discontinuity designs with a discrete running variable. Let $D = \mathbf{1}[X \geqslant 0]$ 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*



(a) Budget Constraint and Utility (b) Bunching Interval

FIGURE 2.1: Example 2 Bunching and Taxable Income Elasticity

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*

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

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.*

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

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 (2.2) 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 $cov(Z, U) = 0$, (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 $\xi \geqslant 0$ measures the level of relaxation. The corresponding identified set of θ accounting for the relaxation by ξ is

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

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 (2.2), where

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

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

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

In their empirical application, the authors implicitly assume that either $\pi_\beta = 0$ or $|\pi_\beta| \geqslant \varepsilon > 0$ 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 (2020a) 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 (2020a) suggests reporting the misspecification robust identified set

$$[\theta_L, \theta_U] \cup \left\{ \frac{\sigma_U \theta_L + \sigma_L \theta_U}{\sigma_L + \sigma_U} \right\}, \tag{2.6}$$

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 (2.2) 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 (2020a) proposes a CI for (2.6), 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

$$d_{ij}\left(\beta_3^{-1}E\left[r_{ij}\,|\mathcal{J}_{ij}\right]-\beta_3^{-1}\beta_1-\beta_3^{-1}\beta_2 dist_j-v_{ij}\right)\geqslant 0. \tag{2.7}$$

Based on (2.7), 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 (2.1). Further details of the moment conditions and counterfactuals are given in Appendix A.2.2.

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 et al. (2016), Bombardini et al. (2023), Crawford and Yurukoglu (2012), Ciliberto et al. (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*

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

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

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

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

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 (2.1). □

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 2.4 applies. □

2.3 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 2.2.2. A similar inference procedure applies to general, potentially non-connected, union bounds with the form (2.1).

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

$$\begin{pmatrix} \hat{\lambda}_\ell \\ \hat{\lambda}_u \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \lambda_\ell \\ \lambda_u \end{pmatrix}, \Sigma_n\right), \ \Sigma_n = \begin{bmatrix} \Sigma_{\ell,n} & \Sigma'_{\ell u,n} \\ \Sigma_{\ell u,n} & \Sigma_{u,n} \end{bmatrix} \tag{2.8}$$

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 2.3.4.

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

$$H_0 : \min_{b \in \mathcal{B}} \lambda_{\ell,b} \leqslant \theta \leqslant \max_{b \in \mathcal{B}} \lambda_{u,b}. \tag{2.9}$$

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

$$CI^{\mathrm{m}}(\hat{\lambda}_n, \Sigma_n; \alpha) = \left[\inf_{\phi(\theta, \hat{\lambda}_n, \Sigma_n) = 0} \theta, \sup_{\phi(\theta, \hat{\lambda}_n, \Sigma_n) = 0} \theta\right]. \tag{2.10}$$

2.3.1 The Test Statistic

The test statistic has a max-min form

$$\hat{T}(\theta) = \max\left\{\min_{b \in \mathcal{B}} \mathcal{Z}_{\ell,b}, \ \min_{b \in \mathcal{B}} \mathcal{Z}_{u,b}\right\} \tag{2.11}$$

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

$$\mathcal{Z}_{\ell,b} = \frac{\hat{\lambda}_{\ell,b} - \theta}{\sigma_{\ell,b}},\ \text{and}\ \mathcal{Z}_{u,b} = \frac{\theta - \hat{\lambda}_{u,b}}{\sigma_{u,b}}. \tag{2.12}$$

Observing that H_0 in (3.2) is equivalent to

$$H_0 : \max\left\{\min_{b\in\mathcal{B}}(\lambda_{\ell,b} - \theta),\ \min_{b\in\mathcal{B}}(\theta - \lambda_{u,b})\right\} \leqslant 0, \tag{2.13}$$

and the test statistic is constructed by replacing λ_ℓ and λ_u in (2.13) 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

$$CI^{\text{sim}} = \left[\min_{b\in\mathcal{B}} \hat{\lambda}_{\ell,b} - \sigma_{\ell,b}\Phi^{-1}(1 - \frac{\alpha}{2}),\ \max_{b\in\mathcal{B}} \hat{\lambda}_{u,b} + \sigma_{u,b}\Phi^{-1}(1 - \frac{\alpha}{2})\right], \tag{2.14}$$

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

$$b_\ell = \arg\min_{b\in\mathcal{B}} \lambda_{\ell,b}, \quad b_u = \arg\max_{b\in\mathcal{B}} \lambda_{u,b}. \tag{2.15}$$

and observe that

$$\begin{aligned}
P\left(\theta \notin CI^{\text{sim}}\right) &= P\left(\max\left\{\min_{b\in\mathcal{B}} \mathcal{Z}_{\ell,b},\ \min_{b\in\mathcal{B}} \mathcal{Z}_{u,b}\right\} > \Phi^{-1}(1 - \frac{\alpha}{2})\right) \\
&\leqslant P\left(\max\left\{\mathcal{Z}_{\ell,b_\ell},\ \mathcal{Z}_{u,b_u}\right\} > \Phi^{-1}(1 - \frac{\alpha}{2})\right) \\
&\leqslant P\left(\mathcal{Z}_{\ell,b_\ell} > \Phi^{-1}(1 - \frac{\alpha}{2})\right) + P\left(\mathcal{Z}_{u,b_u} > \Phi^{-1}(1 - \frac{\alpha}{2})\right) \\
&= P\left(\frac{\hat{\lambda}_{\ell,b_\ell} - \lambda_{\ell,b_\ell}}{\sigma_{\ell,b}} + \frac{\lambda_{\ell,b_\ell} - \theta}{\sigma_{\ell,b}} > \Phi^{-1}(1 - \frac{\alpha}{2})\right) \qquad (2.16) \\
&\quad + P\left(\frac{\lambda_{u,b_u} - \hat{\lambda}_{u,b_u}}{\sigma_{u,b}} + \frac{\theta - \lambda_{u,b_u}}{\sigma_{u,b}} > \Phi^{-1}(1 - \frac{\alpha}{2})\right) \qquad (2.17)
\end{aligned}$$

$$\leqslant \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.$$

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 $P(A \cup B) \leqslant P(A)+P(B)$. The final inequality holds under the null hypothesis (3.2).

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.

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

In such cases, either (2.16) or (2.17) 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.

$$\min_{b\in\mathcal{B}\setminus b_\ell} \frac{\lambda_{\ell,b} - \lambda_{\ell,b_\ell}}{\sigma_{\ell,b}} \gg 0, \quad \min_{b\in\mathcal{B}\setminus b_u} \frac{\lambda_{u,b_u} - \lambda_{u,b}}{\sigma_{u,b}} \gg 0, \quad \frac{\lambda_{u,b_u} - \lambda_{\ell,b_\ell}}{\min\{\sigma_{\ell,b_\ell}, \sigma_{u,b_u}\}} \approx 0. \tag{2.18}$$

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 (2.18) but is more efficient under other DGPs.

2.3.2 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 (2.18),

$$P\left(E_\ell \cup E_u\right) \approx 1 \tag{2.19}$$

where[5]

$$\begin{aligned} E_\ell &= \left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,\hat{b}_\ell}\right\} \cap \left\{\lambda_{\ell,\hat{b}_\ell} \leqslant \theta\right\}, \\ E_u &= \left\{\hat{T}(\theta) = \mathcal{Z}_{u,\hat{b}_u}\right\} \cap \left\{\lambda_{u,\hat{b}_u} \geqslant \theta\right\}, \\ \hat{b}_\ell &= \underset{b\in\mathcal{B}_\ell}{\arg\min}\mathcal{Z}_{\ell,b},\ \hat{b}_u = \underset{b\in\mathcal{B}_u}{\arg\min}\mathcal{Z}_{u,b}. \end{aligned} \tag{2.20}$$

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

$$P\left(\hat{T}(\theta) > \hat{c}(\theta)\,|E_\ell \cup E_u\right) \leqslant \alpha^{\mathrm{c}} < \alpha, \tag{2.21}$$

the unconditional rejection rate is bounded above by α following from (2.19). Therefore, I construct a conditional critical value based on the conditional distribution

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

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

Lemma 8. *Under H_0 and (2.8). Let b_1, b_2 satisfy $\lambda_{\ell,b_1} \leqslant \theta \leqslant \lambda_{u,b_2}$, then*

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

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

[5] 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 $\hat{b}_u$.

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 (2.12).

Loosely speaking, Lemma 8 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:

$$\hat{c}^{\mathrm{c}}(\theta,\alpha^{\mathrm{c}})=\begin{cases}\Phi^{-1}\left(\alpha^{\mathrm{c}}\Phi\left(t_{\ell,1}(\theta,\hat{b}_\ell)\right)+(1-\alpha^{\mathrm{c}})\Phi\left(t_{\ell,2}(\theta,\hat{b}_\ell)\right)\right) & \text{if } \mathcal{Z}_{\ell,\hat{b}_\ell}\geqslant\mathcal{Z}_{u,\hat{b}_u}\\ \Phi^{-1}\left(\alpha^{\mathrm{c}}\Phi\left(t_{u,1}(\theta,\hat{b}_u)\right)+(1-\alpha^{\mathrm{c}})\Phi\left(t_{u,2}(\theta,\hat{b}_u)\right)\right) & \text{if } \mathcal{Z}_{\ell,\hat{b}_\ell}<\mathcal{Z}_{u,\hat{b}_u}\end{cases}\tag{2.22}$$

where $\alpha^{\mathrm{c}}\in(\frac{1}{2}\alpha,\alpha)$ 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 9. *Assume that*

$$P\left(\mathcal{Z}_{\ell,\hat{b}_\ell} = \mathcal{Z}_{u,\hat{b}_u}\right) = 0. \tag{2.23}$$

Under H_0 and (2.8), it holds that

$$P\left(\hat{T}(\theta) > \hat{c}^c(\theta,\alpha^c)\middle| E_\ell \cup E_u\right) \leqslant \alpha^c. \tag{2.24}$$

Under (2.23), 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 $\lambda_{b_1} \leqslant \theta \leqslant \lambda_{b_2}$. Hence (2.24) follows directly from Lemma 8. Condition (2.23) holds in most examples previously discussed and is assumed in Proposition 9 for simplicity. Under more favorable DGPs diverging from (2.18), 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

$$\begin{aligned} t_{\ell,1}(\theta,\hat{b}_\ell) &\leqslant \left(1+\rho_{\ell u}(\hat{b}_\ell,b_u)\right)^{-1}\left(\mathcal{Z}_{u,b_u} + \rho_{\ell u}(\hat{b}_\ell,b_u)\mathcal{Z}_{\ell,\hat{b}_\ell}\right) \\ &= \left(1+\rho_{\ell u}(\hat{b}_\ell,b_u)\right)^{-1}\left(\frac{\hat{\lambda}_{\ell,\hat{b}_\ell} - \hat{\lambda}_{u,b_u}}{\sigma_{u,b_u}} + \left(\rho_{\ell u}(\hat{b}_\ell,b_u) - \frac{\sigma_{\ell,\hat{b}_\ell}}{\sigma_{u,b_u}}\right)\mathcal{Z}_{\ell,\hat{b}_\ell}\right) \approx -\infty, \end{aligned} \tag{2.25}$$

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,

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

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

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

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 10. *(Simple Union Bounds) Consider a simple union bound*

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

and the estimator satisfies

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

The test statistic has the form

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

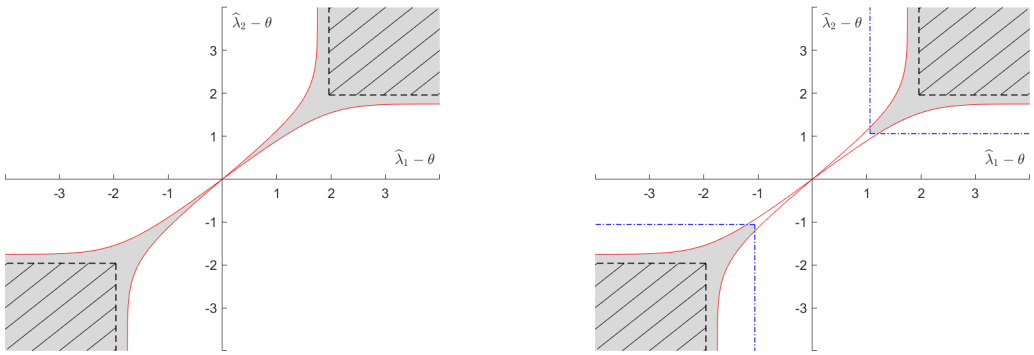
Without loss of generality, assume that $\hat{T}(\theta) = \hat{\lambda}_1 - \theta$*. In this case, the conditional critical value is*

$$\begin{aligned}\hat{c}^c(\theta; \alpha^c) &= \Phi^{-1}\left((1-\alpha^c)\Phi\left(\hat{\lambda}_2 - \theta\right) + \alpha^c \Phi\left(\theta - \hat{\lambda}_2\right)\right) \\ &< \Phi^{-1}(1-\alpha^c) < \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \end{aligned} \tag{2.26}$$

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* $\hat{c}^c(\theta; \alpha^c) \to \Phi^{-1}(1-\alpha^c)$*, 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 2.2a, I plot the rejection region for the simple and conditional critical values with* $\alpha = 0.05$ *and* $\alpha^c = 0.04$*. 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.* □



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

FIGURE 2.2: Example 10 - 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 $\alpha^{\mathrm{c}} = 0.04$.

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 (2.18). 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.

2.3.3 The Modified Conditional Critical Value

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

$$\hat{c}^{\mathrm{m}}(\theta;\alpha) = \tilde{c}^{\mathrm{m}}(\theta,\hat{c}^{\mathrm{t}};\alpha) = \max\left\{\hat{c}^{\mathrm{c}}(\theta,\alpha^{\mathrm{c}}),\hat{c}^{\mathrm{t}}\right\} \tag{2.27}$$

where $\hat{c}^{\mathrm{t}}$ is defined later in (2.31).

To illustrate the construction of the lower truncation, let $\widetilde{CI}^{\mathrm{m}}(c)$ be the confidence interval based on (2.10) 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.

$$c^{\mathrm{t}}(\theta) = \inf\left\{ c \in \mathbb{R}_+ : \sup_{\lambda \in \Lambda_0(\theta)} p(c;\theta,\lambda) \leqslant \alpha \right\}, \tag{2.28}$$

where Λ_0 is the set of feasible λ satisfying H_0:

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

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

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

from the discussion in Section 2.3.1. In fact $\hat{c}^{\mathrm{t}}(\theta)$ is usually significantly smaller than c^{sim}. The intuition is that by virtue of Lemma 9, truncation is unnecessary for DGPs such that

$$P\left(E_\ell \cup E_u; \lambda\right) \geqslant \frac{1-\alpha}{1-\alpha^c} \tag{2.29}$$

with $\alpha^{\mathrm{c}} < \alpha$, where E_ℓ, E_u are defined in (2.20). Thus we only need to consider truncation in more favorable DGPs deviating from (2.18), 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

$$\begin{aligned}
p(c;\theta,\lambda) \leqslant & \max\left\{ P\left([\theta_\ell, \theta_m] \nsubseteq \widetilde{CI}^{\mathrm{m}}(c); \lambda \right), P\left([\theta_m, \theta_u] \nsubseteq \widetilde{CI}^{\mathrm{m}}(c); \lambda \right) \right\} \\
\leqslant & \max\left\{ P\left(\hat{T}(\theta_\ell) > \tilde{c}^{\mathrm{m}}(\theta_\ell, c) \text{ or } \left\{ \hat{T}(\theta_m) > \tilde{c}^{\mathrm{m}}(\theta_m, c) \text{ and } \hat{T}(\theta_u) > \tilde{c}^{\mathrm{m}}(\theta_u, c) \right\}; \lambda \right), \right. \\
& \left. P\left(\hat{T}(\theta_u) > \tilde{c}^{\mathrm{m}}(\theta_u, c) \text{ or } \left\{ \hat{T}(\theta_m) > \tilde{c}^{\mathrm{m}}(\theta_m, c) \text{ and } \hat{T}(\theta_\ell) > \tilde{c}^{\mathrm{m}}(\theta_\ell, c) \right\}; \lambda \right) \right\} \\
=: & \bar{p}(c, \lambda).
\end{aligned} \tag{2.30}$$

Therefore, it is valid, but conservative, to replace $p(c;\theta,\lambda)$ in (2.28) 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[6]

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

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

$$\hat{c}^{\mathrm{t}} = \inf_c \left\{ c \geqslant 0 : \sup_{\lambda \in \hat{\Lambda}_\eta} \bar{p}(c,\lambda) + \eta \leqslant \alpha \right\}, \tag{2.31}$$

and $\bar{p}(c,\lambda)$ is defined in (2.30). 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 (2.29). 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 research.

Remark 11. *The relaxation in (2.30) 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.*[7] *Conversely, if the identified set is very small, then the set's coverage will be similar to the*

[6] Section 3.5.2 gives a more efficient $\hat{\Lambda}_\eta$ set.

[7] To see this, note that by construction $\hat{c}^{\mathrm{m}} \geqslant 0$, 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 $P\left(\hat{\lambda}_{\ell,\hat{b}_\ell} \leqslant \theta_m \leqslant \hat{\lambda}_{u,\hat{b}_u}\right) \to 1$.

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*

$$p(c;\theta,\lambda) \leqslant \max\Big\{P\left(\left[\theta_\ell, \theta_{1/4}\right] \nsubseteq \widetilde{CI}^m(c^t);\lambda\right), P\left(\left[\theta_{1/4}, \theta_m\right] \nsubseteq \widetilde{CI}^m(c^t);\lambda\right)$$
$$P\left(\left[\theta_m, \theta_{3/4}\right] \nsubseteq \widetilde{CI}^m(c^t);\lambda\right), P\left(\left[\theta_{3/4}, \theta_u\right] \nsubseteq \widetilde{CI}^m(c^t);\lambda\right)\Big\}$$

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 12. *(Simple Union Bounds, Cont.) For simplicity, in this example I let* $\eta = 0$*. With* $\alpha = 0.05$ *and* $\alpha^c = 0.04$*, we can calculate that* $c^t = 1.06$*. In Figure 2.2b, 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 2.2a. 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* $\hat{c}^t$ *removes some counter-intuitive values from the rejection region close to* H_0*, e.g.* $\left(\hat{\lambda}_1 - \theta, \hat{\lambda}_2 - \theta\right) = (\varepsilon, \varepsilon) \approx (0,0)$*.* □

2.3.4 Size and Power Properties

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

Assumption 13. *(Known Singularity) There are known* $|\mathcal{B}| \times J$ *matrices* A_ℓ*,* A_u *such that*

$$\lambda_\ell = A_\ell \delta_P,\ \lambda_u = A_u \delta_P \tag{2.32}$$
$$\hat{\lambda}_\ell = A_\ell \hat{\delta}_n,\ \hat{\lambda}_u = A_u \hat{\delta}_n$$

$$\hat{\Sigma}_n = \begin{bmatrix} A_\ell' & A_u' \end{bmatrix}' \hat{\Omega}_n \begin{bmatrix} A_\ell' & A_u' \end{bmatrix}$$

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

Assumption 14. *(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 15. *(Full Rank) Let* $\mathcal{S}$ *denote the set of matrices with eigenvalues bounded below by* $\underline{e} > 0$ *and above by* $\bar{e} \geqslant \underline{e}$. *For all* $P \in \mathcal{P}$, $\Omega_P \in \mathcal{S}$.

Assumption 16. *(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 17. *(Confidence Set of* $(\lambda_\ell, \lambda_u)$*) For all* $\eta \in [0, \frac{\alpha}{4})$, *the confidence set* $\hat{\Lambda}_\eta$ *satisfies*

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

Assumptions 13, 14, 15 and 16 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 $\hat{c}^{\mathrm{c}}(\theta, \alpha)$. These assumptions hold for the examples in Section 2.2.2 with finite $\mathcal{B}$ under mild conditions. Assumption 17 requires that $\hat{\Lambda}$ is a uniformly valid $1-\eta$ confidence set of $(\lambda_\ell, \lambda_u)$, e.g. $\hat{\Lambda}$ implied by (A.2) in Section A.1.1.

Theorem 18. *(Uniform Coverage) Suppose Assumptions 13, 14, 15, 16, and 17 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*

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

Remark 19. *Theorem 18 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 16 fails. The main idea is that we can rewrite the union bounds as the union of several sub-union bounds, with Assumption 16 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. □

Remark 20. *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*

$$\liminf_n \inf_{P\in\mathcal{P}} P\left(\hat{\mathcal{B}} = \mathcal{B}\right) = \liminf_n \inf_{P\in\mathcal{P}} P\left(d_H\left(\hat{\mathcal{B}}, \mathcal{B}\right) < \varepsilon\right) = 1. \tag{2.33}$$

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 (2.33) 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 (2.14), and (ii) the adjusted bootstrap CI proposed in Ye et al. (2023).

Theorem 21. *(Power Comparison with Simple CI) Suppose Assumptions 13, 14, 15, and 16 hold. And* $\hat{\Lambda}_\eta$ *is defined as in (A.1). 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*

$$\limsup_{P\in\mathcal{P}_n} \max_{b_1\in\mathcal{B}} \min_{b_2\in\mathcal{B}} \rho_\ell(b_1, b_2) < \rho_1^*(\alpha, \alpha^c), \tag{2.34}$$

$$\limsup_{P\in\mathcal{P}_n} \rho_\ell(b_\ell, b_u) < \rho_2^*(\alpha, \eta), \tag{2.35}$$

where $\rho_1^*(\alpha, \alpha^c)$ *and* $\rho_2^*(\alpha, \eta)$ *are defined in Lemma 69 and Lemma 66, respectively.*[8]

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

$$\mathcal{P}_n = \left\{ P \in \mathcal{P} : \lambda_{u,b_u} - \lambda_{\ell,b_\ell} \geqslant \frac{\kappa_n}{\sqrt{n}} \right\}. \tag{2.36}$$

It holds that

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

$$\liminf_n \inf_{P\in\mathcal{P}_n} P\left(CI^m\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right) \subseteq CI^{sim}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha'\right)\right) = 1. \tag{2.37}$$

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*

$$\liminf_{a_n\to\infty} P_{a_n}\left(\theta_{a_n} \notin CI^m\left(\hat{\lambda}_{a_n}, \hat{\Sigma}_{a_n}/a_n; \alpha\right)\right) - P_{a_n}\left(\theta_{a_n} \notin CI^{sim}\left(\hat{\lambda}_{a_n}, \hat{\Sigma}_{a_n}/a_n; \alpha\right)\right) > 0. \tag{2.38}$$

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

The first part of Theorem 21 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:

$$\rho_1^*(0.05, 0.04) = 0.84, \qquad \rho_1^*(0.10, 0.08) = 0.83,$$

$$\rho_2^*(0.05, 0.001) \approx 1, \qquad \rho_2^*(0.10, 0.001) \approx 1.$$

[8] 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 $\hat{\rho}_\ell$.

The values are large and thus the restriction (2.34) is not binding in most applications.

The second part of Theorem 21 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 (2.25).

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 22. *(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

$$\liminf_{n\to\infty} \inf_{P \in \mathcal{P}} P\left(\theta_n \notin CI^m\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) = 1 \tag{2.39}$$

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

Theorem 22 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 61, while it is rejected by CI^{YKHS} with probability bounded above by α. Hence, CI^{m} has large power improvement upon CI^{YKHS}.

2.4 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:

$$E_P\left[m_j(X_i; \beta, \varphi_P)\right] \leqslant 0,\ j = 1, ..., J_1, \tag{2.40}$$

$$E_P\left[m_j(X_i; \beta, \varphi_P)\right] = 0,\ j = J_1 + 1, ..., J_1 + J_2, \tag{2.41}$$

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 (2.40) and (2.41). Additionally, I assume that

$$\varphi_P = \varphi^\dagger\left(E_P\left[m_\varphi\left(X_i\right)\right]\right), \tag{2.42}$$

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

$$\lambda_{P,k}(\beta, \varphi) = \lambda_k^\dagger\left(E_P\left[m_k\left(X_i; \beta, \varphi\right)\right]\right) \quad k = \ell, u, \tag{2.43}$$

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 2.5.2 based on Dickstein and Morales (2018).

It is important to note that (2.42) and (2.43) 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.

2.4.1 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

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

Then, define

$$\bar{m}_{\mathcal{J}}(\beta, \varphi) = \frac{1}{n}\sum_{i=1}^{n} m_{\mathcal{J}}(X_i; \beta, \varphi) ,$$

$$\bar{m}_k(\beta, \varphi) = \frac{1}{n}\sum_{i=1}^{n} m_k(X_i; \beta, \varphi) , \quad k = \ell, u,$$

$$\bar{m}_{\varphi} = \frac{1}{n}\sum_{i=1}^{n} m_{\varphi}(X_i) .$$

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

$$\hat{\lambda}_k(\beta) = \lambda_k^{\dagger}(\bar{m}_k(\beta, \hat{\varphi})) , \tag{2.44}$$

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

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

$$\mathbb{G}_n(\beta) = \sqrt{n}\begin{bmatrix} \bar{m}_{\mathcal{J}}(\beta, \hat{\varphi}) - E_P[m_{\mathcal{J}}(X; \beta, \varphi_P)] \\ \hat{\lambda}_{\ell}(\beta) - \lambda_{\ell,P}(\beta, \varphi_P) \\ \hat{\lambda}_u(\beta) - \lambda_{u,P}(\beta, \varphi_P) \end{bmatrix} . \tag{2.45}$$

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

$$G_P(\beta) = \begin{bmatrix} \mathcal{I}_J, & 0, & 0, & \nabla_{\varphi'} E\left[m(X;\beta,\varphi_P)\right]\nabla'\varphi^\dagger \\ 0, & \nabla'\lambda_\ell^\dagger, & 0, & \nabla_{\varphi'}\lambda_{P,\ell}(\beta,\varphi_P)\nabla'\varphi^\dagger \\ 0, & 0, & \nabla'\lambda_u^\dagger, & \nabla_{\varphi'}\lambda_{P,u}(\beta,\varphi_P)\nabla'\varphi^\dagger \end{bmatrix}. \tag{2.46}$$

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 (2.46), $\nabla'\lambda_k^\dagger$ 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 essay, 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

$$\hat{\sigma}_j(\beta) = \sqrt{\hat{\Sigma}_{jj}},\ \forall j = 1,...,J, \qquad \hat{\sigma}_\ell(\beta) = \sqrt{\hat{\Sigma}_{J+1}(\beta)}, \qquad \hat{\sigma}_u(\beta) = \sqrt{\hat{\Sigma}_{J+2}(\beta)}, \tag{2.47}$$

$$\hat{\sigma}_{u\ell}(\beta) = \sqrt{\begin{bmatrix} 0_{1\times J}, & 1, & -1 \end{bmatrix}\hat{\Sigma}(\beta)\begin{bmatrix} 0_{1\times J}, & 1, & -1 \end{bmatrix}'}$$

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 (2.47) with $\hat{\Sigma}$ replaced by Σ.

Lastly, let

$$D_{P,k}(\beta) = \frac{\nabla_{\beta'}\lambda_{P,k}(\beta,\varphi_P)}{\sigma_{P,k}(\beta,\varphi_P)},\ k = \ell, u, \tag{2.48}$$

$$D_{P,j}(\beta) = \frac{\nabla_{\beta'} E\left[m_j(X;\beta,\varphi_P)\right]}{\sigma_{P,j}(\beta,\varphi_P)},\ j = 1,...,J, \tag{2.49}$$

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 $\hat{D}_j(\beta)$.

2.4.2 The Confidence Interval

The confidence interval has the following form

$$CI_n = \left[\inf_{\beta \in \hat{\mathcal{B}}} \hat{\lambda}_\ell(\beta) - \frac{\hat{\sigma}_\ell(\beta)}{\sqrt{n}} \hat{c}_n(\beta), \quad \sup_{\beta \in \hat{\mathcal{B}}} \hat{\lambda}_u(\beta) + \frac{\hat{\sigma}_u(\beta)}{\sqrt{n}} \hat{c}_n(\beta) \right], \tag{2.50}$$

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

$$\hat{\mathcal{B}} = \left\{ \beta \in \overline{\mathcal{B}} : \frac{\sqrt{n}\bar{m}_j(\beta, \hat{\varphi})}{\hat{\sigma}_j(\beta)} \leqslant \hat{c}_n(\beta), \ \forall j = 1, ..., J \right\}, \tag{2.51}$$

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.[9] 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

$$\hat{c}_n(\beta) = \max \left\{ \hat{c}_u(\beta), \ \hat{c}_\ell(\beta) \right\}, \tag{2.52}$$

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

$$\hat{c}_k(\beta) = \inf \left\{ c \in \mathbb{R}_+ : P \left(\min_{\Delta \in \hat{\mathbf{\Delta}}_n(\beta, c)} \mathbb{Z}_k^s(\beta) + \hat{D}_k(\beta)\Delta \leqslant c, \text{ and } \right. \right. \tag{2.53}$$

$$\left. \left. \min_{\Delta \in \hat{\mathbf{\Delta}}_n(\beta, c)} \mathbb{Z}_{[k]}^s(\beta) + \hat{D}_{[k]}(\beta)\Delta + \hat{\zeta}_0(\beta) \leqslant c \middle| \{X_i\}_{i=1}^n \right) \geqslant 1 - \alpha \right\}, \tag{2.54}$$

and the random feasible set for Δ is

$$\begin{aligned} \hat{\mathbf{\Delta}}_n(\beta, c) = \Big\{ & \Delta \in \sqrt{n}(\overline{\mathcal{B}} - \beta) \cap \rho \left[-1, 1\right]^d : \\ & \mathbb{Z}_j^s(\beta) + \hat{D}_j(\beta)\Delta + \hat{\zeta}_j(\beta) \leqslant c, j = 1, \ldots, J \Big\}. \end{aligned} \tag{2.55}$$

The elements in (2.53), (2.54), and (2.55) are listed below.

[9] To see this, (2.60), (2.62), (2.63) have similar structure with non-positive last term $\frac{E\left[m_j(\beta, \varphi_P)\right]}{\sigma_{P,j}(\beta, \varphi_P)/\sqrt{n}}$, $\frac{\lambda_{P,\ell}(\beta) - \theta}{\sigma_{P,\ell}(\beta)/\sqrt{n}}$, $\frac{\theta - \lambda_{P,u}(\beta)}{\sigma_{P,u}(\beta)/\sqrt{n}}$ respectively.

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

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

2. $\hat{\zeta}_j(\beta)$, $\hat{\zeta}_0(\beta)$ in (2.54) and (2.55) are generalized moment selection (GMS) type functions defined as

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

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

   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 (2.54) is negligible.

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

$$\rho = \Phi^{-1}\left(\frac{1}{2} + \frac{1}{2}\left(1 - \eta / \binom{J_1 + J_2 + 2}{d}\right)^{1/d}\right) \tag{2.56}$$

   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 (2.53), (2.54), (2.55) 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

$$\tilde{\beta} \in \overline{\mathcal{B}} \cap \left\{ \beta + \frac{\rho}{\sqrt{n}} [-1,1]^d \right\} \tag{2.57}$$

such that (i) θ is covered

$$\hat{\lambda}_\ell(\tilde{\beta}) - \frac{\hat{\sigma}_\ell(\tilde{\beta})}{\sqrt{n}} \hat{c}_n(\tilde{\beta}) \leqslant \theta \leqslant \hat{\lambda}_u(\tilde{\beta}) + \frac{\hat{\sigma}_u(\tilde{\beta})}{\sqrt{n}} \hat{c}_n(\tilde{\beta}) \tag{2.58}$$

and (ii) the moment conditions are not rejected

$$\frac{\sqrt{n}\bar{m}_j(\tilde{\beta})}{\hat{\sigma}_j(\tilde{\beta})} \leqslant \hat{c}_n(\tilde{\beta}), \quad \forall j = 1, ..., J. \tag{2.59}$$

That said, we only need to calibrate $\hat{c}(\tilde{\beta})$ such that (2.58) and (2.59) 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 (2.57), $\tilde{\beta}$ is restricted to a local neighborhood of β, and the tuning parameter ρ defines its boundary. This allows us to linearize (2.58) and (2.59), so that they can be approximated by (2.53, 2.54) and (2.55), respectively. As for (2.59), a first-order Taylor expansion gives

$$\frac{\sqrt{n}\bar{m}_j(\tilde{\beta})}{\hat{\sigma}_j(\tilde{\beta})} \approx \frac{\mathbb{G}_{n,j}(\beta)}{\sigma_{P,j}(\beta)} + D_{P,j}(\beta)\Delta + \frac{\sqrt{n}E\left[m_j(\beta, \varphi_P)\right]}{\sigma_{P,j}(\beta)} \tag{2.60}$$

where $\Delta = \sqrt{n}(\tilde{\beta} - \beta)$, $\mathbb{G}_n$, D_P, and σ_P are defined in (2.45), (2.49) and below (2.47), respectively. In (2.55), 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, (2.55) is a valid approximation of the moment restrictions in (2.59). Similarly, (2.58) is equivalent to

$$\frac{\hat{\lambda}_\ell(\tilde{\beta}) - \theta}{\hat{\sigma}_\ell(\tilde{\beta})/\sqrt{n}} \leqslant \hat{c}(\tilde{\beta}), \qquad \frac{\theta - \hat{\lambda}_u(\tilde{\beta})}{\hat{\sigma}_u(\tilde{\beta})/\sqrt{n}} \leqslant \hat{c}_n(\tilde{\beta}). \tag{2.61}$$

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

$$\frac{\hat{\lambda}_\ell(\tilde{\beta}) - \theta}{\hat{\sigma}_\ell(\tilde{\beta})/\sqrt{n}} \approx \frac{\mathbb{G}_{n,\ell}(\beta)}{\sigma_{P,\ell}(\beta)} + D_{P,\ell}(\beta)\Delta + \frac{\lambda_{P,\ell}(\beta) - \theta}{\sigma_{P,\ell}(\beta)/\sqrt{n}} \tag{2.62}$$

$$\frac{\theta - \hat{\lambda}_u(\tilde{\beta})}{\hat{\sigma}_u(\tilde{\beta})/\sqrt{n}} \approx \frac{\mathbb{G}_{n,u}(\beta)}{\sigma_{P,u}(\beta)} + D_{P,u}(\beta)\Delta + \frac{\theta - \lambda_{P,u}(\beta)}{\sigma_{P,u}(\beta)/\sqrt{n}}. \tag{2.63}$$

The approximations for (2.62, 2.63) 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

$$\frac{\sqrt{n}\left(\lambda_{P,u}(\beta) - \lambda_{P,\ell}(\beta)\right)}{\kappa_n \max\left\{\sigma_{P,u}(\beta), \sigma_{P,\ell}(\beta), \sigma_{P,u\ell}(\beta)\right\}} \to \infty, \tag{2.64}$$

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 (2.61) is not binding with probability approaching one. Otherwise, we can replace both terms with zero if (2.64) does not hold, which introduces a conservative distortion, which in turn leads to the restriction in (2.53) and (2.54).

Remark 23. *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 essay, 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 essay 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 (2.49) 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 (2.49) as it is easier*

to calculate. Moreover, the analogous Jacobian for λ_k defined in (2.48) follows directly from the Taylor expansion in (2.62)-(2.63), where the denominators are fixed. □

Remark 24. *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.* □

2.4.3 Asymptotic Results

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

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

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

Assumption 25. *$\overline{\mathcal{B}} \subset \mathbb{R}^d$ is a compact hyperrectangle with nonempty interior.*

Assumption 25 restricts the shape of the parameter space. It is satisfied in most applications and guarantees that the calculation in (2.53) can be obtained with simple linear programs.

Assumption 26. *Let*

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

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

$$\left\|G_P(\beta) - G_P(\tilde{\beta})\right\| = M \left\|\beta - \tilde{\beta}\right\|,$$

$$\left\|D_P(\beta) - D_P(\tilde{\beta})\right\| \leqslant M \left\|\beta - \tilde{\beta}\right\|,$$

$$\|G_P(\beta)\| \leqslant M, \qquad \|D_P(\beta)\| \leqslant M.$$

Assumption 27. *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

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

Assumption 26 imposes that the derivatives D_P and G_P are sufficiently smooth in their arguments, and Assumption 27 requires consistent estimators for the Jacobian D_P and G_P. Under these two assumptions, we can approximate the Taylor expansion in (2.60), (2.62), (2.63) with feasible analogs.

Assumption 28. *For some constants $\omega > 0$, $\varepsilon > 0$, all distributions $P \in \mathcal{P}$ satisfy the following condition. Let*

$$\mathcal{J}_1(P,\beta;\varepsilon) = \left\{j \in \{1,\cdots,J_1\} : \frac{E_P\left[m_j\left(X_i;\beta,\varphi_P\right)\right]}{\sigma_{P,j}(\beta,\varphi_P)} \geqslant -\varepsilon\right\}, \tag{2.65}$$

$$\mathcal{J}_k(P,\beta;\varepsilon) = \begin{cases}\{u,\ell,J_1+1,...,J_1+J_2\} & \text{if } \tau_{0,P,\beta} \geqslant -\varepsilon\\ \{k,J_1+1,...,J_1+J_2\} & \text{otherwise}\end{cases} \tag{2.66}$$

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 $k = \ell, u$,

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

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 28 imposes that the covariance matrix for those moments is non-singular. In Assumption 73 in Appendix A.4.2, 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 29. *Suppose Assumptions 26, 27, 25, 72 hold. In addition, suppose Assumption 28 or 73 holds. Let* $0 < \alpha < 1/2$. *Then*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\bigcup_{\beta\in\mathcal{B}}\left[\lambda_{P,\ell}(\beta),\lambda_{P,u}(\beta)\right]} P\left(\theta \in CI_n\right) \geqslant 1-\alpha.$$

Assumptions 72, which is a regularity condition for the moments, can be found in Appendix A.4.2.

2.5 Simulation

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

2.5.1 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 2.3, I consider two existing procedures for union bounds: (i) the adjusted bootstrap in Ye et al. (2023), (ii) the simple confidence interval in (2.14), and (iii) the inference procedure in Rambachan and Roth (2023).[10] All three methods are uniformly valid. All tuning parameters are set at the values in the papers in which they are proposed.

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

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

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) P.

[10] 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.

Christensen et al. (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 2.3. 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 2.3, I plot the rejection rate near the upper bound. The lower bound is similar and thus omitted. The horizontal axis is the value of θ, while the vertical axis is the rate that θ is not included in the CI. The asterisks represent the identified set, and the nominal rejection rate is 10%. The modified conditional CI is the red curve and it has proper size control in all simulation designs. The simple CI is the black dotted curve and it has significantly lower power than modified conditional CI in all designs.

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 2.3a, 2.3d, and 2.3j. In some DGPs, Rambachan and Roth (2023) may perform worse than the simple CI, e.g. in Figure 2.3h. When there is only one large violation, for example in Figure 2.3c-2.3l, 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 2.3g and 2.3h. This is consistent with the slower convergence rate of the YKHS CI to the identified set

and Theorem 22.

2.5.2 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

$$r_h = X_1 + X_2 + X_3,$$

$$r_f = \varphi_f r_h + e.$$

The expected profit of export is given by

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

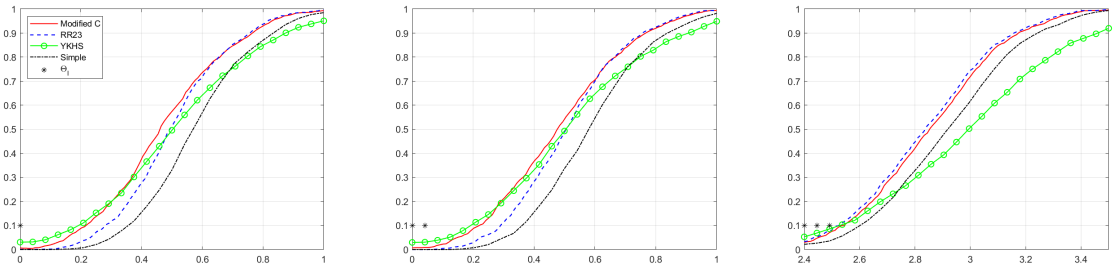
where η is the elasticity, $\mathcal{J}$ is the information set the firms have when making the export decision, and $(X_1, X_2, X_3, \nu) \sim \mathcal{N}(0, \mathcal{I}_4)$. Then the decision is given by

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

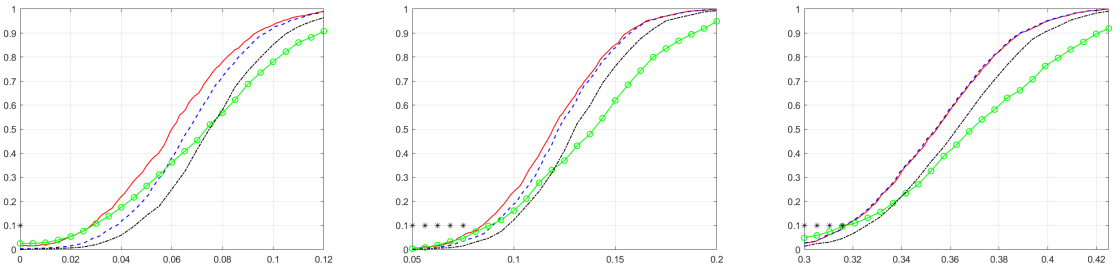
where $\beta_1 = \tilde{\beta}_1^{-1}$, $\beta_2 = \tilde{\beta}_1^{-1}\tilde{\beta}_2$. This reparameterization is just for simplicity. The researcher observes (X, r_h, d, dr_f) and a subset of the information set $Z \subseteq \mathcal{J}$. The parameters are $(\varphi_f, \beta_1, \beta_2)$, and $\eta = 2$ is known to both the researcher and firms.[11] 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 A.2.1.

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 2.1, I report the calibrated

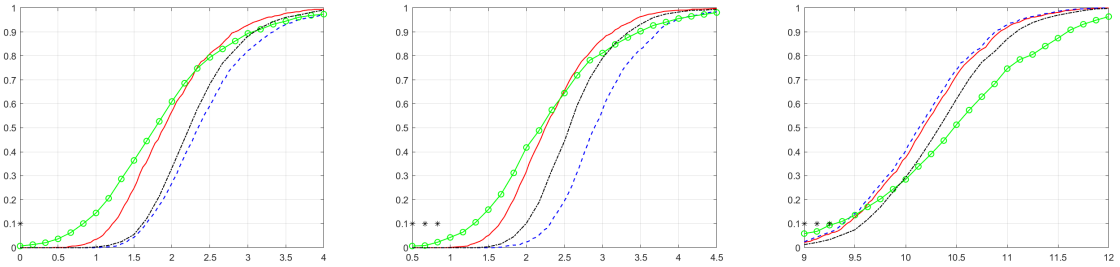
[11] Note that this is in fact a normalization since η and β_1 cannot be identified separately.



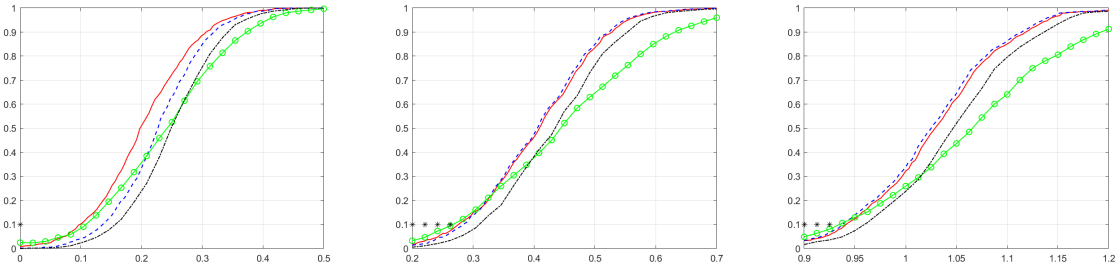
(a) Parallel Trends with Σ_1 (b) Small Violation with (γ_1, Σ_1) (c) Large Violation with Σ_1



(d) Parallel Trends with Σ_2 (e) Small Violation with (γ_2, Σ_2) (f) Large Violation with Σ_2



(g) Parallel Trends with Σ_3 (h) Small Violation with (γ_3, Σ_3) (i) Large Violation with Σ_3



(j) Parallel Trends with Σ_4 (k) Small Violation with (γ_4, Σ_4) (l) Large Violation with Σ_4

FIGURE 2.3: 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 2.3a, 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 P. Christensen et al., 2023 with $\underline{T} = 15$.

Table 2.1: 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]

projection CI using the inference proposed in Section 2.4, 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.

2.6 Empirical Illustration

2.6.1 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

$$\log(\text{emp}_{rt}) = \sum_{\tau=2011, \tau \neq 2014}^{2016} \gamma_\tau \overline{GAP}_r \mathbf{1}\left[\tau = t\right] + \alpha_r + \xi_t + \varepsilon_{rt} \tag{2.67}$$

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 2.4a shows the estimated coefficients $\{\hat{\gamma}_\tau\}$ from specification (2.67). 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 2.4a 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.

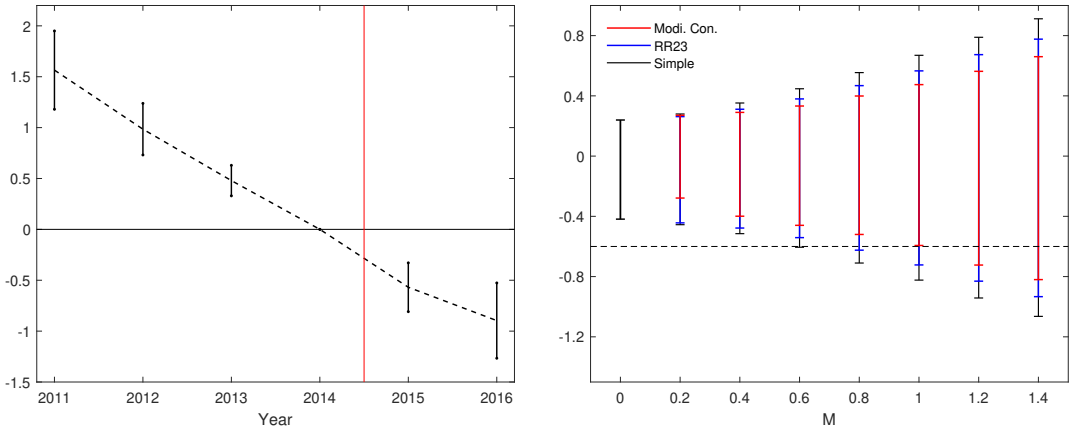
In particular, the authors conduct the sensitivity analysis using the second differences relative magnitudes (SDRM) relaxation. This approach assumes that

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

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 2.4a. 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 $100(\gamma_{2015} - \xi_{2015})\%$ in expectation.

In Figure 2.4b, I report the 95% confidence intervals for different values of M constructed based on three different methods: the modified conditional CI proposed in Section 2.3, the hybrid CI in Rambachan and Roth (2023), and the simple CI in (2.14).[12] We can clearly see that the modified conditional CI is the shortest and the simple CI is the widest for all

[12] 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.



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

FIGURE 2.4: Empirical Illustration for finite $\mathcal{B}$ with $\alpha = 0.05$

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 (2.67) 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 2.4b. 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.

2.6.2 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 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 A.2.2. In this model, the structural parameter β is partially identified by moment inequalities as in (2.40) 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 2.2, 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 $\hat{\varphi}$, mak-

Table 2.2: 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]

ing 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 interval 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 2.4. 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.

2.7 Conclusion

In this essay, 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.

3. Local linearization based subvector inference in moment inequality models

This essay introduces a bootstrap-based profiling inference method for subvectors in moment inequality models. [1]

3.1 Introduction

A moment inequality model has the form

$$\begin{aligned} E_F\big[m_j(W_i;\theta_0)\big] &\geqslant 0,\ j=1,...,p \\ E_F\big[m_j(W_i;\theta_0)\big] &= 0,\ j=p+1,...,k \end{aligned} \tag{3.1}$$

where $\{W_i\}_{i=1}^n$ is an independent and identically distributed (i.i.d.) sequence of random variables with distribution F and $m=(m_1,...,m_k)' : \mathbb{R}^d \times \Theta \to \mathbb{R}^k$ is a known measurable function of a finite-dimensional parameter vector $\theta \in \Theta$. Frequently the parameter θ_0 in this model is not known to be point identified, a feature that complicates both estimation and inference substantially. Most of the recent papers on inference for moment inequalities focus on confidence sets for the entire parameter vector, e.g. D. W. K. Andrews and Shi (2013), D. W. K. Andrews and Soares (2010), Chernozhukov et al. (2007), and Rosen (2008). See Canay and Shaikh (2017) and Molinari (2020) for more references. However, applied researchers are often interested in a few specific components (a subvector) of the whole parameter vector, i.e. $c'\theta_0$ where c is a $d_\theta \times d_r$ matrix with full column rank.

Bugni et al. (2017) (BCS17) propose a bootstrap-based test procedure for $H_0 : c'\theta_0 = r$. Similar to Romano and Shaikh (2008), they calculate the profiled test statistic, as well as its bootstrap approximation, by minimizing the test statistic over the *whole* null parameter space to profile out the nuisance parameters. Building on their work, this essay uses the same test statistic and improves upon BCS17 by applying *local* linearization methods to calculate the critical value. The challenge of local linearization is that the a naive bootstrapped minimizer $\hat{\theta}^*_{n,r}$ may be near a different point of the identified set from the original minimizer $\hat{\theta}_{n,r}$. To address this issue, a penalty term is added to the bootstrapped test

[1] This essay was originally published in the Journal of Econometrics, Volume 238, Issue 1, Pages 105594, [2024]

statistic to keep $\hat{\theta}^*_{n,r}$ close to $\hat{\theta}_{n,r}$, and therefore, close to the same point in the identified set. This enables linearization of the moment conditions and the constraints around this point. In this way, the bootstrap iterates over quadratic programming as opposed to general nonlinear problems. The local linearization based test (Test LL) controls asymptotic size uniformly over a large class of data generating processes and can be inverted to construct the confidence interval. As for power properties, Test LL weakly dominates projection-based procedures in any finite sample, and strictly dominates projection-based procedures in terms of asymptotic power.

Kaido et al. (2019) (KMS19) introduce a calibrated projection-based subvector inference method, where they calibrate the critical value so that the coverage of the subvector is above the prespecified probability. KMS19 also linearize constraints locally around the evaluated points, and the critical value can be calculated by analyzing linear programs. However, KMS conducts the test for a grid of θ; while LL, as a profile based method, conducts the test for a grid of r only. The grid search over a relatively large dimensional space can be time-consuming, which makes the LL method more computationally efficient, especially with d_θ much larger than d_r. G. Cox and Shi (2019) propose a simple test for moment inequalities based on the rank of the active inequalities in finite samples, but it is required that the nuisance parameters enter the moment conditions linearly to adapt it for subvector inference. Belloni et al. (2018) study subvector inference with many moment inequalities. Chen et al. (2018) provide Monte Carlo confidence sets for the identified sets, rather than the true value, of both the full parameter vector and subvectors. Kline and Tamer (2016) develop a Bayesian approach that can be used for subvector inference in partially identified models, but they do not provide uniform validity results.

The Monte Carlo simulation studies the finite sample properties of Test LL and compares it to BCS17 and KMS19, with all three methods implemented with the E-A-M algorithm proposed in KMS19. The simulation study, conducted on a simple entry game, considers two sample sizes with $n = 1000, 4000$ and three model sizes with the number of moment conditions $k = 8, 16, 24$ and the number of parameters $d_\theta = 7, 11, 15$. Test LL improves

upon the computing time of BCS17 (resp. KMS19) by a factor ranging from 30 to 170 (resp. 3 to 27). The computational advantage of Test LL is more prominent with a larger sample and model size, reaching the highest factors for the highest dimensional parameters. The simulation results also confirm the good size and power properties.

3.2 Local linearization based subvector inference

In this section, I set up the model and illustrate the local linearization based test.

3.2.1 The model and notation

I assume that the distribution F belongs to a baseline distribution space $\mathcal{P}$, and I maintain the following mild assumption on Θ.

Assumption 30. *The parameter space Θ is a non-empty, convex and compact subset of $\mathbb{R}^{d_\theta}$ with $d_\theta < \infty$.*

Definition 31. *For any distribution $F \in \mathcal{P}$, the corresponding identified set $\Theta_I(F)$ is given by*

$$\Theta_I(F) := \left\{ \theta \in \Theta : \begin{array}{ll} E_F\big[m_j(W_i;\theta)\big] \geqslant 0, & j = 1, ..., p \\ E_F\big[m_j(W_i;\theta)\big] = 0, & j = p+1, ..., k \end{array} \right\}.$$

By definition, the identified set $\Theta_I(F)$ is the set of parameter values $\theta \in \Theta$ that are observationally equivalent to the true value θ_0. The objective of this essay is to construct a confidence interval for $c'\theta_0$ where c is a $d_\theta \times d_r$ matrix. I do this by inverting a test of the hypotheses

$$H_0 : c'\theta_0 = r \quad \text{v.s.} \quad H_1 : c'\theta_0 \neq r, \tag{3.2}$$

where r is a $d_r \times 1$ vector. The hypothesis test is based on a profiled test statistic $J_n(r)$ such that H_0 is rejected if $J_n(r)$ is too large. Specifically, it takes the form

$$\phi_n^{LL}(r) := \mathbf{1}\left[J_n(r) > cv_n^{LL}(r, 1-\alpha)\right],$$

where $\mathbf{1}[\cdot]$ is the indicator function, $\alpha \in (0,1)$ is the significance level, and $cv_n^{LL}(r, 1-\alpha)$ is the critical value I define later.

Before presenting the test procedure, I introduce more notation. For vector a, b, denote $\|a\| := \sqrt{a'a}$ and denote $d(a,b) := \|a-b\|$ as the Euclidean distance between a and b. For point a and set B, let $d(a,B) := \inf_{b\in B} \|a-b\|$. For a scalar α, an element a and a set A, let $a + A = \{a + \tilde{a} : \tilde{a} \in A\}$ and $\alpha A = \{\alpha a : a \in A\}$. The ε-expansion of $\Theta_1 \subseteq \Theta$ is defined as

$$\Theta_1^\varepsilon := \{\theta \in \Theta : d(\theta, \Theta_1) \leqslant \epsilon\}.$$

Let

$$\Theta(r) := \{\theta \in \Theta : c'\theta = r\}$$

be the set of parameters that satisfy the subvector restriction, and

$$\Theta_I(r,F) := \{\theta \in \Theta_I(F) : c'\theta = r\}. \tag{3.3}$$

be the set of parameters that satisfy both the moment conditions and the subvector restriction. $\Theta_I(r,F)$ is compact and non-empty under H_0. Define the projection operator $T^r : \Theta \to \Theta$ that projects θ on $\Theta_I(r,F)$, i.e.

$$T^r\theta \in \underset{\tilde{\theta}\in\Theta_I(r,F)}{\arg\min}\, d(\tilde{\theta}, \theta),$$

where $T^r\theta$ is well defined under H_0. If the arg min set contains more than one element, $T^r\theta$ is one of the minimizers, and the choice does not affect the results. The operator T^r depends on F, but I leave this dependence implicit to simplify notation. Denote by

$$\Sigma_F(\theta) = E_F\left[\Big(m(W,\theta) - E_F\big[m(W,\theta)\big]\Big)\Big(m(W,\theta) - E_F\big[m(W,\theta)\big]\Big)'\right]$$

$$D_F(\theta) = \operatorname{diag}(\Sigma_F(\theta))$$

the variance-covariance matrix and the diagonal matrix of variance. Let $[x]_+ = \max(x,0)$, $[x]_- = \max(-x,0)$ and $\mathbb{R}_+ = [0,+\infty)$.

Definition 32. *The parameter space of (r,F) is defined by*

$$\mathcal{L} := \{(r,F) : F \in \mathcal{P}, r \in \{c'\theta : \theta \in \Theta\}\},$$

and the null parameter space of (r,F) is

$$\mathcal{L}_0 := \{(r,F) : F \in \mathcal{P}, r \in \{c'\theta : \theta \in \Theta_I(F)\}\}.$$

3.2.2 The profiled test statistic

Denote by

$$\bar{m}_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} m(W_i,\theta),$$

$$\hat{\Sigma}_n(\theta) := \frac{1}{n}\sum_{i=1}^{n}\big(m(W_i,\theta)-\bar{m}_n(\theta)\big)\big(m(W_i,\theta)-\bar{m}_n(\theta)\big)',$$

the sample mean and sample variance-covariance matrix of the moment functions in (3.1). Let

$$\hat{D}_n(\theta) := \operatorname{diag}\big(\hat{\Sigma}_n(\theta)\big)$$

be the diagonal matrix of sample variances, and

$$\hat{\Omega}_n(\theta) := \hat{D}_n(\theta)^{-\frac{1}{2}}\hat{\Sigma}_n(\theta)\hat{D}_n(\theta)^{-\frac{1}{2}}$$

be the sample correlation matrix.

Define the profiled test statistic by

$$J_n(r) := \inf_{\theta\in\Theta(r)} S\big(\hat{D}_n(\theta)^{-\frac{1}{2}}\sqrt{n}\bar{m}_n(\theta),\hat{\Omega}_n(\theta)\big) \tag{3.4}$$

where S is a test function specified by the econometrician and required to satisfy several regularity assumptions, see Assumptions 87, 88 in the Appendix. These assumptions are standard in the literature, see, e.g. D. W. K. Andrews and Guggenberger (2009) and D. W. K. Andrews and Soares (2010). Two widely used test functions are the modified method of moments (MMM) and the quasi-likelihood ratio (QLR) statistics given by

$$S^{\mathrm{MMM}}(m,\Sigma) = \sum_{j=1}^{p}\left[m_j/\Sigma_{jj}^{\frac{1}{2}}\right]_-^2 + \sum_{j=p+1}^{k}\left(m_j/\Sigma_{jj}^{\frac{1}{2}}\right)^2,$$

$$S^{\mathrm{QLR}}(m,\Sigma) = \inf_{t=\left(t_1,0_{k-p}\right):t_1\in\mathbb{R}_+^p}(m-t)'\Sigma^{-1}(m-t).$$

The test statistic takes the infimum of test function $S(\cdot)$ over the restricted parameter space $\Theta(r)$.

3.2.3 Local linearization based critical value

The critical value $cv_n^{\mathrm{LL}}(r, 1-\alpha)$ is the $1-\alpha$ quantile of

$$J_n^{LL}(r) := \inf_{\hat{\theta}_{n,r} \in \hat{\Theta}_I(r)} S\left(v_n^*(\hat{\theta}_{n,r}) + \varphi\left(\xi_n(\hat{\theta}_{n,r})\right) + \hat{G}_n(\hat{\theta}_{n,r})\Delta(\hat{\theta}_{n,r}), \hat{\Omega}_n(\hat{\theta}_{n,r})\right). \tag{3.5}$$

where

$$\Delta(\hat{\theta}_{n,r}) \in \underset{\Delta \in \mathbf{\Delta}_n^r(\hat{\theta}_{n,r})}{\operatorname{argmin}} \left(S\left(v_n^*(\hat{\theta}_{n,r}) + \varphi\left(\xi_n(\hat{\theta}_{n,r})\right) + \hat{G}_n(\hat{\theta}_{n,r})\Delta, \hat{\Omega}_n(\hat{\theta}_{n,r})\right) + \frac{\lambda_n}{n}\|\Delta\|^2\right). \tag{3.6}$$

The elements in (3.5) and (3.6) are listed below.

1. $v_n^*(\cdot)$ is a stochastic process defined as

$$v_n^*(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{D}_n(\theta)^{-\frac{1}{2}}\big(m(W_i, \theta) - \bar{m}_n(\theta)\big)\zeta_i, \quad \zeta_i | \{W_i\}_{i=1}^n \sim N(0,1) \text{ i.i.d.} \tag{3.7}$$

2. $\xi_n(\theta) = \frac{\sqrt{n}}{\kappa_n}\hat{D}_n(\theta)^{-\frac{1}{2}}\bar{m}_n(\theta)$ is a slackness function with $\kappa_n \to \infty$ and $\kappa_n = O\left(\frac{n^{1/4}}{(\ln n)^{1/\chi}}\right)$ where χ is defined in Assumption 88.5. Following D. W. K. Andrews and Soares (2010), the suggested choice is $\kappa_n = \sqrt{\ln n}$.
3. $\varphi(\cdot) = (\varphi_1(\cdot), ..., \varphi_k(\cdot))$ is a generalized moment selection (GMS) function with

$$\varphi_j(\xi) = \begin{cases} [\xi_j]_+ & \text{for } j = 1, ..., p \\ 0 & \text{for } j = p+1, ..., k. \end{cases} \tag{3.8}$$

4. $\hat{G}_n(\cdot)$ is a consistent estimator of the Jacobian $G_F(\theta) = \nabla_{\theta'}\left(D_F(\theta)^{-\frac{1}{2}} E_F\left(m(\theta)\right)\right)$.
5. λ_n is a non-random sequence satisfying $\lambda_n/\sqrt{n} \to \infty$ and $\lambda_n(\ln \kappa_n^2)^{\frac{2}{\chi}} = O(n)$ where χ is defined in Assumption 88.5. A valid choice is $\lambda_n = \sqrt{n}\kappa_n^2$.
6. $\hat{\Theta}_I(r)$ is a non-empty subset of the minimizers

$$\varnothing \neq \hat{\Theta}_I(r) \subseteq \underset{\theta \in \Theta(r)}{\operatorname{argmin}}\ S\left(\hat{D}_n(\theta)^{-\frac{1}{2}}\sqrt{n}\bar{m}_n(\theta), \hat{\Omega}_n(\theta)\right). \tag{3.9}$$

More generally, $\hat{\Theta}_I(r)$ can in fact be a subset of approximate minimizers satisfying

$$\sup_{\theta \in \hat{\Theta}_I(r)} d\left(\theta, \Theta_I(r, F)\right) = \sup_{\theta \in \hat{\Theta}_I(r)} d\left(\theta, T^r\theta\right) = o_p\left(n^{-1/2}(\ln \kappa_n^2)^{1/\chi}\right) \tag{3.10}$$

with χ defined in Assumption 88.5.

7. $\mathbf{\Delta}_n^r(\cdot)$ is a set such that $\mathbf{\Delta}_n^r(\theta) \subseteq \sqrt{n}(\Theta(r) - T^r\theta)$ for all $\theta \in \hat{\Theta}_I(r)$ with probability approaching one.

To illustrate the construction of the critical value, first, omit the outer inf operator in (3.5) and focus on a single $\hat{\theta}_{n,r} \in \hat{\Theta}_I(r)$. The first step estimator, $\hat{\theta}_{n,r}$, needs to be recalculated for each bootstrap draw. The problem with doing a naive bootstrap is that it may be near a different point of the identified set rather than $T^r\hat{\theta}_{n,r}$. A modified bootstrapped version of $\hat{\theta}_{n,r}$ is given by $T^r\hat{\theta}_{n,r} + \frac{1}{\sqrt{n}}\Delta(\hat{\theta}_{n,r})$ with $\Delta(\hat{\theta}_{n,r})$ defined in (3.6). In that expression, the first element $S(\cdot)$ is a local linearization of the test function around $T^r\hat{\theta}_{n,r}$.[2] The second element is a penalty term that keeps the bootstrap version of $\hat{\theta}_{n,r}$ near $T^r\hat{\theta}_{n,r}$, the same point in the identified set. The feasible set of $T^r\hat{\theta}_{n,r} + \frac{1}{\sqrt{n}}\Delta(\hat{\theta}_{n,r})$ is $\Theta(r)$, hence we need to choose $\Delta(\hat{\theta}_{n,r}) \in \sqrt{n}\left(\Theta(r) - T^r\hat{\theta}_{n,r}\right)$. However, $T^r\hat{\theta}_{n,r}$ is unknown and thus we approximate the boundary constraint by $\mathbf{\Delta}_n^r(\hat{\theta}_{n,r})$. When the bootstrapped minimizer is plugged into the linearized statistic, the result is (3.5). Lastly, note that this local optimization holds for all $\hat{\theta}_{n,r} \in \hat{\Theta}_I(r)$, thus we implement the outer inf operator in (3.5) to improve power.

The penalty term in (3.6) is a new contribution of this essay, and it validates the local linearization approximation of the moment conditions, which greatly improves the computation. When Test LL is implemented with the MMM or QLR test function and a polyhedral $\mathbf{\Delta}_n^r(\theta)$, the inner minimization in (3.6) is a simple quadratic programming problem. Also, note that $\hat{\Theta}_I(r)$ is not required to contain all the minimizers for the coverage property, and in practice, $\hat{\Theta}_I(r)$ contains only a few elements. In this way, the bootstrap iterates over quadratic programming problems, which can be solved with high-speed solvers like CVXGEN in Mattingley and Boyd (2012).

Next, I present the assumptions that are important for proper coverage.

---

[2] In (3.6), we can replace $T^r\hat{\theta}_{n,r}$ with $\hat{\theta}_{n,r}$ since they are close enough.

Assumption 33. *For all* $(r, F) \in \mathcal{L}_0$ *and* $\theta \in \Theta(r)$,

$$S\Big(E_F\big(m(W,\theta)\big), \Sigma_F(\theta)\Big) \geqslant c \min\{\delta, d(\theta, \Theta_I(r,F))\}^\chi$$

for some constants $c, \delta > 0$ *and for* χ *as in Assumption 88.5.*

Assumption 33 is borrowed from BCS17. It states that in a neighborhood of the null identified set $\Theta_I(r,F)$, the population version of the test function can be bounded below by a polynomial of the deviation of θ from $\Theta_I(r,F)$. It is similar to the polynomial minorant condition in Chernozhukov et al. (2007) (Eqs. (4.1) and (4.5)), but logically independent. This assumption rules out some cases where the entire boundary is a smooth manifold, focusing on a simpler class of models with a well-behaved $\Theta_I(r,F)$.[34] However, it holds under many cases of practical interest. Kaido et al. (2022) extensively analyze the constraint qualifications in partial identification, including the minorant type condition, and so I refer to their paper for detailed discussion. Assumption 33, together with other regularity assumptions, guarantees that $\hat{\theta}_{n,r}$ is close enough to $T^r\hat{\theta}_{n,r}$, so that we can replace $T^r\hat{\theta}_{n,r}$ with $\hat{\theta}_{n,r}$ in the linear approximation.

Assumption 34. *For all distributions* $F \in \mathcal{P}$*, it holds that*

1. *For each* j*, there exist* $G_{F,j}(\theta) = \nabla_{\theta'}\left(D_{F,j}(\theta)^{-\frac{1}{2}} E_{F,j}\left(m(W,\theta)\right)\right)$ *and its estimator* $\hat{G}_{n,j}(\cdot)$ *such that*

$$\sup_{\theta\in\Theta}\left\|\hat{G}_{n,j}(\theta) - G_{F,j}(\theta)\right\| = O_p\left(\frac{\kappa_n}{\sqrt{n}}\right)$$

uniformly in $\mathcal{P}$*, where* κ_n *is the GMS tuning parameter.*

[3] To give a simple example excluded by Assumption 33, let $W = (Y, X_1, X_2)$. $\mathcal{P} = \{N([0,0,\mu], I_3) : \mu \in (0,1]\}$, and $m(W;\theta) = E[Y - X_1\theta_1 - X_2\theta_2] \geqslant 0$. $\Theta = [-1,1]^2$ and $H_0 : \theta_1 = 0$, it is easy to get $\Theta_I(r,F) = \{0\} \times [-1,0]$. Consider the MMM test statistic. For all $\theta_2 > 0$, $d((0,\theta_2), \Theta_I(r,F)) = \theta_2$ and $S\Big(E_F\big(m(W_i,\theta)\big), \Sigma_F(\theta)\Big) = [-\mu\theta_2]_-^2/(1+\theta_2^2) \leqslant \mu^2\theta_2^2$. In this case, $\chi = 2$ and we can not find c, δ such that Assumption 33 holds.

[4] To give a simple example excluded by Assumption 33, let $E[W] = 1$, $\mathrm{var}(W) = 1$, $\Theta = [-2,2]^2$, $m(W;\theta) = E[W - \theta_1^2 - \theta_2^2] \geqslant 0$ and $H_0 : \theta_1 = 1$. It is easy to get $\Theta_I(r,F) = \{(1,0)\}$. Consider the MMM test statistic. For all $\theta_2 > 0$, $d((1,\theta_2), \Theta_I(r,F)) = \theta_2$ and $S\Big(E_F\big(m(W_i,\theta)\big), \Sigma_F(\theta)\Big) = [-\theta_2^2]_-^2 = \theta_2^4$. In this case, $\chi = 2$ and we can not find c, δ such that Assumption 33 holds.

2. *There exists* $M_G, \bar{M}_G < \infty$, *such that for all* $\theta_1, \theta_2 \in \Theta$,

$$\max_{j=1,\ldots,k} \sup_{F \in \mathcal{P}} \|G_{F,j}(\theta_1) - G_{F,j}(\theta_2)\| \leqslant M_G \|\theta_1 - \theta_2\|$$

and

$$\max_{j=1,\ldots,k} \sup_{F \in \mathcal{P}} \sup_{\theta \in \Theta_I(F)} \|G_{F,j}(\theta)\| \leqslant \bar{M}_G.$$

Assumption 34 requires that the normalized population moments are differentiable and that the derivative can be consistently estimated uniformly in θ and in F. For instance, in a separable model with $D_F(\theta) = D_F$ invariant to θ, we can use

$$\hat{G}_{n,j}(\theta) = \hat{D}_{n,j}^{-\frac{1}{2}} \nabla_{\theta'} \bar{m}_j(\theta).$$

See KMS19 Supplemental Appendix F for such estimators in some canonical moment inequality examples.

Lastly, I give a sufficient primitive assumption for the set $\mathbf{\Delta}_n^r(\theta)$.

Assumption 35. *Set* $\mathbf{\Delta}_n^r(\theta)$ *satisfies:*

1. $0 \in \mathbf{\Delta}_n^r(\theta)$.
2. *Let* $\varepsilon_n = n^{-1/2}(\ln \kappa_n^2)^{1/\chi}$ *where* χ *satisfies Assumption 88.5. There exists* $a > 0$ *such that*

$$\liminf_n \inf_{(r,F) \in \mathcal{L}_0} \inf_{\theta \in \Theta(r) \cap \Theta_I(r,F)^{a\varepsilon_n}} P\left(\mathbf{\Delta}_n^r(\theta) \subseteq \sqrt{n}(\Theta(r) - T^r\theta)\right) = 1. \tag{3.11}$$

$\mathbf{\Delta}_n^r(\theta)$ is a key choice for implementing the method and sets the boundary for the local search to get the critical value. Note that $\Theta_I(r,F)^{a\varepsilon_n}$ is a non-random expansion of $\Theta_I(r,F)$ and

$$\hat{\theta}_{n,r} \in \hat{\Theta}_I(r) \subseteq \Theta(r) \cap \Theta_I(r,F)^{a\varepsilon_n}$$

with probability approaching one. Thus Assumption 35 guarantees that

$$\mathbf{\Delta}_n^r(\hat{\theta}_{n,r}) \subseteq \sqrt{n}(\Theta(r) - T^r\hat{\theta}_{n,r})$$

holds asymptotically. Assumption 35 is trivially satisfied with $\mathbf{\Delta}_n^r(\theta) = \{0_{d_\theta}\}$, but the idea is to choose $\mathbf{\Delta}_n^r(\theta)$ as large as possible, while still satisfying Assumption 35, to improve the power. To give a simple example of $\mathbf{\Delta}_n^r(\theta)$, let $\Theta = [0,1]^2$, $c = (1,0)$ and $\Theta(r) = \{r\} \times [0,1]$. Note that for $\theta = (\theta_1, \theta_2) \in \Theta(r) \cap \Theta_I(r,F)^{\varepsilon_n}$, $d(\theta, T^r\theta) \leqslant \varepsilon_n$, thus

$$\begin{aligned}
\sqrt{n}\left(\Theta(r) - T^r\theta\right) &= \sqrt{n}\left(\Theta(r) - \theta\right) + \sqrt{n}\left(\theta - T^r\theta\right) \\
&= \{0\} \times \left[\sqrt{n}(0 - \theta_2) + \sqrt{n}\big(\theta_2 - (T^r\theta)_2\big), \sqrt{n}(1 - \theta_2) + \sqrt{n}\big(\theta_2 - (T^r\theta)_2\big)\right] \\
&\supseteq \{0\} \times \left[\sqrt{n}(-\theta_2 + \varepsilon_n), \sqrt{n}(1 - \theta_2 - \varepsilon_n)\right]
\end{aligned}$$

where the first two lines are straightforward and the last line follows from $|\theta_2 - (T^r\theta)_2| \leqslant d(\theta, T^r\theta) \leqslant \varepsilon_n$. Hence,

$$\mathbf{\Delta}_n^r(\theta) = \left(\{0\} \times \left[\sqrt{n}(-\theta_2 + \varepsilon_n), \sqrt{n}(1 - \theta_2 - \varepsilon_n)\right]\right) \cup \{0_2\} \tag{3.12}$$

satisfies Assumption 35. Also, note that Assumption 35 allows for data-dependent choices of tuning parameters that enter $\mathbf{\Delta}_n^r(\theta)$ since $\mathbf{\Delta}_n^r(\theta) \subseteq \sqrt{n}(\Theta(r) - T^r\theta)$ only needs to hold with probability approaching one. Section 3.4.1 gives more examples.

Theorem 36. *Let Assumptions 30, 33, 34, 35, 86, 87, 88, 89 hold. Then for all $\alpha \in (0, \frac{1}{2})$,*

$$\limsup_{n\to\infty} \sup_{(r,F)\in\mathcal{L}_0} E_F\left[\phi_n^{\mathrm{LL}}(r)\right] \leqslant \alpha. \tag{3.13}$$

Assumptions 86, 87, 88, 89 can be found in Appendix B.1.2. Assumptions 86 and 87 are borrowed from BCS17. Assumption 86 gives regularity conditions on the baseline distribution space. It is the same as BCS17 Definition 4.2 and thus I refer to their paper for detailed discussion. Assumption 87 is a continuity assumption on the limit distribution of the test statistic. Assumption 88 is the maintained assumption on the test function S. Assumption 89 restricts the rates of tuning parameters.

Remark 37. *The set $\hat{\Theta}_I(r)$ is also used in BCS17 with a similar requirement, and BCS17 Remark 4.1 applies here. The key restriction for set $\hat{\Theta}_I(r)$ is that*

$$\sup_{\theta\in\hat{\Theta}_I(r)} d\left(\theta, \Theta_I(r,F)\right) = o_p\left(n^{-1/2}(\ln \kappa_n^2)^{1/\chi}\right)$$

uniformly, or sufficiently

$$P\left(\hat{\Theta}_I(r) \subseteq \Theta(r) \cap \Theta_I(r,F)^{\varepsilon_n}\right) \to 1$$

uniformly with $\varepsilon_n = an^{-1/2}(\ln \kappa_n^2)^{1/\chi}$ *for some* $a > 0$*. This has two implications. First, we do not need to consistently estimate* $\Theta_I(r,F)$ *and* $\hat{\Theta}_I(r)$ *is not required to contain all the minimizers of* $J_n(r)$*. The intuition is that with fewer elements in* $\hat{\Theta}_I(r)$*, the critical value defined in (3.5) is larger and the coverage is higher. And thus for the asymptotic size control, it suffices to include only one minimizer and construct the critical value by minimizing locally around the single point. Second,* $\hat{\Theta}_I(r)$ *can also be a subset of approximate minimizers defined as*

$$\hat{\Theta}_I(r) \subseteq \left\{\theta \in \Theta(r) : S\big(\hat{D}_n(\theta)^{-\frac{1}{2}}\sqrt{n}\bar{m}_n(\theta), \hat{\Omega}_n(\theta)\big) \leqslant J_n(r) + \delta_n\right\}$$

with $\delta_n \geqslant 0$ *and* $\delta_n = o_p(1)$*. I give detailed guidance on calculating* $\hat{\Theta}_I(r)$ *in Section 3.4.2.*

3.3 Comparison with BCS critical value

BCS17 use the same profiled test statistic as in (3.4). In this section, I compare the LL critical value with BCS critical value in terms of the choice of GMS functions, power properties and computation.

3.3.1 BCS critical value

Let $v_n^*(\cdot)$ defined as in (3.7). And to simplify notation, let

$$Q_n\left(\tilde{\Theta}, \tilde{\varphi}(\cdot), \tilde{\mathbf{\Delta}}(\cdot)\right) = \inf_{\theta \in \tilde{\Theta}} S\left(v_n^*(\theta) + \tilde{\varphi}(\theta) + \hat{G}_n(\theta)\Delta(\theta), \hat{\Omega}_n(\theta)\right) \tag{3.14}$$

where

$$\Delta(\theta) \in \underset{\Delta \in \tilde{\mathbf{\Delta}}(\theta)}{\arg\min}\, S\left(v_n^*(\theta) + \tilde{\varphi}(\theta) + \hat{G}_n(\theta)\Delta(\theta), \hat{\Omega}_n(\theta)\right) + \frac{\lambda_n}{n}\|\Delta\|^2.$$

The LL critical value is given by the $1-\alpha$ quantile of

$$J_n^{\mathrm{LL}}(r) = Q_n\left(\hat{\Theta}_I(r), \varphi(\xi_n(\cdot)), \mathbf{\Delta}_n^r(\cdot)\right). \tag{3.15}$$

The critical value of the minimum resampling test in BCS17 is given by the $1-\alpha$ quantile of

$$J_n^{\mathrm{BCS}}(r) = \min\{J_n^{\mathrm{DR}}(r), J_n^{\mathrm{PR}}(r)\}$$

with

$$J_n^{\mathrm{DR}}(r) = Q_n\left(\hat{\Theta}_I(r), \varphi^{\mathrm{BCS}}(\xi_n(\cdot)), \{0_{d_\theta}\}\right), \tag{3.16}$$

$$J_n^{\mathrm{PR}}(r) = Q_n\left(\Theta(r), \xi_n(\cdot), \{0_{d_\theta}\}\right), \tag{3.17}$$

where $\varphi_n^{\mathrm{BCS}}(\cdot)$ is a GMS function satisfying the following assumption.

Assumption 38. *Given the GMS function $\varphi^{BCS}(\cdot)$, there is a function $\varphi^* : \mathbb{R}^k_{[\pm\infty]} \to \mathbb{R}^k_{[+\infty]}$ that takes the form $\varphi^*(\xi) = \left(\varphi_1^*(\xi_1), \ldots, \varphi_p^*(\xi_p), \mathbf{0}_{k-p}\right)$ and, for all $j = 1, \ldots, p$,*

1. *$\varphi_j^*(\xi_j) \geqslant \varphi_j(\xi_j)$ for all $\xi_j \in \mathbb{R}_{[+\infty]}$*
2. *$\varphi_j^*(\cdot)$ is continuous*
3. *$\varphi_j^*(\xi_j) = 0$ for all $\xi_j \leqslant 0$ and $\varphi_j^*(\infty) = \infty$.*

Assumption 38 is satisfied when φ^{BCS} is any of the functions $\varphi^{(1)} - \varphi^{(4)}$ described in D. W. K. Andrews and Soares (2010). For instance, in the main text and the simulation in BCS17, they use

$$\varphi_j^{(1)}(\xi) = \begin{cases} \infty, & \text{if } \xi_j > 1 \text{ and } j = 1, ..., p \\ 0, & \text{if } \xi_j \leqslant 1, \text{ or } j = p+1, ..., k \end{cases}. \tag{3.18}$$

Here and after, denote Test BCS (resp. DR, PR) as the test with the test statistic in (3.4) and $1-\alpha$ quantile of $J_n^{\mathrm{BCS}}(r)$ (resp. $J_n^{\mathrm{DR}}(r)$, $J_n^{\mathrm{PR}}(r)$) being the critical value.

Comparing (3.15), (3.16) and (3.17), we can see that J_n^{PR} minimizes over the whole null parameter space $\Theta(r)$, but J_n^{LL} and J_n^{DR} only search around $\hat{\Theta}_I(r)$. Compared to Test DR, Test LL further searches around $\hat{\Theta}_I(r)$ by minimizing over Δ, while Test DR imposes $\Delta = 0$.

3.3.2 GMS functions

For Test LL, I only consider $\varphi(\xi)$ defined in (3.8), which is $\varphi^{(3)}$ in D. W. K. Andrews and Soares (2010). Other GMS functions allowing for $\varphi_j(\xi) = \infty$ for $\xi \in \mathbb{R}^k$ may lead to

overrejection. To see this, recall that we use

$$\upsilon_n^*(\hat{\theta}_{n,r}) + \varphi_n(\xi_n(\hat{\theta}_{n,r})) + \hat{G}_n(\hat{\theta}_{n,r})\Delta(\hat{\theta}_{n,r})$$

as a linear approximation to

$$D_F(\theta)^{-\frac{1}{2}}\sqrt{n}\bar{m}_n(\theta) \approx \nu_n(T^r\hat{\theta}_{n,r}) + h_{n,F}(T^r\hat{\theta}_{n,r}) + G_F(\bar{\theta}_{n,r})\sqrt{n}(\theta - T^r\hat{\theta}_{n,r})$$

with $\Delta(\hat{\theta}_{n,r}) = \sqrt{n}(\theta - T^r\hat{\theta}_{n,r}) = o_p(n^{\frac{1}{4}})$. Note that by construction, θ can be far away from $T^r\hat{\theta}_{n,r}$, in the sense that the jth moment inequality condition can be violated at θ even if it is slack at $T^r\hat{\theta}_{n,r}$. To be more specific, even if $h_{n,F,j}(T^r\theta_{n,r})/\kappa_n \to +\infty$, it is not necessary that

$$h_{n,F,j}(T^r\hat{\theta}_{n,r}) + G_{F,j}(\bar{\theta}_{n,r})\sqrt{n}(\theta - T^r\hat{\theta}_{n,r}) \geqslant 0.$$

Hence, we can not simply omit the jth moment by setting $\varphi_j = \infty$. Using a GMS function with $\varphi_j(\xi) < \infty$ for $\xi \in \mathbb{R}^k$, $j = 1, ..., p$ introduces a penalty when violating the inequalities, which mimics the behavior of the profiled test statistic.

This also suggests that the GMS function depends closely on how the feasible set $\mathbf{\Delta}_n^r(\theta)$ was chosen. If we choose a smaller $\mathbf{\Delta}_n^r(\theta)$ so that it belongs to a fixed compact set, then $\Delta(\hat{\theta}_{n,r}) = O_p(1)$ and $\theta = T^r\hat{\theta}_{n,r} + \frac{1}{\sqrt{n}}\Delta(\hat{\theta}_{n,r})$ is close enough to $T^r\hat{\theta}_{n,r}$. Therefore, the jth moment inequality condition is slack at θ as long as it is large enough at $T^r\hat{\theta}_{n,r}$. That is, $h_{n,j}(T^r\theta_{n,r})/\kappa_n > 1$ implies

$$h_{n,F,j}(T^r\theta_{n,r}) + G_{F,j}(\theta_{n,r})\Delta(\theta_{n,r}) \to +\infty.$$

In this case, we can ignore the jth moment in the approximation, and $\varphi^{(1)}$ is a valid choice. Test DR can be viewed as an extreme case here with $\mathbf{\Delta}_n^r(\theta) = \{0_{d_\theta}\}$.

In general, the power increases (weakly) in $\mathbf{\Delta}_n^r(\theta)$ and we want it to be as large as possible. Hence, I work with $\varphi^{(3)}$ and do not impose further boundary restrictions on $\mathbf{\Delta}_n^r(\theta)$.

3.3.3 Power properties

In this section, I compare Test LL with Test DR and Test PR in terms of power properties.

Comparison with Test DR

Test DR is closely related to Test LL. If we assume the GMS functions φ_n are the same in Test LL and Test DR, we have

$$\begin{aligned} J_n^{LL}(r) &= \inf_{\theta \in \hat{\Theta}_I(r)} S\left(\upsilon_n^*(\theta) + \varphi(\xi_n(\theta)) + \hat{G}_n(\theta)\Delta(\theta), \hat{\Omega}_n(\theta)\right) \\ &\leqslant \inf_{\theta \in \hat{\Theta}_I(r)} \left(S\big(\upsilon_n^*(\theta) + \varphi(\xi_n(\theta)) + \hat{G}_n(\theta)\Delta(\theta), \hat{\Omega}_n(\theta)\big) + \frac{\lambda_n}{n}||\Delta(\theta)||^2\right) \\ &\leqslant \inf_{\theta \in \hat{\Theta}_I(r)} S\big(\upsilon_n^*(\theta) + \varphi(\xi_n(\theta)), \hat{\Omega}_n(\theta)\big) = J_n^{\mathrm{DR}}(r) \end{aligned}$$

where the first line is by (3.5), the first inequality is straightforward and the second one follows from (3.6) and $0 \in \boldsymbol{\Delta}_n^r(\theta)$. Thus the power of Test LL is greater than Test DR for all sample sizes and all alternatives.

Remark 39. *It is also common practice to use a projection-based method in applied work. The projection-based method first computes the confidence set for the entire parameter θ by*

$$\mathrm{CS}_n^\theta(1-\alpha) \equiv \left\{\theta \in \Theta : S\big(\hat{D}_n(\theta)^{-\frac{1}{2}}\sqrt{n}\bar{m}_n(\theta), \hat{\Omega}_n(\theta)\big) \leqslant cv_n^\theta(\theta, 1-\alpha)\right\}$$

where $cv_n^\theta(\theta, 1-\alpha)$ is such that $\mathrm{CS}_n^\theta(1-\alpha)$ has the correct asymptotic size; and then rejects H_0 if $\Theta(r) \cap \mathrm{CS}_n^\theta(1-\alpha) = \varnothing$. BCS17 proves that Test DR has higher power than this projection-based test in finite samples for all alternatives. Thus, for all samples, it follows that $\phi_n^{\mathrm{BP}}(r) \leqslant \phi_n^{\mathrm{DR}}(r) \leqslant \phi_n^{\mathrm{LL}}(r)$, where $\phi_n^{\mathrm{BP}}(r)$, $\phi_n^{\mathrm{DR}}(r)$ and $\phi_n^{\mathrm{LL}}(r)$ be the projection-based test, Test DP in BCS17 and Test LL, respectively. By Remark 4.6 in BCS17, under a condition similar to Bugni et al. (2015)(BCS15) Assumption A.9, Test DR has strictly higher asymptotic power than the projection-based test. Under the same condition, Test LL has strictly higher asymptotic power than the projection-based test.

Example 40 illustrates a case where Test LL has strictly better asymptotic power than Test DR with MMM test function and the same GMS function φ in (3.8).

Example 40. *Let* $\{W_i\}_{i=1}^n = \{(W_{1,i}, W_{2,i}, W_{3,i})\}_{i=1}^n$ *be an i.i.d. sequence of random variables with distribution* $F_n, \Sigma_{F_n} = I_3, E_{F_n}[W_1] = \mu_1\kappa_n/\sqrt{n}, E_{F_n}[W_2] = \mu_2/\sqrt{n}$, *and* $E_{F_n}[W_3] = 0$ *for some* $\mu_1 > 0, \mu_2 \leqslant 0$. *Consider the following model with* $\Theta = [-1,1]^3$ *:*

$$E_{F_n}[m_1(W_i,\theta)] = E_{F_n}[W_{1,i} - \theta_1] \geqslant 0,$$
$$E_{F_n}[m_2(W_i,\theta)] = E_{F_n}[W_{2,i} - \theta_2] \geqslant 0,$$
$$E_{F_n}[m_3(W_i,\theta)] = E_{F_n}[W_{3,i} - \theta_3] = 0.$$

We are interested in testing the hypotheses

$$H_0 : c'\theta = (\theta_1, \theta_2) = (0,0) \text{ v.s. } H_1 : c'\theta = (\theta_1, \theta_2) \neq (0,0).$$

Below I use $Z = (Z_1, Z_2, Z_3) \sim N(0, I_3)$. *Simple algebra shows that*

$$J_n(r) = \inf_{\theta_3 \in [-1,1]} \left[\sqrt{n}\hat{\sigma}_{n,1}^{-1}\bar{W}_{n,1}\right]_-^2 + \left[\sqrt{n}\hat{\sigma}_{n,2}^{-1}\bar{W}_{n,2}\right]_-^2 + \left(\sqrt{n}\hat{\sigma}_{n,3}^{-1}(\bar{W}_{n,3} - \theta_3)\right)^2$$
$$\xrightarrow{d} [Z_2 + \mu_2]_-^2 .$$

Test LL Let $\hat{\Theta}_I = \{(0,0,\bar{W}_{n,3})\}$. *Let* $\varepsilon_n = \sqrt{\frac{\ln \kappa_n^2}{n}}$ *and*

$$\boldsymbol{\Delta}_n^r(\hat{\theta}_n) = \left\{(0,0,\tilde{\Delta}_3) : \tilde{\Delta}_3 \in \sqrt{n}\left[-1 + \varepsilon_n - \bar{W}_3, 1 - \varepsilon_n - \bar{W}_3\right]\right\} \cup \{0_3\}$$

in the spirit of (3.12). This test uses the (conditional) $(1-\alpha)$ *quantile of*

$$J_n^{LL}(r) = \left[\upsilon_{n,1}^* + \varphi_1(\xi_n(\hat{\theta}_n))\right]_-^2 + \left[\upsilon_{n,2}^* + \varphi_2(\xi_n(\hat{\theta}_n))\right]_-^2 + \left(\upsilon_{n,3}^* - \Delta_3\right)^2$$

$$\Delta_3 \in \underset{\tilde{\Delta}:(0,0,\tilde{\Delta}) \in \boldsymbol{\Delta}_n^r(\hat{\theta}_n)}{\arg\min} \left[\upsilon_{n,1}^* + \varphi_1(\xi_n(\hat{\theta}_n))\right]_-^2 + \left[\upsilon_{n,2}^* + \varphi_2(\xi_n(\hat{\theta}_n))\right]_-^2 + \left(\upsilon_{n,3}^* - \tilde{\Delta}\right)^2 + \frac{\lambda_n}{n}\tilde{\Delta}^2.$$

υ_n^* *is defined in (3.7) and it is invariant to* θ *in this additively separable model, thus I omit the argument. It is straightforward to get* $\Delta_3 = \upsilon_{n,3}^*/(1 + \lambda_n/n)$ *with probability approaching one and*

$$J_n^{LL}(r) \xrightarrow{d} [Z_1 + \mu_1]_-^2 + [Z_2]_-^2 .$$

Test DR We have

$$J_n^{DR}(r) = \left[\upsilon_{n,1}^* + \varphi_1(\xi_n(\hat{\theta}_n))\right]_-^2 + \left[\upsilon_{n,2}^* + \varphi_2(\xi_n(\hat{\theta}_n))\right]_-^2 + \left(\upsilon_{n,3}^*\right)^2$$

$$\xrightarrow{d} [Z_1 + \mu_1]_-^2 + [Z_2]_-^2 + Z_3^2.$$

Thus we have $J_n^{DR}(r) > J_n^{LL}(r)$, and Test LL has strictly higher power than Test DR.

Comparison with Test PR

I first consider a simple separable linear model

$$E[m_j(W;\theta)] = E[W_j + G_j\theta] \geqslant 0 \qquad j = 1, ..., p$$

$$E[m_j(W;\theta)] = E[W_j + G_j\theta] = 0 \qquad j = p+1, ..., k$$

where G_j is $1 \times d_\theta$ nonrandom known vector.

Test PR uses the (conditional) $(1-\alpha)$ quantile of

$$J_n^{\mathrm{PR}}(r) = \inf_{\theta \in \Theta(r)} S\left(\upsilon_n^* + \frac{\sqrt{n}}{\kappa_n}\hat{D}_n^{-\frac{1}{2}}(\bar{W} + G\theta), \hat{\Omega}_n\right) \tag{3.19}$$

$$= \inf_{\hat{\theta}_n \in \hat{\Theta}_I} \inf_{\theta \in \Theta(r)} S\left(\upsilon_n^* + \frac{\sqrt{n}}{\kappa_n}\hat{D}_n^{-\frac{1}{2}}(\bar{W} + G\hat{\theta}_{n,r}) + \frac{\sqrt{n}}{\kappa_n}\hat{D}_n^{-\frac{1}{2}}G(\theta - \hat{\theta}_{n,r}), \hat{\Omega}_n\right) \tag{3.20}$$

$$= \inf_{\hat{\theta}_n \in \hat{\Theta}_I(r)} S\left(\upsilon_n^* + \xi_n(\hat{\theta}_{n,r}) + \hat{D}_n^{-\frac{1}{2}}G\Delta^{\mathrm{PR}}(\hat{\theta}_{n,r}), \hat{\Omega}_n\right) \tag{3.21}$$

where

$$\Delta^{\mathrm{PR}}(\theta) \in \underset{\Delta \in \frac{\sqrt{n}}{\kappa_n}(\Theta(r) - \theta)}{\arg\min} S\left(\upsilon_n^* + \xi_n(\theta) + \hat{D}_n^{-\frac{1}{2}}G\Delta, \hat{\Omega}_n\right).$$

Note that (3.19) follows from the definition in (3.17) and (3.20) simply rearranges terms by adding and subtracting $\frac{\sqrt{n}}{\kappa_n}\hat{D}_n^{-\frac{1}{2}}\hat{\theta}_{n,r}$. In (3.21), I replace the optimal value of $\frac{\sqrt{n}}{\kappa_n}(\theta - \hat{\theta}_{n,r})$ with $\Delta^{\mathrm{PR}}(\hat{\theta}_{n,r})$.

Recall that Test LL uses the (conditional) $(1-\alpha)$ quantile of

$$J_n^{\mathrm{LL}}(r) = \inf_{\hat{\theta}_n \in \hat{\Theta}_I(r)} S\left(\upsilon_n^* + \varphi(\xi_n(\hat{\theta}_{n,r})) + \hat{D}_n^{-\frac{1}{2}}G\Delta^{\mathrm{LL}}(\hat{\theta}_{n,r}), \hat{\Omega}_n\right) \tag{3.22}$$

$$\Delta^{\text{LL}}(\theta) \in \underset{\Delta \in \mathbf{\Delta}_n^r(\theta)}{\arg\min} S\left(\upsilon_n^* + \varphi(\xi_n(\theta)) + \hat{D}_n^{-\frac{1}{2}} G\Delta, \hat{\Omega}_n\right) + \frac{\lambda_n}{n}||\Delta||^2. \tag{3.23}$$

The differences between (3.21) and (3.22) are the slackness function and the choice of Δ. The two slackness functions are asymptotically equivalent under the null and local alternatives, in the sense that

$$\xi_n(\hat{\theta}_{n,r}) = \varphi_n(\xi_n(\hat{\theta}_{n,r})) + o_p(1).$$

Δ^{PR} and Δ^{LL} differ in (i) the feasible set, $\Delta^{\text{LL}}(\hat{\theta}_{n,r}) \in \mathbf{\Delta}_n^r(\hat{\theta}_{n,r}) \subseteq \sqrt{n}(\Theta(r) - T\hat{\theta}_{n,r})$, while $\Delta^{\text{PR}} \in \frac{\sqrt{n}}{\kappa_n}\left(\Theta(r) - \hat{\theta}_{n,r}\right)$; (ii) the maximum divergence rate, $\Delta^{\text{LL}} = O_p(\sqrt{\frac{n}{\lambda_n}}) = o(n^{\frac{1}{4}})$ due to the penalty term in the objective function, while $\Delta^{\text{PR}} = O_p(\frac{\sqrt{n}}{\kappa_n})$ due to the feasible set. However, if $\hat{\theta}_{n,r}$ is far away from the boundary of Θ and $\Delta^{\text{PR}}, \Delta^{\text{LL}} = O_p(1)$, the feasible set constraint is not binding. In this case, Test LL and Test PR have the same asymptotic power under the null and the local alternatives. I use Example 40 to illustrate.

Example 41 (continued)**.** *Test PR.*

$$\begin{aligned} J_n^{PR}(r) = \inf_{\theta_3 \in [-1,1]} & \left[\upsilon_{n,1}^* + \frac{\sqrt{n}}{\kappa_n}\hat{\sigma}_{n,1}^{-1}\bar{W}_{n,1}\right]_-^2 + \left[\upsilon_{n,2}^* + \frac{\sqrt{n}}{\kappa_n}\hat{\sigma}_{n,2}^{-1}\bar{W}_{n,2}\right]_-^2 \\ & + \left(\upsilon_{n,3}^* + \frac{\sqrt{n}}{\kappa_n}\hat{\sigma}_{n,3}^{-1}(\bar{W}_{n,3} - \theta_3)\right)^2 \\ & \xrightarrow{d} [Z_1 + \mu_1]_-^2 + [Z_2]_-^2 \end{aligned}$$

with $\Delta^{PR} = \hat{\sigma}_{n,3}\upsilon_{n,3}^* = O_p(1)$*. The asymptotic distribution of* $J_n^{PR}(r)$ *is the same as* $J_n^{LL}(r)$*. Note that in this example,*

$$\hat{\theta}_{n,r} = (0, 0, \bar{W}_{n,3}) \xrightarrow{p} 0_3$$

is an interior point of Θ *and recall that* $\Delta^{LL} = \left(0, 0, \upsilon_{n,3}^*/(1 + \lambda_n/n)\right) = O_p(1)$*. As expected, under the null (*$\mu_2 = 0$*) and local alternative (*$\mu_2 < 0$*), Test LL and Test PR have the same asymptotic power.*

However, in a non-separable or nonlinear model, the previous comparison does not hold as the random process υ_n^*, Jacobian $\hat{G}$ and correlation matrix $\hat{\Omega}$ change with θ, in which

case the power comparison of Test PR and Test LL is unclear. Below I give one example where Test LL is dominated by Test PR.

Example 42. *Let* $\{W_i\}_{i=1}^n = \{(W_{1,i}, W_{2,i}, W_{3,i})\}_{i=1}^n$ *be an i.i.d. sequence of random variables with distribution* $F_n, \Sigma_{F_n} = I_3, E_{F_n}[W_1] = E_{F_n}[W_2] = 2 - \mu/\sqrt{n}, E_{F_n}[W_3] = 0$ *for some* $\mu \geqslant 0$. *Consider the following model with* $\Theta = [-2,2]^2$ *:*

$$E_{F_n}[m_1(W_i,\theta)] = E_{F_n}[W_{1,i} - \theta_1 + \theta_2] \geqslant 0,$$
$$E_{F_n}[m_2(W_i,\theta)] = E_{F_n}[W_{2,i} - \theta_1 - \theta_2] \geqslant 0,$$
$$E_{F_n}[m_3(W_i,\theta)] = E_{F_n}\left[W_{3,i} - \theta_1 + 1 + \eta\left(\theta_2^2 - 1\right)\right] = 0,$$

and $\eta > 0$ *is a known constant that measures the nonlinearity. We are interested in testing the hypotheses*

$$H_0 : c'\theta = \theta_1 = 1 \text{ v.s. } H_1 : c'\theta = \theta_1 \neq 1.$$

Below I use $Z = (Z_1, Z_2, Z_3) \sim N(0, I_3)$. *Under* H_0 *with* $\mu = 0$, *the identified set is* $\Theta_I(1) = \{(1,1),(1,-1)\}$. *Under local alternatives,*

$$J_n(r) = \min_{\theta_2 \in [-2,2]} n\left[\frac{\bar{W}_1 - 1 + \theta_2}{\hat{\sigma}_1}\right]_-^2 + n\left[\frac{\bar{W}_2 - 1 - \theta_2}{\hat{\sigma}_2}\right]_-^2 + n\left(\frac{\bar{W}_3 + \eta\left(\theta_2^2 - 1\right)}{\hat{\sigma}_3}\right)^2$$
$$\xrightarrow{d} \min_{j=1,2} \frac{[2\eta(Z_j - \mu) + Z_3]_-^2}{4\eta^2 + 1}$$

where $\frac{[2\eta(Z_1-\mu)+Z_3]_-^2}{4\eta^2+1}$ *is the local minimum around* $\theta = (1,-1)$ *and* $\frac{[2\eta(Z_2-\mu)+Z_3]_-^2}{4\eta^2+1}$ *is the local minimum around* $\theta = (1,1)$. *If the two local minimums are both zero, we have two minimizers, otherwise, the minimizer is unique. The calculation details are given in Appendix* **??**.

Test PR . We can show that

$$J_n^{PR}(1) \big| W_n \xrightarrow{d} \min_j \frac{[2\eta Z_j + Z_3]_-^2}{4\eta^2 + 1}.$$

Test LL. In this simple example, I use

$$\hat{\Theta}_I(r) = \underset{\theta \in \Theta(r)}{\arg\min}\, n\left[\frac{\bar{W}_1 - 1 + \theta_2}{\hat{\sigma}_1}\right]_-^2 + n\left[\frac{\bar{W}_2 - 1 - \theta_2}{\hat{\sigma}_2}\right]_-^2 + n\left(\frac{\bar{W}_3 + \eta\left(\theta_2^2 - 1\right)}{\hat{\sigma}_3}\right)^2$$

which is the set of all minimizers.[5] *The bootstrapped test statistic* $J_n^{LL}(1)$ *depends on* $\hat{\Theta}_I(1)$. *If* $\hat{\Theta}_I(1)$ *is singleton with* $\hat{\theta}_2$ *close to* -1, *we let* $\delta_n = 1$ *and*

$$J_n^{LL}(1)\big|\,\{\delta_n = 1, W_n\} \xrightarrow{d} \frac{[2\eta Z_1 + Z_3]_-^2}{4\eta^2 + 1}.$$

If $\hat{\Theta}_I(1)$ *is singleton with* $\hat{\theta}_2$ *close to* 1, *we let* $\delta_n = 2$ *and*

$$J_n^{LL}(1)\big|\,\{\delta_n = 2, W_n\} \xrightarrow{d} \frac{[2\eta Z_2 + Z_3]_-^2}{4\eta^2 + 1}.$$

And if $\hat{\Theta}_I(1)$ *contains more than one elements, we let* $\delta_n = 3$ *and*

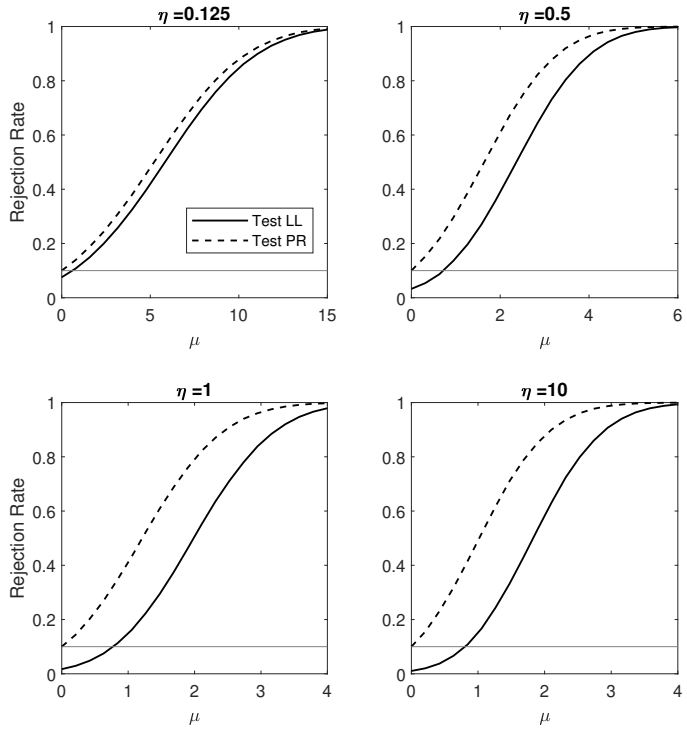
$$J_n^{LL}(1)\big|\,\{\delta_n = 3, W_n\} \xrightarrow{d} \min_{j=1,2} \frac{[2\eta Z_j + Z_3]_-^2}{4\eta^2 + 1}.$$

See Appendix **??** *for a specific form of* δ_n.

For all $\eta > 0$, *the asymptotic distribution of* $J_n^{LL}(1)$ *first order stochastically dominates* $J_n^{PR}(1)$ *and Test PR has strictly higher power than Test LL. The power loss comes from the fact that we are unable to consistently estimate* $\Theta_I(r)$ *at a fast enough rate, and thus, with positive probability,* $\hat{\Theta}_I(r)$ *can only approximate a strict subset of* $\Theta_I(r)$. *In Figure 3.1, I plot the local power of Test LL and Test PR for* η *from 0.125 to 10. We can see that the power loss increases with the nonlinearity* η, *especially for* η *smaller than one. However, when* η *changes from one to ten, the power curves do not change significantly, and thus the power loss stablizes for large* η.

[5] The asymptotic power is the same if we use $\hat{\Theta}_I(r)$ which includes only one minimizer. The reason is that if there are two distinct minimizers, then $J_n(r) = 0$ with probability approaching one. In this case, we do not reject H_0 and the critical value does not matter.

FIGURE 3.1: Example 42 local power with coverage rate $1 - \alpha = 0.9$



3.3.4 Computational properties

To test $H_0 : c'\theta_0 = r$ for a given r, the most computationally demanding step in BCS17 is Test PR. First, Test PR bootstraps a statistic that minimizes a nonlinear objective function. Second, we need to calculate $\upsilon^*_{n,j}(\theta)$ for (a large number of) $\theta \in \Theta(r)$. This is computationally expensive as it requires evaluating the moment function $m(W_i, \theta)$ for all observations $i = 1, ..., n$. Test LL addresses the first point by approximating the objective function and the constraints by linear counterparts so that the inner minimization problem in (3.6) is simply quadratic programming. And Test LL addresses the second point by conducting the outer minimization in (3.5) over $\hat{\Theta}_I(r)$, so that we only need to calculate $\upsilon^*_{n,j}(\theta)$ for $\theta \in \hat{\Theta}_I(r)$, which is much smaller than $\Theta(r)$.

Note that both Test DR and Test LL involve a potentially nonlinear constraint $\hat{\Theta}_I(r)$, which seems computationally costly. However, $\hat{\Theta}_I(r)$ is not required to contain all the

minimizers and the number of elements in $\hat{\Theta}_I(r)$ trades off the computation and power properties. In practice, $\hat{\Theta}_I(r)$ usually contains a finite number of points, and taking minimization over $\hat{\Theta}_I(r)$ can be computationally easy.

For both Test LL and BCS, the confidence intervals are constructed by inverting the test over a grid of r and the computing time increases linearly in the number of grid points. BCS17 calculate the test on evenly spaced points in their simulation. An alternative method is the E-A-M algorithm proposed in KMS19, which improves the implementation of BCS17 by a factor of roughly 4 in the KMS19 simulation. In Section 3.5, I implement both Test BCS and Test LL with the E-A-M algorithm.

The computational gain of Test LL comes at the cost of potential power loss and additional user-chosen tuning parameters. To compute the critical value, Test PR searches over the null set $\Theta(r)$ while Test LL only searches locally around $\hat{\Theta}_I(r)$. Thus we may sacrifice the power compared to Test PR if $\hat{\Theta}_I(r)$ includes only a few elements or if it can not be a consistent estimator for $\Theta_I(r)$, especially when the dimension of θ is relatively high and $\Theta_I(r)$ is large. As for the tuning parameters, in addition to the GMS parameter κ_n, which is also used in BCS, LL needs (i) the penalty parameter λ_n and (ii) the set $\mathbf{\Delta}_n^r(\theta)$ which is a local approximation of $\sqrt{n}(\Theta(r) - T^r\theta)$. To reduce the number of tuning parameters, I suggest setting $\lambda_n = \sqrt{n}\kappa_n^2$ and constructing $\mathbf{\Delta}_n^r(\theta)$ based on shrinking $\Theta(r)$ by $a\varepsilon_n \propto \sqrt{\frac{\ln \kappa_n^2}{n}}$, see Section 3.4.1 for detailed discussion.

3.4 Implementation

In this section, I illustrate the implementation details of constructing $\mathbf{\Delta}_n^r(\theta)$ in (3.6), $\hat{\Theta}_I(r)$ in (3.5), and give a step by step algorithm to construct the confidence interval.

3.4.1 Examples of $\Delta_n^r(\theta)$

Recall that Assumption 35 requires that for some $a > 0$,

$$0 \in \mathbf{\Delta}_n^r(\theta) \subseteq \sqrt{n}(\Theta(r) - T^r\theta)$$

for all $\theta \in \Theta(r) \cap \Theta_I(r,F)^{a\varepsilon_n}$ with $\varepsilon_n = n^{-1/2}(\ln \kappa_n^2)^{\frac{1}{\chi}}$ with probability approaching one. Note that

$$\sqrt{n}(\Theta(r) - T^r\theta) = \sqrt{n}(\Theta(r) - \theta) + \sqrt{n}(\theta - T^r\theta),$$

thus the main idea is to shrink the boundary of $\sqrt{n}(\Theta(r) - \theta)$ by an "upper bound" of $\sqrt{n}(\theta - T^r\theta)$.

First, for a general Θ, define

$$\overline{\Theta(r)} = \left\{\theta \in \mathbb{R}^{d_\theta} : c'\theta = r, \theta \notin \Theta\right\}$$

to be the set of parameters satisfying the subvector restriction but not in Θ, and define

$$\Theta(r)^{-\varepsilon_n} = \left\{\theta \in \Theta(r) : d\left(\theta, \overline{\Theta(r)}\right) \geqslant \varepsilon_n\right\}$$

to be the set of parameters in $\Theta(r)$ but away from the boundary of Θ. A valid choice is

$$\mathbf{\Delta}_n^r(\theta) = \sqrt{n}\left(\Theta(r)^{-\varepsilon_n} - \theta\right) \cup \{0_{d_\theta}\}. \tag{3.24}$$

Assumption 35.1 holds trivially. To verify Assumption 35.2. For all $\theta \in \Theta(r) \cap \Theta_I(r,F)^{\varepsilon_n}$, $\left\|T^r\theta - \theta\right\| \leqslant \varepsilon_n$, thus $\Theta(r)^{-\varepsilon_n} \subseteq \Theta(r) - (T^r\theta - \theta)$, which implies that

$$\sqrt{n}\big(\Theta(r)^{-\varepsilon_n} - \theta\big) \subseteq \sqrt{n}\left(\Theta(r) - (T^r\theta - \theta) - \theta\right) = \sqrt{n}\left(\Theta(r) - T^r\theta\right). \tag{3.25}$$

Example 43. *Let* $\Theta = \left\{\theta : \underline{\theta}_\jmath \leqslant \theta_\jmath \leqslant \bar{\theta}_\jmath, \jmath = 1, ..., v\right\}$. *We are interested in testing* $H_0 : c'\theta = r$ *with* $c = (1, 0_{v-1})'$. *In this sample,*

$$\Theta(r) = \left\{(r, \theta_2, ..., \theta_v) : \underline{\theta}_\jmath \leqslant \theta_\jmath \leqslant \bar{\theta}_\jmath, \quad \jmath = 2, ..., v\right\},$$

$$\overline{\Theta(r)} = \left\{(r, \theta_2, ..., \theta_v) : \theta_\jmath \in (-\infty, \underline{\theta}_\jmath) \cup (\bar{\theta}_\jmath, +\infty), \quad \jmath = 2, ..., v\right\},$$

$$\Theta(r)^{-\varepsilon_n} = \{r\} \times \prod_{\jmath=2}^{v} \left[\underline{\theta}_\jmath + \varepsilon_n, \bar{\theta}_\jmath - \varepsilon_n\right].$$

And (3.24) is equivalent to

$$\mathbf{\Delta}_n^r(\theta) = \left(\{0\} \times \prod_{\jmath=2}^{v} \left[\sqrt{n}(\underline{\theta}_\jmath - \theta_\jmath + \varepsilon_n), \sqrt{n}(\bar{\theta}_\jmath - \theta_\jmath - \varepsilon_n)\right]\right) \cup \{0_v\}. \tag{3.26}$$

We can also scale ε_n in (3.26) so that the test is invariant to the unit of parameters. That is

$$\mathbf{\Delta}_n^r(\theta) = \left(\{0\} \times \prod_{\jmath=2}^{v} \left[\sqrt{n}(\underline{\theta}_\jmath - \theta_\jmath + \epsilon_{n,\jmath}), \sqrt{n}(\bar{\theta}_\jmath - \theta_\jmath - \epsilon_{n,\jmath}) \right] \right) \cup \{0_v\} \tag{3.27}$$

with $\epsilon_{n,\jmath} = \frac{\bar{\theta}_\jmath - \underline{\theta}_\jmath}{2} \varepsilon_n$. In (3.27), Assumption 35 holds with $a = \min_{\jmath=2,...,v} \frac{\bar{\theta}_\jmath - \underline{\theta}_\jmath}{2}$.

If Θ is a polytope, i.e. $\Theta = \{\theta : A_\jmath \theta \leqslant b_\jmath, \jmath = 1, ..., J\}$ where $A_\jmath$ is a $1 \times d_\theta$ row vector, then we can use

$$\mathbf{\Delta}_n^r(\theta) = \left\{ \Delta : c'\Delta = 0, A_\jmath \Delta \leqslant \sqrt{n} \max(b_\jmath - A_\jmath \theta - ||A_\jmath|| \varepsilon_n, 0), \jmath = 1, ..., J \right\}. \tag{3.28}$$

It is straightforward to verify that $0 \in \mathbf{\Delta}_n^r(\theta)$ for all $\theta \in \Theta(r)$ and $\mathbf{\Delta}_n^r(\theta) = \sqrt{n}(\Theta(r) - T^r\theta)$ for all $\theta \in \Theta(r) \cap \Theta_I(r, F)^{\varepsilon_n}$. I illustrate (3.28) with the following two examples.

Example 44. *Let $\Theta = \{(\theta_1, \theta_2, \theta_3) : \theta_1 \geqslant 0, \theta_2 \geqslant 0, \theta_1 + \theta_2 \leqslant 1, 0 \leqslant \theta_3 \leqslant 1\}$. We are interested in testing $H_0 : c'\theta = r$ with $c = (0, 0, 1)$ and $r = 0$. Assume*

$$\Theta_I(r) = \left\{ (\frac{1}{2}, \theta_2, 0) : 0 \leqslant \theta_2 \leqslant \frac{1}{2} \right\}.$$

In this case, $A_1 = (-1, 0, 0)$, $A_2 = (0, -1, 0)$, $A_3 = (1, 1, 0)$, $A_4 = (0, 0, 1)$, $A_5 = (0, 0, -1)$ and $b = (0, 0, 1, 1, 0)$. $\Theta(r) = \{(\theta_1, \theta_2, 0) : \theta_1 \geqslant 0, \theta_2 \geqslant 0, \theta_1 + \theta_2 \leqslant 1\}$.

For $(\theta_1, \theta_2, 0) \in \Theta_I(r, F)^{\varepsilon_n}$, we can easily calculate $\mathbf{\Delta}_n^r(\theta)$ defined in (3.28)

$$\mathbf{\Delta}_n^r(\theta) = \Big\{ (\Delta_1, \Delta_2, 0) : \Delta_1 \geqslant \sqrt{n} min(\varepsilon_n - \theta_1, 0), \Delta_2 \geqslant \sqrt{n} min(\varepsilon_n - \theta_2, 0),$$
$$\Delta_1 + \Delta_2 \leqslant \sqrt{n} \max(1 - \sqrt{2}\varepsilon_n - \theta_1 - \theta_2, 0) \Big\}.$$

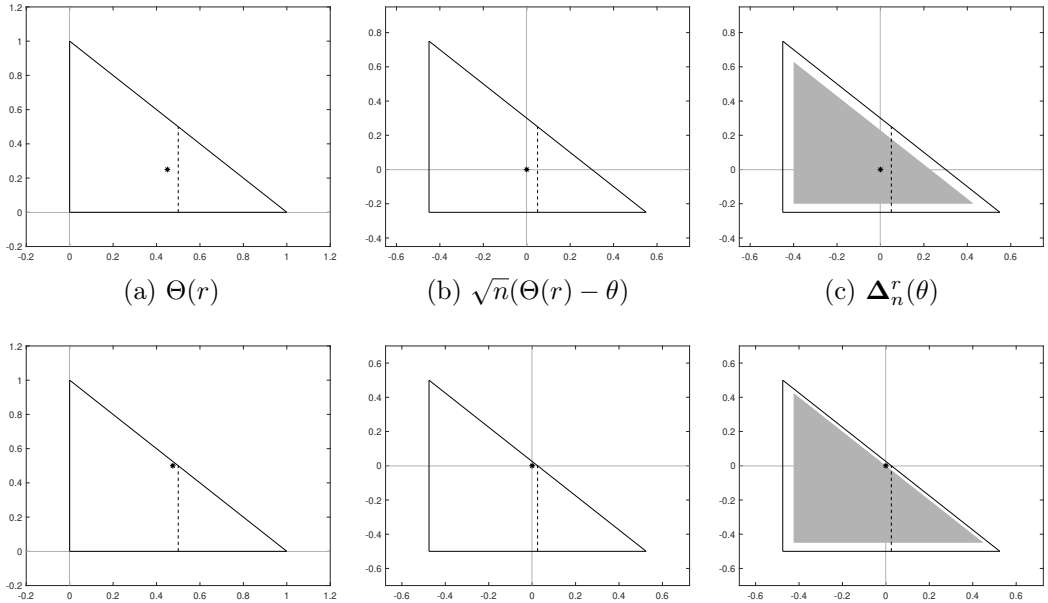
If (θ_1, θ_2) is far away from the boundary of parameter space, for instance, $(\theta_1, \theta_2) = (\frac{1}{2} - \varepsilon_n, \frac{1}{4})$ in Figure 3.2a, then with n large enough,

$$\mathbf{\Delta}_n^r(\theta) = \Big\{ (\Delta_1, \Delta_2, 0) : \Delta_1 \geqslant \sqrt{n}(\varepsilon_n - \theta_1), \Delta_2 \geqslant \sqrt{n}(\varepsilon_n - \theta_2),$$
$$\Delta_1 + \Delta_2 \leqslant \sqrt{n} \left(1 - (\sqrt{2}\varepsilon_n + \theta_1 + \theta_2) \right) \Big\}.$$

This is $\sqrt{n}(\Theta(r)-\theta)$ with each side of the boundary shrunk by width $\sqrt{n}\varepsilon_n$, see Figure 3.2c. However, if (θ_1,θ_2) is close to the boundary of parameter space, for instance, $(\theta_1,\theta_2) = (\frac{1}{2}-\frac{\varepsilon_n}{2},\frac{1}{2})$, then with n large enough,

$$\mathbf{\Delta}_n^r(\theta) = \left\{(\Delta_1,\Delta_2,0) : \Delta_1 \geqslant \sqrt{n}(\varepsilon_n-\theta_1), \Delta_2 \geqslant \sqrt{n}(\varepsilon_n-\theta_2), \Delta_1+\Delta_2 \leqslant 0\right\}.$$

In this case, the hypotenuse boundary is shrunk by $\frac{\sqrt{n}\varepsilon_n}{2\sqrt{2}}$. The intuition is that θ is close to the hypotenuse boundary, thus $T^r\theta$ can not be too far away along direction $(1,1)$. Thus the shrinkage of this bound is smaller, see Figure 3.2f.



(a) $\Theta(r)$ (b) $\sqrt{n}(\Theta(r)-\theta)$ (c) $\mathbf{\Delta}_n^r(\theta)$

(d) $\Theta(r)$ (e) $\sqrt{n}(\Theta(r)-\theta)$ (f) $\mathbf{\Delta}_n^r(\theta)$

FIGURE 3.2: Illustration of $\mathbf{\Delta}_n^r(\theta)$ in Example 44

3.2a is the cross-section plot for the (θ_1,θ_2) at $\theta_3 = 0$. The horizontal axis is θ_1, the vertical axis is θ_2 and the unit is one. The "*" dot is $\theta = (\frac{1}{2}-\varepsilon_n,\frac{1}{4})$. The solid triangle is $\Theta(r)$ and the vertical dash line is $\Theta_I(r)$. 3.2b is 3.2a recentered at θ and scaled by $\sqrt{n}$. The unit for two axes is $\sqrt{n}$. In 3.2c the shaded triangle is $\mathbf{\Delta}_n^r(\theta)$ and the rest is the same as 3.2b. 3.2d, 3.2e and 3.2f are the same, except $(\theta_1,\theta_2) = (\frac{1}{2}-\frac{\varepsilon_n}{2},\frac{1}{2})$.

Example 45 (continued)**.** *In the spirit of (3.27) and (3.28), it is also valid to use*

$$\mathbf{\Delta}_n^r(\theta) = \{0\} \times \prod_{j=2}^{v}\left[\min\left(\sqrt{n}(\underline{\theta}_j-\theta_j+\epsilon_{n,j}),0\right), \max\left(\sqrt{n}(\bar{\theta}_j-\theta_j-\epsilon_{n,j}),0\right)\right] \tag{3.29}$$

with $\epsilon_{n,j} = \frac{\bar{\theta}_j-\underline{\theta}_j}{2}\varepsilon_n$.

3.4.2 Construction of $\hat{\Theta}_I(r)$

As discussed in Remark 37, we can construct $\hat{\Theta}_I(r)$ by

$$\hat{\Theta}_I(r) \subseteq \left\{\theta \in \Theta(r) : S\big(\hat{D}_n(\theta)^{-\frac{1}{2}}\sqrt{n}\bar{m}_n(\theta), \hat{\Omega}_n(\theta)\big) \leqslant J_n(r) + \delta_n\right\} \tag{3.30}$$

with $\delta_n \to 0$ being the tolerance for "approximate minimum". The full set on the right hand side of (3.30) is defined by non-linear restrictions and may contain uncountably many points, but the suggested choice of $\hat{\Theta}_I(r)$ is a subset with only a finite number of "approximate minimizers". In practice, $\hat{\Theta}_I(r)$ can be the set of minimizers in (3.5) generated by different initial values. For instance, we can run the MultiStart minimization algorithm in MATLAB with $\bar{K} \in \mathbb{N}$ different initial values, where $\bar{K}$ is the maximum number of elements. The algorithm will return the minimum $J_n(r)$ and $K \leqslant \bar{K}$ distinct minimizers. This method avoids grid search or nonlinear restrictions.

Note that when $J_n^{\mathrm{LL}}(r)$ is large, we can further improve the efficiency by starting with a singleton $\hat{\Theta}_I(r)$. Specifically, we first calculate $J_n(r)$ and simultaneously get at least one minimizer $\hat{\theta}_{n,r}$. Then we can use this single value to calculate the critical value cv_1^{LL}. If $J_n(r) > cv_1^{\mathrm{LL}}$, there is no need to search for more elements in $\hat{\Theta}_I(r)$ and recalculate the critical value, since the null hypothesis is rejected anyway. If $J_n(r) \leqslant cv_1^{\mathrm{LL}}$, we use MultiSart with $\bar{K}$ initial values, and use $\hat{\Theta}_I(r)$ as the collection of K minimizers and recalculate cv_K^{LL}. The testing result is given by

$$\phi(r) = \mathbf{1}\left[J_n^{\mathrm{LL}}(r) > cv_K^{\mathrm{LL}}\right].$$

In Section 3.4.3, I give a detailed algorithm on how to construct $\hat{\Theta}_I(r)$.

Remark 46. *It plays an important role in the scalability of Test LL that the number of elements in $\hat{\Theta}_I(r)$ is finite and does not grow rapidly with d_θ. $\hat{\Theta}_I(r)$ is chosen by the researcher depending on the dimension of parameters, computational ability, and desired power. This allows the researcher to balance flexibly the computational time and the length of the confidence interval, which is not an obvious option for the existing methods. When*

the dimension of parameters is relatively large, we may use a larger $\bar{K}$ to improve power. Moreover, the computing time increases much slower than a multiple of $\bar{K}$ for the following two reasons. First, as discussed in the previous paragraph, a singleton $\hat{\Theta}_I(r)$ is sufficient when $J_n(r) > cv_1^{LL}(r)$. Thus $\hat{\Theta}_I(r)$ with $K > 1$ elements is only needed for r close to or inside the confidence interval. Second, the MultiStart minimization algorithm and the calculation of $\Delta^{LL}(\theta)$ in (3.23) can be easily parallelized. Therefore, the LL inference procedure has scalability with the sample and model size even with a relatively large $\hat{\Theta}_I(r)$.

3.4.3 A step by step algorithm

In this section, I give step by step algorithm to compute the confidence interval by inverting Test LL. For completeness, I list below the notation used in this section. Define

$$\begin{array}{l} \bar{m}_n(\theta) := \frac{1}{n}\sum_{i=1}^n m(W_i,\theta), \\ \hat{\Sigma}_n(\theta) := \frac{1}{n}\sum_{i=1}^n \big(m(W_i,\theta) - \bar{m}_n(\theta)\big)\big(m(W_i,\theta) - \bar{m}_n(\theta)\big)', \\ \hat{D}_n(\theta) := \operatorname{diag}\big(\hat{\Sigma}_n(\theta)\big), \\ \hat{\Omega}_n(\theta) := \hat{D}_n(\theta)^{-\frac{1}{2}}\hat{\Sigma}_n(\theta)\hat{D}_n(\theta)^{-\frac{1}{2}}. \end{array} \tag{3.31}$$

Let $\kappa_n = \sqrt{\ln(n)}$, $\lambda_n = \sqrt{n}\kappa_n^2$, $\varepsilon_n = \sqrt{\frac{\ln \kappa_n^2}{n}}$ and

$$S\big(m,\Sigma\big) = \sum_{j=1}^{p}\Big[m_j/\Sigma_{jj}^{\frac{1}{2}}\Big]_-^2 + \sum_{j=p+1}^{k}\Big[m_j/\Sigma_{jj}^{\frac{1}{2}}\Big]^2.$$

And $\boldsymbol{\Delta}_n^r(\theta)$ is a set satisfying Assumption 35, for example,

$$\boldsymbol{\Delta}_n^r(\theta) = \sqrt{n}\big(\Theta(r)^{-\varepsilon_n} - \theta\big) \cup \{0_{d_\theta}\}.$$

And I assume there is a grid search algorithm R that takes $\big\{r_\jmath, J_n(r_\jmath), cv_n^{\mathrm{LL}}(r_\jmath, 1-\alpha)\big\}_{\jmath=1}^h$ as input and returns r_{h+1} and s_{h+1} as the next evaluating point and a stopping indicator, respectively. If $s_{h+1} = 1$ then stop the grid search. For instance, the simplest evenly spaced grid search with $K+1$ grid points over $[\underline{r},\ \bar{r}]$ is given by

$$R\left(\big\{r_\jmath, J_n(r_\jmath), cv_n^{\mathrm{LL}}(r_\jmath, 1-\alpha)\big\}_{\jmath=1}^h\right) := \left[\underline{r} + \frac{\bar{r}-\underline{r}}{K} \times (h-1),\ \mathbf{1}\left[h \geqslant K+1\right]\right].$$

The E-A-M algorithm in KMS19 approximates the difference between the test statistic and critical value through a probabilistic model that is updated during the grid search. And it picks the most relevant points as the next point to evaluate. The algorithm stops when the grid search converges, or the maximum iteration is achieved. Lastly, set δ_n to be a small non-negative number, e.g. $\min\left\{\frac{1}{n}, 10^{-4}\right\}$, for the tolerence of the approximate minimizer set $\hat{\Theta}_I(r_h)$ in (3.30).

1. Initialize $h = 1$, $r_1 \in c'\Theta$ and $s_1 = 0$.
2. Estimate the test statistic by minimizing the test function

$$J_n(r_h) = \inf_{\theta \in \Theta(r_h)} S\big(\hat{D}_n(\theta)^{-\frac{1}{2}} \sqrt{n}\bar{m}_n(\theta), \hat{\Omega}_n(\theta)\big),$$

   and get minimizer $\hat{\theta}_{n,r_h}$. Let $\hat{\Theta}_I(r_h) = \left\{\hat{\theta}_{n,r_h}\right\}$ and $Singleton = 1$.[6]
3. Calculate the critical value for r_n
   (a) simulate ζ_i^b i.i.d for $i = 1, ..., n$, $b = 1, ..., B$, such that

$$\zeta_i^b | \{W_i\}_{i=1}^n \sim N(0, 1).$$

   (b) for each b, calculate

$$J_n^{LL,b}(r) = \inf_{\theta \in \hat{\Theta}_n(r_h)} S\left(\upsilon_n^{*,b}(\theta) + \varphi(\xi_n(\theta)) + \hat{G}_n(\theta)\tilde{\Delta}(\theta), \hat{\Omega}_n(\theta)\right),$$

   where $\upsilon_n^{*,b}(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{D}_n^{-\frac{1}{2}}(\theta) \left(m(W_i, \theta) - \bar{m}_n(\theta)\right) \zeta_i^b$,

$$\varphi_j(\xi_n(\theta)) = \begin{cases} \left[\frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\kappa_n \hat{\sigma}_{n,j}(\theta)}\right]_+ & \text{if } j = 1, ..., p \\ 0 & \text{if } j = p+1, ..., k, \end{cases}$$

   and $\hat{G}_n$ is a consistent estimator for $\nabla_{\theta'}\left(D_F(\theta)^{-\frac{1}{2}} E_F\left(m(W_i, \theta)\right)\right)$, and

$$\Delta(\theta) \in \underset{\Delta \in \mathbf{\Delta}_n^r(\theta)}{\operatorname{argmin}} \left(S\left(\upsilon_n^{*,b}(\theta) + \varphi(\xi_n(\theta)) + \hat{G}_n(\theta)\Delta, \hat{\Omega}_n(\theta)\right) + \frac{\lambda_n}{n}\|\Delta\|^2 \right).$$

[6] The procedure to start with a singleton set is a pure computational device and does not suffer from a pre-test distortion.

(c) get the $1-\alpha$ quantile of $\left\{J_n^{LL,b}(r_h)\right\}_{b=1}^{B}$, and denoted it as $cv_n^{\mathrm{LL}}(r_h, 1-\alpha)$.

4. If $Singleton = 1$ and $J_n(r_h) \leqslant cv_n^{\mathrm{LL}}(r_h, 1-\alpha)$,

   (a) run a minimization algorithm with $\bar{K}$ initial values (e.g. MultiStart in MATLAB), and let $\tilde{\Theta}_I(r_h)$ be the set of $\bar{K}$ returned minimizers and

$$\hat{\Theta}_I(r_h) = \left\{\theta \in \tilde{\Theta}_I(r_h) : S\big(\hat{D}_n(\theta)^{-\frac{1}{2}}\sqrt{n}\bar{m}_n(\theta), \hat{\Omega}_n(\theta)\big) \leqslant J_n(r_h) + \delta_n\right\}.$$

   $\hat{\Theta}_I(r_h)$ is the set of approximate minimizers and it rules out the local minimizers far from the global minimizer.

   (b) let $Singleton = 0$

   (c) go to Step 3

   Else,

   (a) go to Step 5

5. Let

$$[r_{h+1}, s_{h+1}] = R\left(\left\{r_\jmath, J_n(r_\jmath), cv_n^{\mathrm{LL}}(r_\jmath, 1-\alpha)\right\}_{\jmath=1}^{h}\right).$$

6. If $s_{h+1} = 0$, let $h = h+1$ and return to Step 2. If $s_{h+1} = 1$, return the confidence interval

$$\left[\min_{\jmath \leqslant h: J_n(r_\jmath) \leqslant cv_n^{\mathrm{LL}}(r_\jmath, 1-\alpha)} r_\jmath, \max_{\jmath \leqslant h: J_n(r_\jmath) \leqslant cv_n^{\mathrm{LL}}(r_\jmath, 1-\alpha)} r_\jmath\right].$$

3.5 Monte Carlo Simulations

In this section, I study the size and power properties by Monte Carlo simulations of a simple two-player entry game.

3.5.1 Model

There are two firms $j = 1, 2$ choosing whether to enter a market $i \in \{1, ..., n\}$. The payoff is given by

$$\pi_{j,i} = A_{j,i}\big(e_{j,i} + \delta_j A_{-j,i} + \sum_{q=1}^{d_X} \beta_{j,q} X_{q,i}\big),$$

where $A_{j,i} = 1$ if j chooses to enter market i. There are d_X type markets and $X_{q,i}$ are observed binary market type indicators with distribution $P(X_{q,i} = 1) = p_q = \frac{1}{d_X}$. Assume p_q is known for simplicity. Let

$$(e_{1,i}, e_{2,i})|X_i \sim N\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

The moment conditions are

$$\begin{aligned}
E_F\big[m_{1,q}(W_i,\theta)\big] =& E_F\left[G_\rho\left([-\beta_{1,q},\infty)\times(-\infty,-\beta_{2,q}-\delta_2]\right)p_q - A_{1,i}(1-A_{2,i})X_{q,1}\right] \geqslant 0 \\
E_F\big[m_{2,q}(W_i,\theta)\big] =& E_F\left[A_{1,i}(1-A_{2,i})X_{q,1} - G_\rho\left([-\beta_{1,q},\infty)\times(-\infty,-\beta_{2,q}-\delta_2]\right)p_q\right. \\
&\left. + G_\rho\left([-\beta_{1,q},-\beta_{1,q}-\delta_1)\times(-\beta_{2,q},-\beta_{2,q}-\delta_2]\right)p_q\right] \geqslant 0 \\
E_F\big[m_{3,q}(W_i,\theta)\big] =& E_F\left[(1-A_{1,i})(1-A_{2,i})X_{q,1} - G_\rho\left((-\infty,-\beta_{1,q}]\times(-\infty,-\beta_{2,q}]\right)p_q\right] = 0 \\
E_F\big[m_{4,q}(W_i,\theta)\big] =& E_F\left[A_{1,i}A_{2,i}X_{q,1} - G_\rho\left([-\beta_{1,q}-\delta_1,\infty)\times[-\beta_{2,q}-\delta_2,\infty)\right)p_q\right] = 0
\end{aligned}$$

where $W_i = (A_i, X_i)$ and $G_\rho(A) = P\left((e_1, e_2) \in A\right)$.

To examine the performance of the inference procedures under different model sizes, I conduct the simulation with $d_X = 2, 4, 6$, where $d_\theta = 7, 11, 15$, the number of moment inequalities $p = 4, 8, 12$ and the number of total moment conditions $k = 8, 16, 24$, respectively. The true parameter is $\theta = (\beta_{1,1}, ..., \beta_{1,d_X}, \beta_{2,1}, ..., \beta_{2,d_X}, \delta_1, \delta_2, \rho)$ where

$$\begin{array}{rcrrrrrr}
\beta_1 & = & (0.1, & 1.1, & -0.1, & -0.5, & 0.8, & -0.5), \\
\beta_2 & = & (0.3, & -0.8, & 0, & -0.5, & -0.9, & -0.2), \\
\delta_1 & = & -0.1, & & \delta_2 = & -0.3, & \rho & = 0.4.
\end{array}$$

In the region of multiple equilibria, the probability of selecting $A_i = (1, 0)$ is 0.5. The parameter space is

$$\Theta = \Big\{\theta = (\beta_1, \beta_2, \delta_1, \delta_2, \rho) \in \mathbb{R}^{2d_X+3} : \beta_{j,q} \in \big[-2, 2\big], \delta_j \in [-2, 0]$$

$$\text{for all } \; j = 1, 2, q \in \big\{1, \ldots, d_X\big\}, \rho \in [0, 0.85]\big\}.$$

And we are interested in constructing a 95% confidence interval (CI) for θ_1. The identified set of θ_1 is reported in the first row of Table 3.1a.

3.5.2 Implementation details

I run all methods with $MC = 1,000$ replications and $B = 600$ bootstrap draws on a server with 16 workers.[7] For Test BCS, I use $\kappa_n = \sqrt{\log(n)}$ as recommended by D. W. K. Andrews and Soares (2010), the MMM test function and $\varphi^{(1)}$ in (3.18) as suggested in BCS17. As for Test LL, I use $\lambda_n = \sqrt{n}\kappa_n^2$, and

$$\mathbf{\Delta}_n^r(\theta) = \{0\} \times \prod_{j=2}^{d_\theta} \left[\min\left(0, \sqrt{n}(\underline{\theta}_j - \theta_j + \epsilon_{n,j}\right), \max\left(0, \sqrt{n}(\bar{\theta}_j - \theta_j + \epsilon_{n,j})\right)\right]. \tag{3.32}$$

with $\epsilon_{n,j} = \frac{(\bar{\theta}_j - \underline{\theta}_j)}{2}\varepsilon_n$ and $\varepsilon_n = \sqrt{\frac{\ln(\kappa_n^2)}{n}}$ as in (3.29). And I use the E-A-M algorithm to calculate the confidence intervals for both profiling methods. As for $\hat{\Theta}_I(r)$, I follow BCS17 and calculate the test statistic by running Matlab command $fmincon$ twice with two distinct initial values, and each run returns a minimizer. $\hat{\Theta}_I(r)$ is the set of distinct minimizers, which has at most two elements. Test LL and Test BCS use the same $\hat{\Theta}_I(r)$. In general, larger $\bar{K}$ gives (weakly) larger power at the cost of increasing computing time. The recommended range of $\bar{K}$ is between two to ten, depending on the model size and computational power.

The calibrated projection-based test proposed in KMS has form

$$\left[\inf_{J_n^{\mathrm{KMS}}(\theta) \leqslant cv_n^{\mathrm{KMS}}(\theta)} c'\theta, \sup_{J_n^{\mathrm{KMS}}(\theta) \leqslant cv_n^{\mathrm{KMS}}(\theta)} c'\theta\right],$$

where $J_n^{\mathrm{KMS}}(\theta)$ and $cv_n^{\mathrm{KMS}}(\theta)$ are the KMS test statistic and critical value evaluated at $\theta \in \Theta$. Let m be the number of initial points to use in multi-start, and I calculate the KMS confidence interval using the E-A-M algorithm with three different m. The recommended value of m is between 10 and 100. I run the algorithm with $m = 20, 40, 80$ for each sample and get three confidence intervals. Note that the seeds to generate the original and bootstrapped data are fixed, thus the difference among these three confidence intervals only

[7] BCS17 set $B = 301$ and KMS19 report the results with both $B = 301$ and $B = 1001$. $B = 600$ is a reasonable choice which makes the Monte Carlo simulations computationally feasible. In practice, researchers can use a larger B, e.g. $B = 1000$, to get a more accurate critical value.

comes from the randomness in the E-A-M algorithm. It turns out that the E-A-M algorithm can be sensitive to the choice of m, especially when d_θ is large. For instance, for $n = 1,000$, $d_X = 6$, the difference between the three lower bounds is larger than 0.1 for around 16% of the samples (with the median length of CI around 1). It is also unclear which is the optimal m. For 19% (resp. 19% , 14%) of the samples, $m = 20$ (resp. $m = 40$, $m = 80$) gives a strictly smaller lower bound than the other two by at least 0.005. Therefore, implementing the E-A-M algorithm with different values of m can reduce sensitivity.[8] I take the union of these three confidence intervals as the final one.

3.5.3 Simulation results

In this section, I present the simulation results.

Confidence intervals and coverage rates

The median 95% CI and coverage rate are reported in Table 3.1. The median CI is the median of the lower bounds to the median of the upper bounds. The "Upper" coverage rate is for the upper bound of the identified set of θ_1, and similarly for "Lower" coverage. The median confidence intervals of LL and BCS are similar, and the coverage rates are higher than the nominal value for all model and sample sizes. The coverage rates for KMS with $N = 4,000,\ d_X = 4, 6$ and $N = 1,000,\ d_X =4$ are close to, but slightly lower than, the nominal rate. It could be a result of the sensitivity of the E-A-M algorithm. The KMS median confidence intervals are strictly shorter than BCS and LL.

Computing time

The average computing time is reported in Table 3.2. The computing time of BCS is in the second and third rows, where "add" represents for "additively separable model" and "gen" represents for "general model". Note that in this simulation, the data and the

[8] It may also help to try a combination of different values of other tuning parameters, e.g. the number of initial points and points to keep from expected improvement maximization. Here I keep other tuning parameters at the default values because of the limit of computation.

Table 3.1: Simulation Results: Median CI and Coverage

(a) Median 95% CI

		$d_X = 2$	$d_X = 4$	$d_X = 6$
Identified Set		[0.044, 0.265]	[0.064, 0.167]	[0.071, 0.159]
$N = 1,000$	LL	[-0.134, 0.817]	[-0.294, 0.926]	[-0.489, 1.052]
	BCS	[-0.149, 0.842]	[-0.282, 0.922]	[-0.404, 0.997]
	KMS	[-0.078, 0.737]	[-0.158, 0.795]	[-0.250, 0.839]
$N = 4,000$	LL	[-0.037, 0.721]	[-0.104, 0.773]	[-0.177, 0.823]
	BCS	[-0.046, 0.737]	[-0.110, 0.780]	[-0.164, 0.816]
	KMS	[-0.013, 0.690]	[-0.048, 0.720]	[-0.083, 0.742]

(b) Coverage Rate

		$d_X = 2$		$d_X = 4$		$d_X = 6$	
		Lower	Upper	Lower	Upper	Lower	Upper
$N = 1,000$	LL	99.3	99.9	99.7	100.0	100.0	100.0
	BCS	99.4	99.9	99.3	100.0	98.9	100.0
	KMS	96.6	97.7	94.4	97.7	96.8	98.3
$N = 4,000$	LL	98.4	99.7	99.6	100.0	99.5	100.0
	BCS	99.2	99.6	99.3	99.9	99.3	100.0
	KMS	96.5	97.5	91.6	98.7	92.3	98.7

parameters are additively separable in the moment conditions, and so $\upsilon_n^*(\theta)$ is invariant to θ. "BCS-add" exploits this feature and calculates $\upsilon_{n,j}^*(\theta)$ only once for each bootstrap. I also report the computing time "BCS-gen" using BCS17 code that works with general models, which calculates $\upsilon_n^*(\theta)$ at each θ. "BCS-gen" is more relevant in a non-separable model with similar sample and model sizes. Note that the confidence intervals calculated by these two methods are numerically the same. The computing time of KMS is reported in the fourth row. For each sample, I run the KMS method with three choices of m, and the reported time is the smallest one corresponding to the largest confidence interval.[9]

LL is computationally attractive and takes less than 81 seconds in all models. BCS-gen is very time-consuming and especially sensitive to the sample size. LL improves upon BCS-

[9] For instance, for sample s, if $m = 40$ gives the largest CI, record $t_s = t_s^{(40)}$. If both $m = 20$ and $m = 40$ give the largest CI, $t_s = \min(t_s^{(20)}, t_s^{(40)})$. Here $t_s^{(m)}$ is the computing time for sample s with m being the number of initial points to use in multi-start.

gen by a factor of up to 170 ($d_X = 6$, $n = 4,000$), where a factor of 9 comes from using a local search to reduce the time evaluating $v_n^*(\theta)$ (compare BCS-gen and BCS-add), and a further factor of 19 from the linearization of moment function (compare BCS-add and LL). The computing time of BCS-add and KMS remains approximately the same as n increases but increases significantly with the model size d_X.

LL improves upon the computing time of KMS by a factor up to 27 ($d_X = 6$, $n = 1,000$). Note that both LL and KMS use local linearization, so conducting the test for a given point is fast.[10] However, LL, as a profile-based method, only conducts the test for a grid of r, which is one dimensional, while KMS, as a calibrated projection-based method, conducts the test for a grid of θ, whose highest dimension is 15. The grid search over a relatively large dimension can be time consuming even with the E-A-M algorithm, and thus the computational advantage of LL over KMS is more prominent with larger d_θ.

Table 3.2: Simulation Results: Average Time

		Time (seconds per simulation)			Ratio to LL		
		$d_X = 2$	$d_X = 4$	$d_X = 6$	$d_X = 2$	$d_X = 4$	$d_X = 6$
$N = 1,000$	LL	30	37	55	1	1	1
	BCS-add	344	762	1742	11	21	31
	BCS-gen	897	2502	5037	30	68	91
	KMS	87	549	1481	3	15	27
$N = 4,000$	LL	37	50	81	1	1	1
	BCS-add	373	774	1538	10	16	19
	BCS-gen	1788	5602	13914	49	113	172
	KMS	115	362	1019	3	7	13

In Table 3.3, I decompose the profiling method and report the computing time per grid point with $B = 600$ without parallelization. The computing time of simple grid search using the profiling method is approximately linear in the bootstrap draws B and the number of evaluated grid points h, thus readers can estimate the computing time using simple grid

[10] Here "a given point" means a given r for LL test, or a given θ for KMS test.

search. For instance, for $n = 1000$, $d_X = 2$, $1,000$ bootstrap draws, a grid search of LL over 100 points, paralleled with 4 workers will take approximately

$$(0.47 + \frac{0.13}{600} \times 1000) \times 100/4 = 17 \ s.$$

The computing time of LL critical value is almost negligible compared to BCS.

Table 3.3: Simulation Results: Computing Time for Profiling Method

		Time (seconds per grid point)			Ratio to LL		
		$d_X = 2$	$d_X = 4$	$d_X = 6$	$d_X = 2$	$d_X = 4$	$d_X = 6$
$N = 1,000$							
Test Stat		0.47	1.20	2.47	-	-	-
Critical Val	LL	0.13	0.28	0.47	1	1	1
	BCS-add	147.22	410.42	785.53	1127	1447	1686
	BCS-gen	378.18	1229.71	2447.61	2896	4337	5254
$N = 4,000$							
Test Stat		0.88	2.55	5.24	-	-	-
Critical Val	LL	0.13	0.22	0.38	1	1	1
	BCS-add	192.46	376.35	772.65	1498	1705	2009
	BCS-gen	818.59	2828.71	6187.08	6371	12812	16087

3.6 Empirical Application

In this section, I apply the test procedures to the binary entry game in Kline and Tamer (2016) (KT16). This empirical application is also revisited in Chen et al. (2018), Chesher and Rosen (2020b) and KMS19 based on moment (in)equality models. The goal is to explain the entry decision in the airline industry and the main setting is summarized below.

All airlines are aggregated into two types: the low cost carrier (LCC) and others (OA). The payoff of j type firm entering market i is determined by two exogenous explanatory variables, namely *market presence* ($x_{j,i}^{\text{pres}}$) and *market size* (x_i^{size}), as well as the opponent's decision. The first one is a market- and firm- specific variable that measures the market presence of type j firm in market i. The market size is defined as the population at the

endpoint city and it only depends on the market. Both exogenous variables are simplified to 1 if the value is larger than the median, and 0 otherwise. Assume that the payoff function is given by

$$A_{j,i}\left(\beta_j^{\text{cons}} + \beta_j^{\text{pres}} x_{j,i}^{\text{pres}} + \beta_j^{\text{size}} x_i^{\text{size}} + \delta_j A_{-j,i} + e_{j,i}\right)$$

where A_j is the decision of j type airline, which equals 1 if it enters and 0 otherwise; A_{-j} is the opponent's decision. β and δ are model parameters. A firm chooses to enter a market if the entering payoff is non-negative. $(e_{1,i}, e_{2,i})$ follows a joint normal distribution with correlation coefficient ρ and marginal mean and variance normalized to 0 and 1, respectively. I refer to KT16 for more details and further discussion. The exogenous variables $(x_{1,i}^{\text{pres}}, x_{2,i}^{\text{pres}}, x_i^{\text{size}})$ can take 8 values, and for each value, there are two moment equalities (for the probability of both entering and both not entering) and two moment inequalities (upper and lower bounds for the probability of only OA entering). This in sum gives 32 moment conditions.

The data is from the second quarter of the 2010 Airline Origin and Destination Survey, which contains $n = 7882$ markets. Let $\theta = (\beta_{\text{LCC}}^{\text{cons}}, \beta_{\text{LCC}}^{\text{size}}, \beta_{\text{LCC}}^{\text{pres}}, \beta_{\text{OA}}^{\text{cons}}, \beta_{\text{OA}}^{\text{size}}, \beta_{\text{OA}}^{\text{pres}}, \delta_{\text{LCC}}, \delta_{\text{OA}}, \rho)$ and $\Theta = \times_{i=1}^{9}[\underline{\theta}_i, \bar{\theta}_i]$, where

$$\underline{\theta} = (\ -8, \quad -2, \quad -2, \quad -8, \quad -2, \quad -2, \quad -4, \quad -4, \quad -0.85\ )$$

$$\bar{\theta} = (\ \ 2, \quad\ 3, \quad\ 10, \quad\ 2, \quad\ 3, \quad\ 10, \quad\ 0, \quad\ 0, \quad\ 0.85\ )$$

following from the bound used in KMS19. The bounds on ρ guarantee that Assumption 34.2 holds, and are also imposed in KMS19. I set $\kappa_n = \sqrt{\log(n)}$, $\lambda_n = \sqrt{n}\kappa_n^2$, $\epsilon_{n,\jmath} = \frac{|\bar{\theta}_\jmath - \underline{\theta}_\jmath|}{2}\sqrt{\frac{\ln \kappa_n^2}{n}}$ and for $\jmath$th parameter,

$$\mathbf{\Delta}_{n,\jmath}^{r}(\theta) = \left\{\Delta : \Delta_\jmath = 0, \Delta_\imath = \left[\sqrt{n}(\underline{\theta}_\imath - \theta_\imath + \epsilon_{n,\imath}), \sqrt{n}(\bar{\theta}_\imath - \theta_\imath - \epsilon_{n,\imath})\right], \jmath \neq \imath\right\} \cup \{0_9\} \quad (3.33)$$

as in (3.27). All confidence intervals are calculated with E-A-M algorithms with $B = 500$ bootstrap draws. KMS is calculated with the number of initial points to use in multi start $m = 40$, which is used in the KMS19 code for the same empirical application. The 95% confidence intervals and computing time for each parameter are presented in Table 3.4.

The moment condition is additively separable and the BCS computing time corresponds to "BCS-add" in Section 3.5.

The estimation results are qualitatively consistent with those in KT16 and KMS19. All bounds are informative. The market size (resp. market presence) has a significant positive effect on the payoff, and the effect is larger for OA (resp. LCC) firms. The entry of each airline significantly reduces the opponent's payoff. LL and BCS confidence intervals are very close, which agrees with the findings in the Monte Carlo simulations. However, different from the simulation results, the confidence intervals of KMS are larger than LL and BCS, especially for $\beta_{\text{LCC}}^{\text{cons}}$, δ_1 and ρ. It is possible that the power advantage of calibrated projection and profiling methods is different under different models. The power comparison with KMS is interesting but outside the scope of this essay and I leave it for future research.

The computing time for of LL is significantly less than both BCS and KMS and it takes less than 20 minutes to get all LL confidence intervals, while it takes 2.5 hours for KMS and 7.5 hours for BCS.

Table 3.4: Results for Empirical Application

(a) 95% confidence intervals

	LL	BCS	KMS
$\beta_{\text{LCC}}^{\text{cons}}$	[-1.11,-1.04]	[-1.23,-1.00]	[-2.06,-0.84]
$\beta_{\text{LCC}}^{\text{size}}$	[0.24, 0.35]	[0.23, 0.38]	[0.18, 0.42]
$\beta_{\text{LCC}}^{\text{pres}}$	[1.78, 1.88]	[1.78, 1.92]	[1.74, 1.95]
$\beta_{\text{OA}}^{\text{cons}}$	[0.47, 0.54]	[0.45, 0.54]	[0.40, 0.60]
$\beta_{\text{OA}}^{\text{size}}$	[0.42, 0.50]	[0.41, 0.51]	[0.33, 0.56]
$\beta_{\text{OA}}^{\text{pres}}$	[0.45, 0.52]	[0.43, 0.55]	[0.39, 0.58]
δ_{LCC}	[-1.25,-1.18]	[-1.29,-1.04]	[-1.45,-0.19]
δ_{OA}	[-1.38,-1.29]	[-1.41,-1.18]	[-1.46,-0.77]
ρ	[0.83, 0.85]	[0.80, 0.85]	[0.21, 0.85]

(b) Computing Time

	Time (seconds)			Ratio to LL	
	LL	BCS	KMS	BCS	KMS
$\beta_{\text{LCC}}^{\text{cons}}$	282	5176	2857	18	10
$\beta_{\text{LCC}}^{\text{size}}$	137	3444	435	25	3
$\beta_{\text{LCC}}^{\text{pres}}$	125	1827	327	15	3
$\beta_{\text{OA}}^{\text{cons}}$	108	1433	1963	13	18
$\beta_{\text{OA}}^{\text{size}}$	93	6886	723	74	8
$\beta_{\text{OA}}^{\text{pres}}$	78	2363	285	30	4
δ_{LCC}	101	2575	362	25	4
δ_{OA}	322	2102	956	7	3
ρ	108	1230	321	11	3
Total	1355	27036	8229	20	6

4. Hypothesis tests with a repeatedly singular information matrix

In this essay, We study score-type tests in likelihood contexts in which the nullity of the information matrix under the null is greater than one. [1]

4.1 Introduction

Rao (1948) score test and Silvey (1959) numerically equivalent Lagrange multiplier (LM) version completed the triad of classical hypothesis tests (see Bera and Bilias (2001) for a survey). Given that they only require estimation of the model parameters under the null, in the late 1970's and early 1980's they became the preferred choice for many specification tests which are nowadays routinely reported by econometric software packages (see the surveys by Breusch and Pagan (1980), Engle (1983), and Godfrey (1988)). In addition to computational considerations, which continue to be very relevant for resampling procedures, two other important advantages of LM tests are that (i) rejections provide a clear indication of the specific directions along which modelling efforts should focus, and (ii) they are often easy to interpret as moment tests, so they remain informative for alternatives they are not designed for. Furthermore, under standard regularity conditions, they are asymptotically equivalent to the Likelihood ratio (LR) and Wald tests under the null and sequences of local alternatives, and thus they share their optimality properties.

One of the crucial regularity conditions for a common asymptotic chi-square distribution for these three tests is a full rank information matrix of the unrestricted model parameters evaluated under the null. Nevertheless, there are empirically relevant situations in which this condition does not hold despite the fact that the model parameters are locally identified. In non-linear instrumental variable models, Sargan (1983) referred to those instances in which the expected Jacobian of the influence functions is singular but the expected Jacobian of the linear combinations of their derivatives that span its nullspace has full rank as second-order identified but first-order underidentified. In a likelihood context, a singular

[1] This essay is a joint work with Dante Amengual and Enrique Sentana.

information matrix implies that there is a linear combination of the average scores which is identically 0, at least asymptotically. In their seminal paper, L. F. Lee and Chesher (1986) studied some popular examples of this situation in economics: i) univariate regression models with sample selectivity; ii) stochastic production frontier models; and iii) certain mixture models.[2]

L. F. Lee and Chesher (1986) proposed to replace the LM test by what they called an "extremum" test. Their suggestion was to study the restrictions that the null imposes on higher-order optimality conditions. Often, the second derivative will suffice, but sometimes it might be necessary to study the third or even higher-order ones. They proved the asymptotic equivalence between their extremum tests and the corresponding LR tests under the null and sequences of local alternatives in unrestricted contexts. Using earlier results by D. Cox and Hinkley (1974), this equivalence intuitively follows from the fact that their tests can often be re-interpreted as standard LM tests of a suitable transformation of the parameter whose first derivative is 0 on average such that the new score is no longer so. In contrast, Wald tests are extremely sensitive to reparametrization under these circumstances. Bera et al. (1998) provide some additional insights. In turn, Rotnitzky et al. (2000) rigorously study the asymptotic distribution of the maximum likelihood (ML) estimators in those contexts. Finally, Bottai (2003) looks at the validity of confidence intervals obtained by inverting the three classical test statistics in this setup.

However, in all the existing literature the nullity of the information matrix, q_r say, is assumed to be 1. When the information matrix is repeatedly singular under the null, in the sense that q_r is two or more, the number of second-order derivatives exceeds the number of parameters effectively affected by the singularity by an order of magnitude. The unbalance gets worse when it becomes necessary to look at higher-order derivatives. Unfortunately, in general there is no reparametrization that leads to a regular information matrix.[3]In

[2] In all their examples, in fact, the average score with respect to one of the parameters of the alternative evaluated at the restricted parameter estimators that impose the null is identically 0 in finite samples.

[3] An exception is the multiplicative seasonal Arma model considered in Amengual et al. (2023).

particular, transforming each of the parameters individually along the lines suggested by L. F. Lee and Chesher (1986) does not usually give rise to a test asymptotically equivalent to the LR. On the contrary, different reparametrizations will typically give rise to different test statistics.

The purpose of this essay is precisely to propose a feasible generalization of the L. F. Lee and Chesher (1986) approach in repeatedly singular contexts that leads to tests asymptotically equivalent to the LR, but which only require estimation under the null. Specifically, we propose a generalized extremum test (GET) which typically maximizes an easy to interpret statistic over a space of dimension $q_r - 1$ when all parameters show the same degree of underidentification, thereby simplifying to the L. F. Lee and Chesher (1986) proposal when the nullity is one. More generally, GET is an LR-type test that compares the log-likelihood function under the null to the maximum over q_r dimensions of its lowest-order expansion under the alternative capable of identifying the restricted parameters. In contrast, LR tests require the maximization over the entire parameter space of an unrestricted log-likelihood function which is extremely flat around its maximum when the null hypothesis is true.[4]These computational advantages are particularly pertinent for bootstrap-type inference, which is especially necessary in our context because the common sup-type asymptotic distribution of the GET and LR tests is often non-standard, and the sample sizes required for this distribution to be reliable unusually large.

Repeatedly singular information matrices are not a mere theoretical curiosity. In fact, we illustrate our proposed testing procedure in detail with several examples of interest that arise in economic and finance applications when testing: 1) exogeneous sample selectivity in multivariate regressions; 2) normality against the flexible semi-nonparametric (SNP) family proposed by Gallant and Nychka (1987); 3) a Gaussian copula against another flexible Hermite expansion; and 4) unpredictability in a multiple regressor version of the purely non-linear model considered by Bottai (2003). Further, in Amengual, Bei, and Sentana

[4] Obviously, both procedures require the estimation of the model under the null, but the restricted maximum likelihood estimator is typically available in closed form in many models subject to specification tests.

(2022) and Amengual et al. (2023) we discuss the application of the test proposed in this essay to two additional examples of substantial empirical interest: testing for multivariate normality against a skew normal distribution, and testing for neglected serial correlation in univariate time series models, respectively.

The structure of the rest of this essay is as follows. In section 2 we obtain our theoretical results first in the case in which all the underidentified parameters have the same degree of underidentification, and then when the degree of underidentification may be different for different parameters. Then, in section 3 we discuss the first two aforementioned examples in detail, assessing the finite sample size and power properties of our proposed tests by means of several extensive Monte Carlo exercises. Finally, we conclude in section 4, relegating proofs, the remaining two examples, and some additional results to the appendices.

4.2 Theoretical results

Consider the estimation of the parameter vector $\boldsymbol{\rho}$ characterizing the distribution of an *i.i.d.* random vector $\mathbf{y}$. Let $l_i(\boldsymbol{\rho}) = \ln f(\mathbf{y}_i; \boldsymbol{\rho})$ denote the log-likelihood function contribution from observation i, so that the log-likelihood function of a sample of size n is $\mathcal{L}_n = \sum_{i=1}^{n} l_i(\boldsymbol{\rho})$.[5]In what follows,

$$s_{\rho_j i}(\boldsymbol{\rho}) = \frac{\partial l_i(\boldsymbol{\rho})}{\partial \rho_j}$$

will denote the contribution of observation i to the score with respect to the j^{th} element of $\boldsymbol{\rho}$ and $\mathbf{S}_{\rho_j n}(\rho) = \sum_{i=1}^{n} \mathbf{s}_{\rho_j i}(\rho)$ their sum.

Let us partition $\boldsymbol{\rho}$ into two blocks: 1) $\boldsymbol{\phi}$, which contains the $p \times 1$ vector of parameters estimated under the null; and 2) $\boldsymbol{\theta}$, which is the $q \times 1$ vector of parameters such that the null hypothesis can be written in explicit form as $H_0 : \boldsymbol{\theta} = \mathbf{0}$. Let $\boldsymbol{\rho}^*$, $\hat{\boldsymbol{\rho}}$ and $\tilde{\boldsymbol{\rho}} = (\tilde{\boldsymbol{\phi}}', \mathbf{0}')'$ denote the true value of the parameter vector, its unrestricted ML estimator (UMLE), and the restricted one (RMLE), respectively, so that $\boldsymbol{\rho}^* = (\boldsymbol{\phi}^*, \mathbf{0})$ under H_0. As usual, $|.|$ and

[5] Although we could easily generalize our results to explicitly deal with dependent data by using standard factorizations of the sample log-likelihood function, we maintain independence to simplify the expressions.

$||.||$ denote absolute value and Euclidean norm, respectively. Finally, we use $e_{\min}(\mathbf{A})$ and $e_{\max}(\mathbf{A})$ for the smallest and largest eigenvalues, respectively, of a symmetric square matrix $\mathbf{A}$.

Using this notation, we henceforth assume:

Assumption 47. *(Regularity conditions)*

(47.1) $\boldsymbol{\rho}$ *takes its value in a compact subset* $\mathbf{P}$ *of* $\mathbb{R}^{p+q}$ *that contains an open neighborhood* $\mathcal{N}$ *of the true value* $\boldsymbol{\rho}^*$ *which generates the observations.*

(47.2) Distinct values of $\boldsymbol{\rho}$ *in* $\mathbf{P}$ *correspond to distinct probability distributions.*

(47.3) $E[\sup_{\boldsymbol{\rho}\in\mathbf{P}} |l_i(\boldsymbol{\rho})|] < \infty$.

(47.4) $E[\partial l_i(\boldsymbol{\phi},\mathbf{0})/\partial\boldsymbol{\phi}\cdot\partial l_i(\boldsymbol{\phi},\mathbf{0})/\partial\boldsymbol{\phi}']$ *has full rank under the null for all* $(\boldsymbol{\phi},\mathbf{0}) \in \mathbf{P}$.

The compactness of $\mathbf{P}$ in Assumption 47.1 together with the continuity of $l_i(\boldsymbol{\rho})$ and Assumptions 47.2 and 47.3 guarantee the existence, uniqueness with probability tending to 1, and consistency of both the UMLE $\hat{\boldsymbol{\rho}}$ and the RMLE $\tilde{\boldsymbol{\rho}}$ (see Newey and McFadden 1994, Theorem 2.5). The "open neighborhood" part of Assumption 47.1 is just used to simplify the expressions and their derivation. Extensions to situations in which the true parameters lie at the boundary of the parameter space under the null are feasible, but at the expense of complicating the notation and blurring the message of this essay. Finally, Assumption 47.4 guarantees the convergence of the RMLE at the usual $n^{-\frac{1}{2}}$ rate.

4.2.1 Repeated singularity of the same order

We first consider the case in which q_1 elements of $\boldsymbol{\theta}$ are first-order identified, while the remaining q_r elements are r^{th}-order identified under the null, a concept that will become precisely defined after we introduce Assumption 49 below. Therefore, if we further partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_r')'$, where $q_1 = \dim(\boldsymbol{\theta}_1)$ and $q_r = \dim(\boldsymbol{\theta}_r)$, so that $q = q_1 + q_r$, then the information matrix under H_0 will be such that its top $(p+q_1)\times(p+q_1)$ block is regular and the rest contains zeros. Consequently, its nullity will be precisely q_r. Often, one needs to reparametrize the model to make sure it satisfies these conditions, an issue we discuss in detail in each of the examples that we consider.

Let $\mathbf{j} \in \mathbb{N}^{p+q}$ denote a $(p+q) \times 1$ vector of indices, $\mathbf{j}! = \prod_{i=1}^{p+q} j_i!$,

$$l_i^{[\mathbf{j}]}(\boldsymbol{\rho}) = \frac{1}{\mathbf{j}!} \frac{\partial^{\boldsymbol{\iota}'_{p+q}\mathbf{j}} l_i(\boldsymbol{\rho})}{\partial \boldsymbol{\rho}^{\mathbf{j}}},$$

where $\boldsymbol{\iota}_m$ is a vector of m ones, and $L_n^{[\mathbf{j}]}(\rho) = \sum_{i=1}^n l_i^{[\mathbf{j}]}(\rho)$. Throughout this subsection, we assume the following conditions hold:

Assumption 48. *(Regularity conditions on the derivatives of the log-likelihood function)*
(48.1) With probability 1, the derivatives $l_i^{[\mathbf{j}]}(\boldsymbol{\rho})$ exist for all $\boldsymbol{\rho}$ in $\mathcal{N}$ and $\boldsymbol{\iota}'_{p+q}\mathbf{j} \leqslant 2r$, and they satisfy $E[\sup_{\boldsymbol{\rho}\in\mathcal{N}} |l_i^{[\mathbf{j}]}(\boldsymbol{\rho})|] < \infty$.
(48.2) For $r \leqslant \boldsymbol{\iota}'_{p+q}\mathbf{j} \leqslant 2r$, $E\{[l_i^{[\mathbf{j}]}(\boldsymbol{\rho})]^2\} < \infty$ for all $\boldsymbol{\rho}$ in $\mathcal{N}$.
(48.3) When $\boldsymbol{\iota}'_{p+q}\mathbf{j} = 2r$ there is some function $g(\mathbf{y})$ satisfying $E[g^2(\mathbf{y})] < \infty$ such that with probability 1, $|L_n^{[\mathbf{j}]}(\boldsymbol{\rho}) - L_n^{[\mathbf{j}]}(\boldsymbol{\rho}^\dagger)| \leqslant ||\boldsymbol{\rho} - \boldsymbol{\rho}^\dagger|| \sum_i g(\mathbf{y}_i)$ for all $\boldsymbol{\rho}$ and $\boldsymbol{\rho}^\dagger$ in $\mathcal{N}$.

We borrow Assumptions 48.1–48.3 from Rotnitzky et al. (2000) with some modifications. The main difference is that they require $(2r+1)^{th}$ differentiability for the Taylor expansions they use to analyze the distribution of the MLE, while we only need $2r^{th}$ differentiability to study the asymptotic distribution of our tests. Assumptions 48.1 and 48.3 guarantee the existence of derivatives and the stochastic equicontinuity of the sample mean of $l_i^{[\mathbf{j}]}(\boldsymbol{\rho})$ with $\boldsymbol{\iota}'_{p+q}\mathbf{j} \leqslant 2r$. In turn, Assumption 48.2 allows us to apply a central limit theorem to $l_i^{[\mathbf{j}]}(\boldsymbol{\rho}^*)$.

Let $\boldsymbol{\theta}_r^{\otimes k} = \underbrace{\boldsymbol{\theta}_r \otimes \boldsymbol{\theta}_r \otimes \cdots \otimes \boldsymbol{\theta}_r}_{k \text{ times}}$ denote the k^{th}-order Kronecker power of the $q_r \times 1$ vector $\boldsymbol{\theta}_r$, and define

$$\frac{\partial^k L_n(\boldsymbol{\rho})}{\partial \boldsymbol{\theta}_r^{\otimes k}} = vec\left\{ \frac{\partial}{\partial \boldsymbol{\theta}_r} \left[\frac{\partial^{k-1} L_n(\boldsymbol{\rho})}{\partial \boldsymbol{\theta}_r^{\otimes (k-1)}} \right]' \right\}.$$

Moreover, let

$$I(\boldsymbol{\phi}) = \begin{bmatrix} I_{\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\phi}) & I_{\boldsymbol{\phi}\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & I_{\boldsymbol{\phi}\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \\ I_{\boldsymbol{\theta}_1\boldsymbol{\phi}}(\boldsymbol{\phi}) & I_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & I_{\boldsymbol{\theta}_1\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \\ I_{\boldsymbol{\theta}_r\boldsymbol{\phi}}(\boldsymbol{\phi}) & I_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & I_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \end{bmatrix} = \lim_{n\to\infty} Var\left\{ \frac{1}{\sqrt{n}} \begin{bmatrix} \mathbf{S}_{\boldsymbol{\phi}n}(\boldsymbol{\phi}, \mathbf{0}) \\ \mathbf{S}_{\boldsymbol{\theta}_1 n}(\boldsymbol{\phi}, \mathbf{0}) \\ \partial^r L_n(\boldsymbol{\phi}, \mathbf{0}) / \partial \boldsymbol{\theta}_r^{\otimes r} \end{bmatrix} \middle| \boldsymbol{\phi}, \mathbf{0} \right\}$$

denote the asymptotic covariance matrix of the relevant influence functions, which may be understood as a generalization of the information matrix.

In addition, let

$$\begin{aligned} V_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) &= \begin{bmatrix} V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \\ V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \end{bmatrix} \\ &= \begin{bmatrix} I_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & I_{\boldsymbol{\theta}_1\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \\ I_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & I_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \end{bmatrix} - \begin{bmatrix} I_{\boldsymbol{\theta}_1\boldsymbol{\phi}}(\boldsymbol{\phi}) \\ I_{\boldsymbol{\theta}_r\boldsymbol{\phi}}(\boldsymbol{\phi}) \end{bmatrix} I^{-1}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}) \left[I_{\boldsymbol{\phi\theta}_1}(\boldsymbol{\phi}) \; I_{\boldsymbol{\phi\theta}_r}(\boldsymbol{\phi}) \right]. \end{aligned}$$

denote the asymptotic residual variance of $\mathbf{S}_{\boldsymbol{\theta}_1 n}(\boldsymbol{\phi}, \mathbf{0})$ and $\partial^r L_n(\boldsymbol{\phi}, \mathbf{0}) / \partial \boldsymbol{\theta}_r^{\otimes r}$ after orthogonalizing these influence functions with respect to $\mathbf{s}_{\boldsymbol{\phi}}$.

Assumption 49. *(Rank conditions for $q_r \geqslant 1$)*

(49.1) For all $(\boldsymbol{\phi}, \mathbf{0}) \in \mathbf{P}$,

$$\frac{\partial^{\boldsymbol{\iota}'_{q_r} \mathbf{j}_{\boldsymbol{\theta}_r}} l_i(\boldsymbol{\phi}, \mathbf{0})}{\partial \boldsymbol{\theta}_r^{\mathbf{j}_{\boldsymbol{\theta}_r}}} = \mathbf{0}$$

with probability 1 for all $\mathbf{j}_{\boldsymbol{\theta}_r} = (j_1, ..., j_{q_r})'$ such that $\boldsymbol{\iota}'_{q_r} \mathbf{j}_{\boldsymbol{\theta}_r} \leqslant r - 1$.

(49.2) The asymptotic covariance matrix of the (scaled by $\sqrt{n}$) sample averages of

$$\left\{ \mathbf{s}_{\boldsymbol{\phi} i}(\boldsymbol{\phi}^*, \mathbf{0}), \mathbf{s}_{\boldsymbol{\theta}_1 i}(\boldsymbol{\phi}^*, \mathbf{0}), \boldsymbol{\theta}_r^{\otimes r\prime} \frac{\partial^r l_i(\boldsymbol{\phi}^*, \mathbf{0})}{\partial \boldsymbol{\theta}_r^{\otimes r}} \right\}$$

has full rank for all possible non-zero values of $\boldsymbol{\theta}_r \in \mathbb{R}^{q_r}$ underlying the vector of coefficients $\boldsymbol{\theta}_r^{\otimes r}$ in the linear combination above.

Intuitively, the rationale for looking at

$$\boldsymbol{\theta}_r^{\otimes r\prime} \frac{\partial^r l_i}{\partial \boldsymbol{\theta}_r^{\otimes r}} = \sum_{\boldsymbol{\iota}'_{q_r} \mathbf{j}_{\boldsymbol{\theta}_r} = r} \frac{r!}{\mathbf{j}_{\boldsymbol{\theta}_r}!} \left(\prod_{k=1}^{q_r} \theta_{rk}^{j_k} \right) \frac{\partial^r l_i(\boldsymbol{\phi}, \mathbf{0})}{\partial \boldsymbol{\theta}_r^{\mathbf{j}_{\boldsymbol{\theta}_r}}}$$

is that it coincides with the r^{th}-order term in the expansion of the log-likelihood function. In that respect, note that although the higher order derivatives $\partial^r l_i / \partial \boldsymbol{\theta}_r^{\otimes r}$ will usually contain many repeated elements thanks to the Clairaut-Schwartz-Young's theorem, the rank deficiency condition in Assumption 49.2 applies to the inner product of $\boldsymbol{\theta}_r^{\otimes r}$ with those

influence functions, so the requirement is that those linear combinations of the elements in $\partial^r l_i/\partial\boldsymbol{\theta}_r^{\otimes r}$ be linearly independent of $\mathbf{s}_{\boldsymbol{\phi}i}(\boldsymbol{\phi},\mathbf{0})$ and $\mathbf{s}_{\boldsymbol{\theta}_1 i}(\boldsymbol{\phi},\mathbf{0})$.

Finally, let

$$Q_n(\boldsymbol{\theta}_r,\boldsymbol{\phi}) = \frac{\boldsymbol{\theta}_r^{\otimes r\prime} D_{rn}(\boldsymbol{\phi}) D'_{rn}(\boldsymbol{\phi})\boldsymbol{\theta}_r^{\otimes r}}{\boldsymbol{\theta}_r^{\otimes r\prime}[V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\boldsymbol{\phi}) - V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}(\boldsymbol{\phi})V^{-1}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi})V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_r}(\boldsymbol{\phi})]\boldsymbol{\theta}_r^{\otimes r}}, \tag{4.1}$$

where

$$D_{rn}(\boldsymbol{\phi}) = \frac{\partial^r L_n(\boldsymbol{\phi},\mathbf{0})}{\partial\boldsymbol{\theta}_r^{\otimes r}} - V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}(\boldsymbol{\phi})V^{-1}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi})S_{\boldsymbol{\theta}_1 n}(\boldsymbol{\phi},\mathbf{0})$$

is the residual in the least squares projection of $\partial^r L_n(\boldsymbol{\phi},\mathbf{0})/\partial\boldsymbol{\theta}_r^{\otimes r}$ onto the linear span of $S_{\boldsymbol{\theta}_1 n}(\boldsymbol{\phi},\mathbf{0})$.[6]In this context, we can proof the following result:

Theorem 50. *If Assumptions 47, 48 and 49 hold, then under* $H_0 : \boldsymbol{\theta} = \mathbf{0}$

$$LR_n = 2\left[L_n(\hat{\boldsymbol{\rho}}) - L_n(\tilde{\boldsymbol{\rho}})\right] = GET_n + O_p(n^{-\frac{1}{2r}}),$$

where

$$GET_n = \frac{1}{n}S'_{\boldsymbol{\theta}_1 n}(\tilde{\boldsymbol{\phi}},\mathbf{0})V^{-1}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\tilde{\boldsymbol{\phi}})S_{\boldsymbol{\theta}_1 n}(\tilde{\boldsymbol{\phi}},\mathbf{0}) + \frac{1}{n}\sup_{\boldsymbol{\theta}_r\neq 0}\begin{cases} Q_n(\boldsymbol{\theta}_r,\tilde{\boldsymbol{\phi}}) \text{ if } r \text{ is odd,} \\ Q_n(\boldsymbol{\theta}_r,\tilde{\phi})\mathbf{1}[\boldsymbol{\theta}_r^{\otimes r\prime} D_{rn}(\tilde{\boldsymbol{\phi}}) \geqslant 0] \text{ if } r \text{ is even.}\end{cases}$$

An important implication of Theorem 50 is that the rate of convergence of the difference between the LR and GET tests is inversely proportional to the order of identification, thereby generalizing the standard result for regular models.

Importantly, expression (4.1), which can be understood as a generalized Rayleigh quotient evaluated at the restricted $q_r^r \times 1$ vector $\boldsymbol{\theta}_r^{\otimes r}$, does not effectively depend on $\boldsymbol{\theta}_r$ when the nullity of the information matrix is 1, so Theorem 50 generalizes the results in L. F. Lee and Chesher (1986) and Rotnitzky et al. (2000) by allowing for the presence of multiple singularities under the null.

[6] Importantly, Assumption 49.2 guarantees that the denominator of $Q_n(\boldsymbol{\theta}_r,\boldsymbol{\phi})$ is positive because $V_{\boldsymbol{\theta}\boldsymbol{\theta}}$ is the covariance matrix of the residuals from the least squares projection of $s_{\boldsymbol{\theta}_1}(\boldsymbol{\phi},\mathbf{0})$ and $\frac{\partial^r l(\boldsymbol{\phi},\mathbf{0})}{\partial\otimes^r\boldsymbol{\theta}_r}$ on the linear span of $s_{\boldsymbol{\phi}}(\boldsymbol{\phi},\mathbf{0})$, while $V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r} - V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}V^{-1}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_r}$ is the residual covariance matrix of the projection of the second residual on the span of the first one, which by the Frisch-Waugh theorem coincides with the residual in the projection of $\frac{\partial^r l(\boldsymbol{\phi},\mathbf{0})}{\partial\otimes^r\boldsymbol{\theta}_r}$ onto the linear span of $s_{\boldsymbol{\phi}}(\boldsymbol{\phi},\mathbf{0})$ and $s_{\boldsymbol{\theta}_1}(\boldsymbol{\phi},\mathbf{0})$.

4.2.2 Repeated singularity of different orders

There are situations in which the degree of identification of the different elements of $\boldsymbol{\theta}$ under the null hypothesis is more heterogeneous than just either one or $r+1$. To characterise them in full, we need to generalize the conditions in Assumptions 48 and 49. Let $\varsigma_{\boldsymbol{\phi}i}(\boldsymbol{\phi})$ and $\varsigma_{\boldsymbol{\theta}i}(\boldsymbol{\phi})$ denote two measurable functions of dimensions $p \times 1$ and $m \times 1$, respectively, so that we can define the empirical process

$$\mathcal{S}_n(\boldsymbol{\phi}) = \begin{bmatrix} \mathcal{S}_{\boldsymbol{\phi},n}(\boldsymbol{\phi}) \\ \mathcal{S}_{\boldsymbol{\theta},n}(\boldsymbol{\phi}) \end{bmatrix} = \sum_{i=1}^{n} \varsigma_i(\boldsymbol{\phi}), \quad \text{where} \quad \varsigma_i(\boldsymbol{\phi}) = \begin{bmatrix} \varsigma_{\boldsymbol{\phi}i}(\boldsymbol{\phi}) \\ \varsigma_{\boldsymbol{\theta}i}(\boldsymbol{\phi}) \end{bmatrix}.$$

Typically, $\varsigma_{\boldsymbol{\phi}i}(\boldsymbol{\phi})$ coincides with the scores with respect to $\boldsymbol{\phi}$, and $\varsigma_{\boldsymbol{\theta}i}(\boldsymbol{\phi})$ with some higher-order derivatives with respect to the elements $\boldsymbol{\theta}$, so that $\mathcal{S}_n$ will serve as the analog to the sample score in regular models. In addition, let

$$\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta}) = \begin{bmatrix} (\boldsymbol{\phi} - \boldsymbol{\phi}^*) + \boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}) \\ \boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \end{bmatrix},$$

where $\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}) \in \mathbb{R}^p$ and $\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \in \mathbb{R}^m$ are non-random vector functions of the parameters that adequately capture their difference from the true values. Finally, let

$$\mathcal{I}(\boldsymbol{\phi}) = \begin{bmatrix} \mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\phi}) & \mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}}(\boldsymbol{\phi}) \\ \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}(\boldsymbol{\phi}) & \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}) \end{bmatrix}$$

denote a non-random positive semidefinite symmetric $(p+m) \times (p+m)$ matrix, which once again will effectively play the role of an information matrix.

Using this notation, we state the following assumptions, many of which are simplified versions of the conditions in Assumption 5 in Meitz and Saikkonen (2021):

Assumption 51. *(LQ approximation) L_n has a "linear-quadratic" expansion given by*

$$L_n(\boldsymbol{\phi}, \boldsymbol{\theta}) - L(\boldsymbol{\phi}^*, \mathbf{0}) = \mathcal{S}_n(\boldsymbol{\phi}^*)' \boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta}) - \frac{1}{2} n \boldsymbol{\lambda}'(\boldsymbol{\phi}, \boldsymbol{\theta}) \mathcal{I}(\boldsymbol{\phi}^*) \boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta}) + R_n(\boldsymbol{\phi}, \boldsymbol{\theta}),$$

where $R_n(\boldsymbol{\phi}, \boldsymbol{\theta})$ is a remainder term. In addition:

(51.1) $\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})$ is continuous in $\boldsymbol{\rho}$, and such that (i) $\boldsymbol{\lambda}(\boldsymbol{\phi}^*, \mathbf{0}) = \mathbf{0}$ and (ii) for all $\epsilon > 0$,

$$\inf_{\|(\boldsymbol{\phi}, \boldsymbol{\theta}) - (\boldsymbol{\phi}^*, \mathbf{0})\| \geqslant \epsilon} \|\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})\| \geqslant \delta_\epsilon \text{ for some } \delta_\epsilon > 0.$$

(51.2) $n^{-\frac{1}{2}}\mathcal{S}_n \xrightarrow{d} \mathcal{S}$ *for some zero-mean* $\mathbb{R}^{p+m}$*-valued Gaussian process with covariance kernel*

$$E\left[\mathcal{S}(\boldsymbol{\phi}_1)\mathcal{S}'(\boldsymbol{\phi}_2)\right] = E\left[\varsigma_i(\boldsymbol{\phi}_1)\varsigma_i'(\boldsymbol{\phi}_2)\right] = \mathcal{K}(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2).$$

(51.3) $\mathcal{I}(\boldsymbol{\phi}^*) = \mathcal{K}(\boldsymbol{\phi}^*, \boldsymbol{\phi}^*)$ *is Lipschitz continuous at a neighborhood of* $\boldsymbol{\phi}^*$ *and satisfies*

$$0 < e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)] < e_{\max}[\mathcal{I}(\boldsymbol{\phi}^*)] < \infty.$$

(51.4) The remainder term $R_n(\boldsymbol{\phi}, \boldsymbol{\theta})$ *satisfies*

$$\sup_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\boldsymbol{P}:\|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\|\leqslant\gamma_n} \frac{|R_n(\boldsymbol{\phi}, \boldsymbol{\theta})|}{1 + n\left\|\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})\right\|^2} = o_p(1)$$

for all sequences of (non-random) positive scalars $\{\gamma_n : n \geqslant 1\}$ *for which* $\gamma_n \to 0$ *as* $n \to \infty$.

(51.5) There exists some function $\mathbf{g}(\mathbf{y})$ *satisfying* $E\left[(\mathbf{g}(\mathbf{y}_i))^2\right] < \infty$ *such that*

$$||\mathcal{S}(\boldsymbol{\phi}^\dagger) - \mathcal{S}(\boldsymbol{\phi}^*)|| \leqslant ||\boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^*|| \sum_{i=1}^{n} \mathbf{g}(\mathbf{y}_i) \tag{4.2}$$

with probability 1 for all $(\boldsymbol{\phi}, \mathbf{0}) \in \mathcal{N}$.

(51.6) If $n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) = O(1)$*, then* $R_n(\boldsymbol{\phi}, \boldsymbol{\theta}) = O_p(n^{-a})$ *for some* a *such that* $\frac{1}{2} \geqslant a > 0$.

Assumption 51 states that the likelihood ratio can be expressed as the sum of a linear-quadratic approximation and a residual term, R_n. The linear-quadratic part, though, represents a higher-order expansion of the likelihood ratio around $\boldsymbol{\theta} = \mathbf{0}$. Assumption 51.1 captures the local identification condition at the true parameter value. Assumption 51.2 is analogous to the information matrix equality, while Assumption 51.3 to the standard non-singular information matrix assumption. In turn, Assumption 51.4 ensures that the residual is dominated by the leading terms, and thus, negligible asymptotically, while Assumption 51.5 enables us to substitute the true parameter $\boldsymbol{\phi}^*$ with the restricted estimator $\tilde{\boldsymbol{\phi}}$ after an appropriate adjustment for sampling variability. Finally, Assumption 51.6 allows

us to obtain the convergence rate of the linear-quadratic approximation, with a typically associated to the slowest rate of convergence of the parameter estimators under the null.

We can then prove the following result:

Theorem 52. *If Assumptions 47 and 51.1– 51.5 hold, then under* $H_0 : \boldsymbol{\theta} = \mathbf{0}$

$$LR = 2[L_n(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}) - L_n(\tilde{\boldsymbol{\phi}}, \mathbf{0})] = GET_n + o_p(1),$$

where

$$\begin{aligned} GET_n &= \sup_{\boldsymbol{\theta}} \{2[\mathcal{S}_{\boldsymbol{\theta},n}(\tilde{\boldsymbol{\phi}}) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}(\tilde{\boldsymbol{\phi}})\mathcal{I}^{-1}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\tilde{\boldsymbol{\phi}})\mathcal{S}_{\boldsymbol{\phi},n}(\tilde{\boldsymbol{\phi}})]'\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \\ &\quad -n\boldsymbol{\lambda}'_{\boldsymbol{\theta}}(\boldsymbol{\theta})[\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\tilde{\boldsymbol{\phi}}) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}(\tilde{\boldsymbol{\phi}})\mathcal{I}^{-1}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\tilde{\boldsymbol{\phi}})\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}}(\tilde{\boldsymbol{\phi}})]\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\}. \end{aligned}$$

If, in addition, Assumption 51.6 holds, then

$$LR = 2[L_n(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}) - L(\tilde{\boldsymbol{\phi}}, \mathbf{0})] = GET_n + O_p(n^{-a}).$$

As expected, we can easily show that our first theorem is a special case of this second theorem when the higher-order identification is of the same order for all the parameters involved regardless of the parity of r. More importantly, the proof of this theorem shows that we can interpret $L_n(\tilde{\boldsymbol{\phi}}, \mathbf{0}) + GET_n$ as a Taylor approximation of order $2r$ to the log-likelihood function around $\tilde{\boldsymbol{\rho}}$, which means that GET_n is effectively an LR-type test that compares the log-likelihood function under the null to the maximum of its lowest-order approximation under the alternative capable of identifying the restricted parameters.

Although GET cannot be directly understood as a moment test, a by-product of our most general theorem is a set of influence functions $S_n(\boldsymbol{\phi}, \mathbf{0})$ that can be used for that purpose after taking into account the sampling uncertainty in estimating $\boldsymbol{\phi}$ under the null. In fact, we can prove that this moment test, which converges in distribution to a χ^2_m under the null, where $m = dim[\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})]$, provides an upper bound to GET_n, albeit a rather loose one in many cases.

4.2.3 Distribution under local alternatives

Let us now consider the distribution of the test statistic under the following sequences of local alternatives:

$$H_{1n} : \sqrt{n} \left[\begin{array}{c} \boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_n) \\ \boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_n) \end{array} \right] = \left[\begin{array}{c} \boldsymbol{\lambda}_{\boldsymbol{\phi},\infty} \\ \boldsymbol{\lambda}_{\boldsymbol{\theta},\infty} \end{array} \right] = \boldsymbol{\lambda}_\infty \in \mathbb{R}^{\dim(\boldsymbol{\lambda}_{\boldsymbol{\theta}})}.$$

To do so, we need to assume that

Assumption 53. *(Cone cover) The sequence of sets*

$$\Lambda_n = \left\{ \sqrt{n} \boldsymbol{\lambda}_\theta(\boldsymbol{\theta}_n) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \right\}$$

covers a closed cone $\Lambda \subseteq \mathbb{R}^{dim(\boldsymbol{\lambda}_\theta)}$ *(with* $\Lambda_n \to \Lambda$*) so that there is a sequence of closed balls* B_{k_n} *of radius* $k_n \to \infty$ *centered at the origin such that* $\Lambda_n \cap B_{k_n} = \Lambda \cap B_{k_n}$*.*

Let $P_{\boldsymbol{\theta}_n}$ and P_0 denote the probability measures corresponding to H_{1n} and H_0, respectively. Then, we can prove the following result:

Theorem 54. *(Distribution under local alternatives)*

(54.1) $P_{\boldsymbol{\theta}_n}$ *is contiguous with respect to* P_0*.*

(54.2) Under H_{1n}*,*

$$\frac{1}{\sqrt{n}} \mathcal{S}_n(\boldsymbol{\phi}^*) \xrightarrow{d} N[\mathcal{I}(\boldsymbol{\phi}^*)\boldsymbol{\lambda}_\infty, \mathcal{I}(\boldsymbol{\phi}^*)].$$

(54.3) Under H_{1n} *and Assumption 53,*

$$GET_n \xrightarrow{d} \sup_{\boldsymbol{\lambda} \in \Lambda} \left\{ 2 \left[S + \left(\mathcal{I}_{\theta\theta}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta}\phi}(\boldsymbol{\phi}^*) \mathcal{I}_{\boldsymbol{\phi}\phi}^{-1}(\boldsymbol{\phi}^*) \mathcal{I}_{\phi\theta}(\boldsymbol{\phi}^*) \right) \boldsymbol{\lambda}_{\theta,\infty} \right]' \boldsymbol{\lambda} \right.$$
$$\left. - \boldsymbol{\lambda} \left[\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta}\phi}(\boldsymbol{\phi}^*) \mathcal{I}_{\boldsymbol{\phi}\phi}^{-1}(\boldsymbol{\phi}^*) \mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}}(\boldsymbol{\phi}^*) \right] \boldsymbol{\lambda} \right\}$$

where

$$S \sim \mathcal{N}[0, \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta}\phi}(\boldsymbol{\phi}^*) \mathcal{I}_{\boldsymbol{\phi}\phi}^{-1}(\boldsymbol{\phi}^*) \mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}}(\boldsymbol{\phi}^*)].$$

Therefore, the distribution of the empirical process underlying our tests converges to a Gaussian random element with a non-zero mean, and consequently, our test statistic to the supremum of a non-central χ^2-type process, despite the fact that our sequence of local alternatives written in terms of the model parameters converges at rates which are different from the usual ones. In fact, there may be different drifting sequences with the same limit, as we will see in section 4.3.2.3. In any event, we would like to emphasize that our proposed test is consistent against fixed alternatives because GET_n will diverge in those circumstances.

4.3 Examples

In this section, we discuss the application of our proposed tests to the first two examples of empirical interest that we mentioned in the introduction. Specifically, we derive a test for irrelevant sample selectivity in multivariate regression models, for which Theorem 50 suffices, and a test for normality against SNP alternatives, which requires our more general Theorem 52.

4.3.1 Example 1: Testing for selectivity in multivariate regressions

Arguably, the study of the determinants and consequences of non-random sample selection that followed Heckman (1974) seminal paper is one of the most important contributions of econometrics in the last fifty years. Nevertheless, the empirical analysis of a dataset would be much simpler if the sample from which it comes could be treated as if it were randomly generated even though it is not necessarily so. As is well known, this will happen when the unobserved determinants of the sample selection are independent of the unobserved determinants of the variables of interest conditional on the set of predetermined explanatory variables, or in simpler terms, when the selection is exogenous rather than endogenous. In the rest of this subsection, we shall develop a test of irrelevant sample selectivity in a multivariate regression context that highlights the hidden difficulties researchers often inadvertently encounter, but which can be easily overcome by the use of the GET procedures that we propose.

4.3.1.1 The model and its log-likelihood function

Consider the following multivariate version of the regression model with selectivity considered by L. F. Lee and Chesher (1986):

$$\mathbf{y} = \mathbf{y}^* d, \tag{4.3}$$

where d is a sample selection binary variable whose value is determined by an observed vector of exogenous regressors $\mathbf{w}$ and some unobserved determinant u_S according to the following equation written in terms of the usual indicator function

$$d = \mathbf{1}(\mathbf{w}'\boldsymbol{\varphi}^S + u_S \geqslant 0), \tag{4.4}$$

while the K partially observed variables $\mathbf{y}^* = (y_1^*, ..., y_K^*)'$ follow the multivariate regression

$$y_k^* = \boldsymbol{\varphi}_k^{M\prime}\mathbf{x} + \varphi_k^D u_k, \quad k = 1, \ldots, K, \tag{4.5}$$

$$\begin{pmatrix} \mathbf{u} \\ u_S \end{pmatrix} |\mathbf{x}, \mathbf{w} \sim N\left\{\mathbf{0}, \begin{bmatrix} \boldsymbol{R}\left(\boldsymbol{\varphi}^L\right) & \boldsymbol{\vartheta} \\ \boldsymbol{\vartheta}' & 1 \end{bmatrix}\right\}, \tag{4.6}$$

with $\mathbf{x}$ being a vector of exogenous regressors that may partially overlap with $\mathbf{w}$, so that $\boldsymbol{\varphi}^D = (\varphi_1^D, \ldots, \varphi_K^D)'$ contains the standard deviations of the regression shocks, $\boldsymbol{\varphi}^L$ the correlations between them, and $\boldsymbol{\vartheta}$ the correlations between those shocks and the unobserved component of the selection equation, whose variance we normalize to 1 without loss of generality.

Therefore, the contribution of a single observation to the sample log-likelihood function will be given (up to a constant term) by

$$(1-d)\ln\Phi\left(-\mathbf{w}'\boldsymbol{\varphi}^S\right) + d\ln\Phi\left[\frac{\mathbf{w}'\boldsymbol{\varphi}^S + \boldsymbol{\vartheta}'\mathbf{u}(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)}{\sqrt{1 - \boldsymbol{\vartheta}'\boldsymbol{R}^{-1}\left(\boldsymbol{\varphi}^L\right)\boldsymbol{\vartheta}}}\right] \tag{4.7}$$

$$-\frac{d}{2}\left[2\sum_{k=1}^K \ln\varphi_k^D + \ln\{\det[\boldsymbol{R}\left(\boldsymbol{\varphi}^L\right)]\} + \mathbf{u}'(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)\boldsymbol{R}^{-1}\left(\boldsymbol{\varphi}^L\right)\mathbf{u}(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)\right],$$

where $\boldsymbol{\varphi}^M = (\boldsymbol{\varphi}_1^{M\prime}, \ldots, \boldsymbol{\varphi}_K^{M\prime})'$, $\mathbf{u}(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D) = [u_1(\boldsymbol{\varphi}_1^M, \varphi_1^D), ..., u_K(\boldsymbol{\varphi}_K^D, \varphi_K^D)]'$, and

$$u_k(\varphi_k^M, \varphi_k^D) = \frac{y_k - \boldsymbol{\varphi}_k^{M\prime}\mathbf{x}}{\varphi_k^D}.$$

4.3.1.2 The null hypothesis of lack of selectivity and the GET test statistic

Under the null that the unobserved selectivity determinants are uncorrelated with the regression residuals, one can efficiently estimate the multivariate regression coefficients φ^M together with the covariance matrix parameters φ^D and φ^L without selection bias from the non-zero values of $\mathbf{y}$ only using equation by equation OLS without the need to consider the model for d. However, under the alternative, those OLS estimators will be biased because of the sample selectivity, which justifies testing the null hypothesis $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$.

For simplicity, consider the case in which $w = 1$ and the regression equations contain a constant term. Straightforward algebra shows that if we evaluate the scores at $\vartheta = \mathbf{0}$, then

$$s_{\vartheta_k} - M_1(\varphi^S)\varphi_k^D s_{\varphi_{k1}^M} = 0 \tag{4.8}$$

for $k = 1, \ldots, K$, where φ_{k1}^M contains the intercept in the conditional mean of y_k^*, and

$$M_1\left(\varphi^S\right) = \Phi^{-1}\left(\varphi^S\right)\phi\left(\varphi^S\right) \tag{4.9}$$

is the usual inverse Mills ratio. As L. F. Lee and Chesher (1986) explain in their univariate example, analogous singularities will arise for example when the observed selectivity determinants $\mathbf{w}$ are given by a set of dummy variables and $\mathbf{x}$ contains those dummy variables too. In general, singularities will be present whenever Heckman (1976) selectivity correction is perfectly collinear with the regressors that appear in the conditional means of the y^*'s even though the log-likelihood function in (4.7) is able to locally identify all the model parameters.

In addition to the K singularities in (4.8), there are $K(K+1)/2$ linear combinations of the scores and the elements of the Hessian corresponding to ϑ that are 0 too, which effectively means that we need to look at third-order derivatives. To do so, it is convenient to reparametrize from $\boldsymbol{\varphi}$ and $\boldsymbol{\vartheta}$ to $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ as we explain in the proof of Proposition 55 below, so that all the elements of the score and the Hessian matrix corresponding to $\boldsymbol{\theta}$ become identically 0 under the null. Fortunately, we can then show that the third-order derivatives with respect to $\boldsymbol{\theta}$, which are only zero on average under the null, will have a

full-rank asymptotic covariance matrix, so that we can apply Theorem 50 in this context. Somewhat remarkably, we can show the following result:

Proposition 55. *The difference between LR test of* $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ *in model (4.4)-(4.6) based on a random sample of* n *observations on* ($\mathbf{y}$,d) *and the following test statistic*

$$GET_n = \sup_{\boldsymbol{\upsilon} \neq \mathbf{0}} \frac{1}{\sum_{i=1}^{n} d_i} \left[\sum_{i=1}^{n} d_i H_3 \left(\frac{\boldsymbol{\upsilon}' \mathbf{v}_i(\tilde{\boldsymbol{\varphi}}^M, \tilde{\boldsymbol{\varphi}}^D)}{\sqrt{\boldsymbol{\upsilon}' \boldsymbol{\upsilon}}} \right) \right]^2 \tag{4.10}$$

is $O_p(n^{-1/6})$*, where* $H_3(z) = (z^3 - 3z)/\sqrt{6}$ *is the third-order normalized Hermite polynomial of a standardized variable* z*,* $\boldsymbol{\upsilon}$ *is a real vector of dimension* K *and* $\mathbf{v}(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)$ *denotes an affine transformation of the regression residuals* $\mathbf{u}(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)$ *whose mean vector and covariance matrix are* $\mathbf{0}$ *and* $\mathbf{I}_K$*, respectively, when evaluated at the restricted parameter estimators.*

In simpler terms, our test statistics numerically coincides with the supremum of the moment tests for univariate skewness based on the third Hermite polynomial over all possible linear combinations of the OLS residuals that have 0 mean and unit variance in the sample of observations with $d = 1$. In fact, the standardization is unnecessary because the moment test for univariate skewness is numerically invariant to affine transformations of the observations, which in turn confirms that the test statistic (4.10) is homogeneous of degree 0 in $\boldsymbol{\upsilon}$. Thus, when $K = 1$ our proposed test reduces to the test for selectivity derived by L. F. Lee and Chesher (1986) in the univariate case, which simply assesses the symmetry of the regression residuals by looking at the sample mean of their third powers.

The rationale is also analogous in the multivariate case. Equations (4.3)-(4.6) imply that the OLS residuals should be approximately multivariate normally distributed when the unobserved component of the sample selection is independent of the shocks to the observed variables. Under the alternative, in contrast, asymmetry becomes a common feature, as in the multivariate skew normal distribution we discussed in Amengual, Bei, and Sentana (2022). Intuitively, if we orthogonalize the regression residuals with respect to the

unobserved component of the selectivity equation, we end up with $\boldsymbol{\vartheta} u_S$ as a common component, whose distribution conditional on $d = 1$ is asymmetric even though the unconditional distribution of u_S is symmetric.

4.3.1.3 Local power analysis

Although the null distribution of the test statistic (4.10) is non-standard, we can still say something about the determinants of its local power. Consider the following sequence of local alternatives:

$$\lim_{n\to\infty} n^{1/6}\boldsymbol{\theta}_n = \boldsymbol{\theta}_\infty$$

where the rate of convergence is $1/6$ rather than $1/2$ because of the need for a third-order expansion of the log-likelihood function. Then, we can show that

Proposition 56. *The local power of the test in Proposition 55 only depends on the magnitude of the quadratic form*

$$\boldsymbol{\vartheta}'_\infty \boldsymbol{R}^{-1}\left(\boldsymbol{\varphi}^L\right)\boldsymbol{\vartheta}_\infty.$$

Intuitively, once we orthogonalize the multivariate regression residuals $\mathbf{u}$ by premultiplying by the inverse square root matrix $\boldsymbol{R}^{-1/2}\left(\boldsymbol{\varphi}^L\right)$, the "direction" of the vector $\boldsymbol{R}^{-1/2}\left(\boldsymbol{\varphi}^L\right)\boldsymbol{\vartheta}$ is irrelevant, what matters is its magnitude. As a result, in our simulations we can choose $\boldsymbol{R}\left(\boldsymbol{\varphi}^L\right) = \mathbf{I}_K$ and $\boldsymbol{\vartheta}_\infty$ proportional to the first vector of the canonical basis without loss of generality.

4.3.1.4 Simulation evidence

For simplicity, we let $w = x_1 = 1$ and $x_2 \sim N(0,1)$. Given that the MLE of the multivariate regression coefficients is equation by equation OLS, and that we are studying the case in which all regressions contain an intercept, the sample mean of the multivariate regression residuals $\hat{\mathbf{u}}$ will be a vector of K zeros. Similarly, any orthogonalization of the $\hat{\mathbf{u}}'s$ based on the estimated covariance matrix will have the identity matrix as sample covariance matrix because the MLEs of the residual standard deviations $\boldsymbol{\varphi}^D$ and correlations $\boldsymbol{\varphi}^L$ match perfectly the sample variances and covariances of $\hat{\mathbf{u}}$ with denominator $\sum_{i=1}^n d_i$. Therefore,

it is not surprising that the particular square root that orthonormalizes the OLS residuals in the sample is numerically irrelevant. For example, in the bivariate case, we could define v_1 as the standardized value of u_1 and v_2 as the standardized value of the residual in the OLS regression of u_2 on a constant and u_1. But we could also define them the other way round.

We can easily verify that the GET statistic is numerically invariant to the true values of $(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)$, so if we set $K = 2$, we can choose $\boldsymbol{\varphi}_k^M = (0, 1)$, $\boldsymbol{\varphi}^D = \iota_2$ without loss of generality. In turn, we set the selection parameter φ^S to 1 and the correlation coefficient φ^L to 0.25.

If we exploited our knowledge of the values of these two parameters, we could compute exact critical values under the null for any sample size to any degree of accuracy by repeatedly simulating samples from the true distribution. In practice, though, we fix the selection parameter and the correlation coefficient to their estimated values in each sample in what is effectively a parametric bootstrap procedure (see Appendix D.1 in Amengual and Sentana (2015) for details), so that we can automatically compute size-adjusted rejection rates, as forcefully argued by J. Horowitz and Savin (2000).

As alternative hypotheses, we consider $\boldsymbol{\vartheta}' = (0.57, 0.57)$ (H_{a1}) and $\boldsymbol{\vartheta}' = (0.80, 0)$ (H_{a2}). For each design, we generate 10,000 samples of size n and compute the parameter estimators and tests.

In Table 1 we compare the results of our tests with a bootstrap-based LR test. Panels A and B of Table 1 report the results for samples of length 400 and 1,600, respectively. We can verify that the LR test statistic is also numerically invariant to the true values $(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)$, which allows us to approximate its critical value using an analogous parametric bootstrap procedure. For comparison purposes, we also consider a J-test based on the influence functions underlying GET, which we label as GMM. The first three columns of Table 1 report rejection rates under the null at the 1%, 5% and 10% levels, confirming that our simulated critical values work remarkably well for both sample sizes.[7] In turn, the last

[7] Given the number of replications, the 95% asymptotic confidence intervals for the Monte Carlo rejection

six columns present the rejection rates at the 1%, 5% and 10% levels for the alternatives we consider. Our proposed test has similar power to the LR test for the two alternatives, and both these tests outperform the GMM one.

Finally, our results also indicate a Gaussian rank correlation[8]of 0.88 (0.95) between our proposed test statistic and the LR across Monte Carlo simulations of 400 (1,600) observations that satisfy the null, which is in line with the asymptotic equivalence result in Theorem 50. In addition, they indicate that the LR takes about 10 and 20 times as much CPU time to compute as GET does for $n = 400$ and $n = 1,600$, respectively, which makes a huge difference in the calculation of the bootstrap critical values.

4.3.2 Example 2: Testing for normality against SNP alternatives

Gram-Charlier expansions provide flexible and analytically tractable generalizations of the normal distribution. Unfortunately, their truncated versions lead to negative density values, and the parametric restrictions that Jondeau and Rockinger (2001) propose to guarantee positivity are not easy to implement even when the truncation order is low. In contrast, the SNP distributions introduced by Gallant and Nychka (1987) provide a Hermite expansion of the Gaussian density that is positive by construction. Although these authors introduced those distributions for nonparametric estimation purposes, León et al. (2009) treated them as parametric ones, studied their statistical properties, and used them in option valuation. Still, MLE under normality is much simpler than when the distribution of the shocks follows an SNP. For that reason, we shall derive a test of normality that will also highlight the hidden complications that researchers face in this context.

4.3.2.1 The model and its log-likelihood function

The model we consider is

$$y = \mu(\mathbf{x}, \boldsymbol{\alpha}) + \sigma(\mathbf{x}, \boldsymbol{\alpha})\, u \tag{4.11}$$

probabilities under the null are (.80,1.20), (4.57,5.43) and (9.41,10.59) at the 1, 5 and 10% levels.

[8] The Gaussian rank correlation between x_1 and x_2 is the Pearson correlation coefficient between $\Phi^{-1}(u_1)$ and $\Phi^{-1}(u_2)$, where u_1 and u_2 are the usual uniform ranks of the observations and $\Phi^{-1}(.)$ the quantile function of the standard normal (see Amengual, Sentana, and Tian (2022) for details).

where μ and σ are known functions of $\mathbf{x}$ and a finite-dimensional unknown parameter $\boldsymbol{\alpha}$, and u is independent of the predetermined variables in $\mathbf{x}$ with finite mean and variance φ^M and φ^V, respectively. We want to test u is normal against the alternative that it follows an SNP density. Observations are given by $(\mathbf{x}_i, y_i)$, $i = 1, 2, \ldots, n$, where $\mathbf{x}_i$ could include the lagged value of y_i to allow for time-series models such as AR and GARCH. For simplicity, we assume that u_i conditional on $\mathbf{x}_i$ is iid. As we will show in section 4.3.2.5 below, estimation of $\boldsymbol{\alpha}$ does not affect the properties of the test, so we initially assume this parameter vector is known and focus on the case without conditioning variables, in which $\mu(\boldsymbol{\alpha})$ and $\sigma(\boldsymbol{\alpha})$ are 0 and 1 without loss of generality. As a result, researchers only need to estimate φ^M and φ^V under the null.

The probability density function (pdf) of an SNP random variable of order K is given by

$$f(y;\varrho) = \frac{1}{\varphi_2}\phi\left(\frac{y-\varphi^M}{\sqrt{\varphi^V}}\right)\left[\epsilon + \frac{(1-\epsilon)\left\{P\left[\left(\frac{y-\varphi^M}{\sqrt{\varphi^V}}\right);\boldsymbol{\vartheta}\right]\right\}^2}{\int_{-\infty}^{\infty}\{P[u;\vartheta]\}^2\phi(u)du}\right], \tag{4.12}$$

with

$$P[u;\boldsymbol{\vartheta}] = 1 + \sum_{i=1}^{K}\vartheta_i H_i(u), \tag{4.13}$$

where $\phi(\cdot)$ denotes the standard normal pdf, $H_i(u)$ is the normalized Hermite polynomial of order i, which can be defined recursively for $i \geqslant 2$ as

$$H_i(u) = \frac{uH_{i-1}(u) - \sqrt{i-1}H_{i-2}(u)}{\sqrt{i}}, \tag{4.14}$$

with initial conditions $H_0(u) = 1$ and $H_1(u) = u$, $\int_{-\infty}^{\infty}\{P[u;\vartheta]\}^2\phi(u)du = 1 + \sum_{i=1}^{K}\vartheta_i^2$ is a constant which guarantees that the density integrates to 1, and ϵ is an infinitesimal factor used to bound the density below from 0, which Gallant and Nychka (1987) introduced to simplify their proofs. Henceforth, we will set $\epsilon = 0$ for the purposes of developing our testing procedure, but the same method applies with $\epsilon > 0$. Intuitively, a non-negative density is automatically achieved by multiplying the Gaussian density by the square of a

linear combination of Hermite polynomials. As explained by León et al. (2009), the SNP distributions can have non-negligible positive and negative asymmetry and excess kurtosis even with $K = 2$.

4.3.2.2 The null hypothesis of normality and the GET test statistic

To simplify the notation, we focus on the case of $K = 2$. Normality is trivially obtained when $H_0 : \vartheta_1 = \vartheta_2 = 0$. The complication arises because

$$s_{\vartheta_1} - 2\sqrt{\varphi^V} s_{\varphi^M} = 0,$$

$$s_{\vartheta_2} - 2\sqrt{2}\varphi^V s_{\varphi^V} = 0,$$

under H_0, so that the nullity of the information matrix is 2. Hall (1990) highlighted this problem when he considered tests of normality against semi-nonparametric alternatives in which the ϑ coefficients were in turn functions of some exogenous variable. However, his proposed solution was to ignore the parameters involved in the singularity, focusing instead only on those which could be regularly estimated under the null. Unfortunately, his recipe would leave us with no test in the case of the unconditional model (4.12)-(4.13).

In fact, it is easy to prove that ϑ_1 and ϑ_2 have different orders of identification, which means that we need to resort to our Theorem 52. In this context, we can establish the following result after reparametrizing from $(\boldsymbol{\varphi}', \boldsymbol{\vartheta}') = (\varphi^M, \varphi^V, \vartheta_1, \vartheta_2)$ to $(\boldsymbol{\phi}', \boldsymbol{\theta}') = (\phi^M, \phi^V, \theta_1, \theta_2)$ as explained in its proof:

Proposition 57. *The difference between the LR test of $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ in model (4.12)-(4.13) based on a random sample of n observations on $\mathbf{y}$ and the following test statistic*

$$GET_n = n \left\{ \left[\frac{1}{n} \sum_{i=1}^{n} H_3(\tilde{u}_i) \right]^2 + \left[\frac{1}{n} \sum_{i=1}^{n} H_4(\tilde{u}_i) \right]^2 \right\} \tag{4.15}$$

is $O_p(n^{-1/6})$ when the null is true, where $H_3(\tilde{u}_i)$ and $H_4(\tilde{u}_i)$ are the third- and fourth-order normalized Hermite polynomials of the $\tilde{u}_i$'s, which are the values of the y_i's standardized so that their sample mean and variance are 0 and 1, respectively.

Remarkably, this means that the Jarque and Bera (1980) test is asymptotically equivalent to the LR test of normality against SNP densities, although they converge to each other at a much lower rate than in the case of the Pearson family of alternative distributions they considered.

4.3.2.3 Local power analysis

Let $\chi_k^2(\upsilon)$ denote a non-central chi-square random variable with k degrees of freedom and non-centrality parameter υ. We can show that

Proposition 58. *Consider a sequence of parameters $\boldsymbol{\theta}_n$ satisfying*

$$\lim_{n\to\infty} \sqrt{n} \begin{pmatrix} -2\sqrt{3}\theta_{1,n}\theta_{2,n} \\ \sqrt{6}(\frac{1}{9}\theta_{1,n}^4 - \theta_{2,n}^2) \end{pmatrix} = \boldsymbol{\lambda}_{\boldsymbol{\theta},\infty} \in \mathbb{R}^2. \tag{4.16}$$

Under the sequence of DGPs indexed by $\boldsymbol{\theta}_n$,

$$GET_n \xrightarrow{d} \chi_2^2(\boldsymbol{\lambda}'_{\boldsymbol{\theta},\infty}\boldsymbol{\lambda}_{\boldsymbol{\theta},\infty}).$$

To understand this result, it is useful to note that

$$\begin{bmatrix} \sqrt{n}E\left\{H_3\left[\frac{y-E(y)}{\sqrt{Var(y)}}\right]\right\} \\ \sqrt{n}E\left\{H_4\left[\frac{y-E(y)}{\sqrt{Var(y)}}\right]\right\} \end{bmatrix} = \boldsymbol{\lambda}_{\boldsymbol{\theta},\infty} + o(1).$$

Unlike in the multivariate regression model with selectivity, though, we can have two different types of local alternatives compatible with (4.16):

$$H_{l1} : \theta_{1n} = n^{-\frac{1}{4}}h_1,\ \theta_{2n} = n^{-\frac{1}{4}}h_2,$$

$$H_{l2} : \theta_{1n} = n^{-\frac{1}{8}}h_1,\ \theta_{2n} = n^{-\frac{3}{8}}h_2.$$

Interestingly, $\sqrt{n}\theta_{2n}^2$ dominates $\sqrt{n}\theta_{1n}^4/9$ along H_{l1}, so that the SNP distributions under this sequence of local alternatives are platykurtic. In contrast, $\sqrt{n}\theta_{1n}^4/9$ dominates $\sqrt{n}\theta_{2n}^2$ along H_{l2}, so that the corresponding SNP distributions are leptokurtic.

4.3.2.4 Simulation evidence

Despite the fact that we estimate the sample mean and variance of each simulated sample, in this case there are effectively no nuisance parameters involved because both the GET and LR test statistics are numerically invariant to affine transformations of the observations. As a result, we can compute the exact finite sample distribution to any desired degree of accuracy for any sample size by simulating a large number of samples of the same size from a standard normal random variable. For that reason, we can focus directly on studying the power of the different tests.

As alternative hypotheses, we consider $\boldsymbol{\vartheta}' = (0.25, 0.10)$ (H_{a1}) and $\boldsymbol{\vartheta}' = (0.75, 0.05)$ (H_{a2}), setting $\varphi^M = 0$ and $\varphi^V = 1$ without loss of generality. As in the previous example, for each design we generate 10,000 samples of size n. In Table 2 we compare the results of our tests with the LR test. Panels A and B of Table 2 report the results for samples of length 400 and $1,600$, respectively. Given that the LR test statistic is also numerically invariant to the true values (φ^M, φ^V), we once again obtain its exact critical values using an analogous parametric bootstrap procedure. The first three columns of Table 2 report rejection rates under H_{a1} at the same levels, while the last three columns present the rejection rates at the 1%, 5% and 10% levels for H_{a2}. As can be seen, our proposed test has similar power to the LR test for both alternatives.

Finally, our results also indicate that the LR takes around 160 and 100 times as much CPU time to compute as GET does for $n = 400$ and $n = 1,600$, respectively, which considerably slows down the calculation of the bootstrap critical values.

4.3.2.5 Robustness to the estimation of mean and variance parameters

We now extend our previous results to a situation in which the conditional mean and variance of y are parametric functions of the variable in $\mathbf{x}$, as in (4.11). In this context, the objective becomes to test whether the innovation u follows a normal distribution versus an SNP.

The conditional log-likelihood of the i^{th} observation is given by:

$$l_i(\boldsymbol{\alpha}, \vartheta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma_Y^2(x_i, \boldsymbol{\alpha}) - \frac{1}{2}u_i^2(\boldsymbol{\alpha}) + 2\ln\left(1 + \sum_{i=1}^{K} \vartheta_i H_i\left[u_i(\boldsymbol{\alpha})\right]\right) - \ln\left(1 + \sum_{i=1}^{K} \vartheta_i^2\right).$$

To be able to obtain the required higher-order log-likelihood expansions, we assume that the following regularity conditions hold:

Assumption 59. *(Smoothness of the conditional first two moments) The conditional mean and variance functions $\mu_Y(\mathbf{x}_i, \boldsymbol{\alpha})$ and $\sigma_Y(\mathbf{x}_i, \boldsymbol{\alpha})$ that appear in (4.11) are such that*

(59.1) They are eight times continuously differentiable with respect to $\boldsymbol{\alpha}$.

(59.2) For all $k \in N^{d_{\boldsymbol{\alpha}}}$ and $\iota' k = 1, \ldots, 8$, it holds that

$$E\left[\left(\frac{\partial^{\boldsymbol{\iota}'\boldsymbol{k}}\mu_Y(\mathbf{x}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^{\boldsymbol{k}}}\right)^2\right] < \infty, \; E\left[\left(\frac{\partial^{\boldsymbol{\iota}'\boldsymbol{k}}\sigma_Y^2(\mathbf{x}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^{\boldsymbol{k}}}\right)^2\right] < \infty,$$

where $k = (k_1, \ldots, k_{d_{\boldsymbol{\alpha}}})$,

$$\begin{aligned}
\frac{\partial^{\boldsymbol{\iota}'\boldsymbol{k}}\mu_Y(\mathbf{x}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^{\boldsymbol{k}}} &= \frac{\partial^{\boldsymbol{\iota}'\boldsymbol{k}}\mu_Y(\mathbf{x}, \boldsymbol{\alpha})}{\partial \alpha_1^{k_1} \ldots \partial \alpha_{d_\alpha}^{k_{d_{\boldsymbol{\alpha}}}}}, \quad \textit{and} \\
\frac{\partial^{\boldsymbol{\iota}'\boldsymbol{k}}\sigma_Y^2(\mathbf{x}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^{\boldsymbol{k}}} &= \frac{\partial^{\boldsymbol{\iota}'\boldsymbol{k}}\sigma_Y^2(\mathbf{x}, \boldsymbol{\alpha})}{\partial \alpha_1^{k_1} \ldots \partial \alpha_{d_\alpha}^{k_{d_{\boldsymbol{\alpha}}}}}.
\end{aligned}$$

Then, we can prove the following result, which is entirely analogous to Proposition 8 in Amengual, Bei, Carrasco, and Sentana (2022):

Proposition 60. *Under Assumptions (59.1) and (59.2), replacing the true value of $\boldsymbol{\alpha}$ by $\tilde{\boldsymbol{\alpha}}$, its restricted maximum likelihood estimator under H_0, does not alter the expressions of the GET test in Proposition 57 or its asymptotic distribution under the null or sequences of local alternatives.*

4.4 Conclusions

We propose a generalization of the extremum-type tests in L. F. Lee and Chesher (1986) to models in which the nullity of the information matrix under the null hypothesis is larger

than one. In the case of a single singularity, our results are consistent with theirs, as well as with those in Rotnitzky et al. (2000). However, when the information matrix is repeatedly singular, we provide a computationally convenient alternative to the LR test, which is particularly useful for resampling-based calculations of p-values. Specifically, our proposed test statistic is a sup-type test over a space whose dimension is at most the nullity of the information matrix, and often less, while the maximization of the original log-likelihood function is over a space of the same dimension as the vector of parameters, which is usually much larger. In addition, the fact that several log-likelihood derivatives of various orders are 0 under the null implies that the LR requires the estimation of all the parameters that appear under the alternative in a model whose log-likelihood function is extremely flat around its maximum. Intuitively, the substantial computational gains that we find arise because GET is a LR-type test that compares the log-likelihood function under the null to the maximum of its lowest-order approximation under the alternative capable of identifying the restricted parameters.

Despite having many features in common, our results cannot be directly applied to testing normality against finite Gaussian mixtures. Nevertheless, we used them as a powerful lever to derive such tests in Amengual, Bei, Carrasco, and Sentana (2022).

5. Conclusions

This dissertation consists of three essays on inference in non-standard scenarios.

In the first essay, I propose inference procedures for a target object whose identified set is a union of bounds. 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 (2020a) and D. W. K. Andrews and Kwon (2023).

The second essay proposes a bootstrap test procedure for the null hypothesis $H_0 : c'\theta = r$ in moment inequality models, which I refer to as the local linearization based test. The test is profile-based, but to calculate the critical value, it only requires a local search around some preliminary estimator. An appealing feature is that the bootstrap iterates over quadratic programming problems, which significantly reduces the computing time. The new test controls asymptotic size uniformly over a large class of data generating processes and can be applied to construct confidence sets for subvectors by test inversion. Simulation experiments demonstrate good statistical and computational properties.

In the third essay, We propose a generalization of the extremum-type tests in L. F. Lee and Chesher (1986) to models in which the nullity of the information matrix under the null hypothesis is larger than one. Interestingly, the asymptotic distribution of our test statistic is similar to the asymptotic distribution of the usual overidentification test statistic in a GMM model in which the expected Jacobian of the moment conditions is of

reduced rank but the parameters are second-order identified (see Supplemental Appendix E of Amengual et al. (2020) for a formal link to the results in Dovonon and Renault (2013)). An application of our approach to GMM contexts in which not only the expected Jacobian matrix is singular but some higher order Jacobian matrices are singular too would constitute a very interesting extension.

Appendix A. Additional Results for Chapter 2

A.1 Proofs for Section 2.3

A.1.1 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:

$$\hat{\Lambda} = \left\{ (A_\ell \delta,\ A_u \delta) \in \Lambda : \delta \in \hat{\Delta} \right\}, \tag{A.1}$$

$$\hat{\Delta} = \left\{ \delta : \frac{|\hat{\delta}_j - \delta_j|}{\hat{\omega}_j} \leqslant \hat{c}_\eta, \forall j = 1, ..., J \right\} \tag{A.2}$$

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

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

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 $p \geqslant \alpha$, $\underline{c} = c$.

   (c) If $\bar{c} - \underline{c} > \varepsilon$, go to 3b;

   If $\bar{c} - \underline{c} \leqslant \varepsilon$, $\hat{c}^t = (\underline{c} + \bar{c})/2$.

4. Construct the confidence interval by a grid search over an outer set of CI^{h} given in (2.10).

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

- If $\theta \leqslant \max_{b\in\mathcal{B}_u} \hat{\lambda}_{u,b} + \Phi^{-1}(1-\eta)\hat{\sigma}_{u,b}$ 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 $CI^{\mathrm{m}} = \varnothing$, exit the algorithm.

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

(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 $\hat{T}(\theta) \leqslant \hat{c}^{\mathrm{m}}(\theta;\alpha)$, $\theta_2 = \theta$.

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 $\varnothing$. 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 (2020a) and I. Andrews et al. (2019), but this is outside the scope of this paper and I leave it for future research.

A.1.2 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 62 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

$$T_k = \max\left\{\min_{b \in \mathcal{B}} Z_{\ell,b} + \lambda_{k\ell,b},\ \min_{b \in \mathcal{B}} Z_{u,b} + \lambda_{ku,b}\right\} \tag{A.3}$$

be the asymptotic analog of $\hat{T}(\theta_{P_n,k})$, where $(\lambda_{k\ell}, \lambda_{ku})$ are specified in Lemma 62. 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 $\mathcal{B}_{k\ell}$, $\mathcal{B}_{ku}$:

$$b_{k\ell} = \min\left\{\arg\min_{b \in \mathcal{B}_\ell} Z_{\ell,b} + \lambda_{k\ell,b}\right\},$$

$$b_{ku} = \min\left\{\arg\min_{b \in \mathcal{B}_u} Z_{u,b} + \lambda_{ku,b}\right\},$$

$$\mathcal{B}_{k\ell} = \{b \in \mathcal{B}_\ell : \lambda_{k\ell,b} < \infty\},$$

$$\mathcal{B}_{ku} = \{b \in \mathcal{B}_u : \lambda_{ku,b} < \infty\}.$$

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

$$t_{k\ell,1}(b) = \begin{cases} \min_{\tilde{b} \in \mathcal{B}} \left(1 + \rho_{\ell u}(b, \tilde{b})\right)^{-1} \left(Z_{u,\tilde{b}} + \rho_{\ell u}(b, \tilde{b}) Z_{\ell,b} + t^\dagger_{k\ell,1}(b, \tilde{b})\right), & \text{if } \min_{\tilde{b} \in \mathcal{B}} \rho_{\ell u}(b, \tilde{b}) > -1 \\ -\infty & \text{elsewhere} \end{cases} \tag{A.4}$$

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

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

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

where

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

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

$$t^{\dagger}_{k\ell,2}(b,\tilde{b}) = \lambda_{k\ell,\tilde{b}} - \rho_{\ell}(b,\tilde{b})\lambda_{k\ell,b}$$

$$t^{\dagger}_{ku,2}(b,\tilde{b}) = \lambda_{ku,\tilde{b}} - \rho_{u}(\tilde{b},b)\lambda_{ku,b}.$$

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

$$c^{\mathrm{c}}_{k} = \begin{cases} \Phi^{-1}\left(\alpha^{\mathrm{c}}\Phi\left(t_{k\ell,1}(b_{k\ell})\right) + (1-\alpha^{\mathrm{c}})\Phi\left(t_{k\ell,2}(b_{k\ell})\right)\right) & \text{if } Z_{\ell,b_{k\ell}} + \lambda_{k\ell,b_{k\ell}} \geqslant Z_{u,b_{ku}} + \lambda_{ku,b_{ku}} \\ \Phi^{-1}\left(\alpha^{\mathrm{c}}\Phi\left(t_{ku,1}(b_{ku})\right) + (1-\alpha^{\mathrm{c}})\Phi\left(t_{ku,2}(b_{ku})\right)\right) & \text{if } Z_{\ell,b_{k\ell}} + \lambda_{k\ell,b_{k\ell}} < Z_{u,b_{ku}} + \lambda_{ku,b_{ku}} \end{cases} \tag{A.5}$$

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

$$\begin{aligned} p(c) = \max\{&P\left(T_{\ell} > c^{\mathrm{m}}_{\ell}(c) \text{ or } \{T_m > c^{\mathrm{m}}_{m}(c) \text{ and } T_u > c^{\mathrm{m}}_{u}(c)\}\right), \\ &P\left(T_u > c^{\mathrm{m}}_{u}(c) \text{ or } \{T_m > c^{\mathrm{m}}_{m}(c) \text{ and } T_{\ell} > c^{\mathrm{m}}_{\ell}(c)\}\right)\}, \end{aligned} \tag{A.6}$$

where

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

and T_k, c^{c}_{k} are defined in (A.75), (A.77). Lastly, let

$$c^{t} = \inf_{c}\{c \geqslant 0 : p(c) \leqslant \alpha - \eta\}, \tag{A.7}$$

be the asymptotic analog of $\hat{c}^{\mathrm{t}}$ defined (2.31).

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

A.1.3 Proofs for Theorems and Propositions

Proof of Lemma 8.

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

$$\frac{\Phi\left(\hat{T}(\theta)\right) - \Phi\left(t_{\ell,1}(\theta, b_1)\right)}{\Phi\left(t_{\ell,2}(\theta, b_1)\right) - \Phi\left(t_{\ell,1}(\theta, b_1)\right)} \Big| \left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\right\} \overset{\text{FOSD}}{\leq} \text{Unif}(0,1). \tag{A.8}$$

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

$$\left\{\hat{T}(\theta) = \mathcal{Z}_{\ell,b_1}\right\} = \bigcup_{s\in\mathcal{B}} \left\{A_s \mathcal{Z}_{\ell,b_1} \leqslant b_s\right\}. \tag{A.9}$$

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

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

and

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

Therefore,

$$\{A_s \mathcal{Z}_{\ell,b_1} \leqslant b_s\} = \left\{\mathcal{V}_s^- \leqslant \mathcal{Z}_{\ell,b_1} \leqslant \mathcal{V}^+, \mathcal{V}^0 \geqslant 0\right\} \tag{A.10}$$

where

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

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

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

Note that

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

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 9.

Let

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

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 J. D. Lee et al. (2016),

$$F_\mu\left(\mathcal{Z}_{\ell,b_\ell}; t_{\ell,1}(\theta,b_1), t_{\ell,2}(\theta,b_1)\right) \Big| \bigcup_{s\in\mathcal{B}} \{A_s\mathcal{Z}_{\ell,b_1} \leqslant b_s\} \sim \text{Unif}(0,1), \tag{A.11}$$

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

$$F_0\left(z; t_{\ell,1}(\theta, b_1), t_{\ell,2}(\theta, b_1)\right) \leqslant F_\mu\left(z; t_{\ell,1}(\theta, b_1), t_{\ell,2}(\theta, b_1)\right). \tag{A.12}$$

Therefore, we have

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

□

Proof of Proposition 9.

Proof. To simplify notation, let

$$\begin{aligned}
\mathcal{B}_{\ell 0} &= \left\{b \in \mathcal{B}_\ell : \lambda_{\ell,b} \leqslant \theta, P\left(b = \hat{b}_\ell\right) > 0\right\}, \\
\mathcal{B}_{u0} &= \left\{b \in \mathcal{B}_u : \lambda_{u,b} \geqslant \theta, P\left(b = \hat{b}_u\right) > 0\right\}.
\end{aligned}$$

Let $b_1 \in \mathcal{B}_{\ell 0}$, by Lemma 8, it holds that

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

where the second line follows from $F_0(x; t_1, t_2)$ strictly increasing in x, the inequality follows from (A.12) 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 (A.11).

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

$$P\left(\hat{T}(\theta) > \hat{c}^{\mathrm{c}}(\theta,\alpha)\middle|\hat{T}(\theta) = \mathcal{Z}_{u,b_2}\right) \leqslant \alpha. \tag{A.13}$$

Therefore, we have

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

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 (2.23), and the inequality follows from (A.8) and (A.13). □

Proof of Theorem 18.

Proof. By Lemma 62, 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 $c \in \mathbb{R}_+$,

$$\bar{p}\left(c, \lambda_{P_n}, \hat{\Sigma}_n/n\right) \xrightarrow{p} p(c) \tag{A.14}$$

where $\bar{p}\left(c, \lambda_{P_n}, \hat{\Sigma}_n/n\right)$ is defined in (2.30) and $p(c)$ is defined in (A.78). Note that by Assumption 16 and (A.36), $\hat{\Omega}_n \xrightarrow{p} \Omega_0$, thus there is $\tau_n = o(1)$ such that

$$\hat{\Omega}_n = \Omega_0 + o_p(\tau_n). \tag{A.15}$$

Let

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

To show (A.14), note that for all $\varepsilon > 0$,

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

where the first inequality follows from $P(A) \leqslant P(A \cap B) + P(B^c)$, the second inequality follows from (A.15), 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

$$\limsup_n \bar{p}\left(c, \lambda_{P_n}, \Sigma_n/n\right) = p(c). \tag{A.16}$$

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

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

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) $T_\ell \perp c_\ell^{\mathrm{c}}$, $T_m \perp c_m^{\mathrm{c}}$, $T_u \perp c_u^{\mathrm{c}}$ by construction. Thus (A.16) follows from Lemma 63.

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

$$\tilde{p}(c, \lambda_{P_n}, \Sigma_n/n) = \max\left\{P\left(\hat{T}(\lambda_{P_n,\ell,b_\ell}) > \hat{c}^{\mathrm{m}}(\lambda_{P_n,\ell,b_\ell}, c; \alpha^{\mathrm{c}}); \mathcal{N}(\lambda_{P_n}, \Sigma_n)\right),\right. \tag{A.17}$$
$$\left. P\left(\hat{T}(\lambda_{P_n,u,b_u}) > \hat{c}^{\mathrm{m}}(\lambda_{P_n,u,b_u}, c; \alpha^{\mathrm{c}}); \mathcal{N}(\lambda_{P_n}, \Sigma_n)\right)\right\}$$

and we have

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

where the last equality follows from Lemma 63. And

$$\begin{aligned}\tilde{p}(c, \lambda_{P_n}, \Sigma_n/n) =& \max\left\{P\left(\hat{T}(\theta_{n,\ell}) > \tilde{c}^{\mathrm{m}}(\theta_{n,\ell}, c; \alpha^{\mathrm{c}}); \mathcal{N}(\lambda_{P_n}, \Sigma_n)\right),\right.\\ &\left. P\left(\hat{T}(\theta_{n,u}) > \tilde{c}^{\mathrm{m}}(\theta_{n,u}, c; \alpha^{\mathrm{c}}); \mathcal{N}(\lambda_{P_n}, \Sigma_n)\right)\right\}\\ \to& \max\left\{P\left(T_\ell > c_\ell^{\mathrm{m}}(c)\right),\ P\left(T_u > c_u^{\mathrm{m}}(c)\right)\right\} \qquad \text{(A.18)}\\ =& p(c)\end{aligned}$$

(A.18) follows from continuous mapping theorem.

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

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

where c^{t} is defined in (A.79) and $\hat{c}^t_{P_n}$ is defined in (A.38). Note that by definition

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

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

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

where the last equation is by

$$\bar{p}\left(c^t - \varepsilon, \lambda_{P_n}, \hat{\Sigma}_n/n\right) \xrightarrow{p} p\left(c^t - \varepsilon\right) < \alpha - \eta, \tag{A.19}$$

and (A.19) follows from Step 1 (A.14).

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

$$\begin{aligned}
&\limsup_{n\to\infty} \max\left\{P_n\left(\hat{T}(\theta_\ell) > \hat{c}^{\mathrm{m}}(\theta_\ell, \hat{c}^t) \vee \left\{\hat{T}(\theta_m) > \hat{c}^{\mathrm{m}}(\theta_m, \hat{c}^t) \wedge \hat{T}(\theta_u) > \hat{c}^{\mathrm{m}}(\theta_u, \hat{c}^t)\right\}\right),\right. \\
&\qquad \left. P_n\left(\hat{T}(\theta_u) > \hat{c}^{\mathrm{m}}(\theta_u, \hat{c}^t) \vee \left\{\hat{T}(\theta_m) > \hat{c}^{\mathrm{m}}(\theta_m, \hat{c}^t) \wedge \hat{T}(\theta_\ell) > \hat{c}^{\mathrm{m}}(\theta_\ell, \hat{c}^t)\right\}\right)\right\} \\
=&\limsup_{n\to\infty} \max\left\{P_n\left(\hat{T}(\theta_\ell) > \tilde{c}^{\mathrm{m}}(\theta_\ell, \hat{c}^t) \vee \left\{\hat{T}(\theta_m) > \tilde{c}^{\mathrm{m}}(\theta_m, \hat{c}^t) \wedge \hat{T}(\theta_u) > \tilde{c}^{\mathrm{m}}(\theta_u, \hat{c}^t)\right\}\right),\right. \\
&\qquad \left. 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)\right\} \\
\leqslant&\limsup_{n\to\infty} \max \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (A.20) \\
&\left\{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),\right. \\
&\qquad \left. 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)\right\} \\
&+ \limsup_{n\to\infty} P_n\left(\hat{c}^t_{P_n} \leqslant c^t - \varepsilon\right) \\
=&p(c^t - \varepsilon). \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (A.21)
\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 (A.21) holds at all $\varepsilon > 0$, we can take a sequence of $\varepsilon \to 0$, then by Lemma 68,

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

By construction,

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

Thus by Lemma 62,

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

□

Proof of Theorem 21.

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

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

$$\mathcal{Z}_{\ell,b} = -\mathcal{Z}_b.$$

And the test statistic is

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

Step 1. By Lemma 67, there is $\alpha_1' > \alpha$ such that

$$\liminf_n \inf_{P\in\mathcal{P}} P\left(\hat{c}^{\mathrm{t}} \leqslant \Phi^{-1}(1-\frac{\alpha_1'}{2})\right) = 1. \tag{A.22}$$

By Lemma 70, 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 (2.37) follows from

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

$=1$

Step 2. I show (2.38) with $\theta_n = \theta_\ell - \frac{\kappa}{\sqrt{n}}$. Note that by (2.37), there is $\alpha' > \alpha$ such that

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

Under P_n, we can show that there is a subsequence P_{a_n} such that (A.36), (A.37) 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 $\bar{\mathcal{B}} \neq \varnothing$.

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

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

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

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

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

$$\geqslant P\left(c_2 \geqslant Z_{b^*} + \bar{\lambda}_{b^*} + \frac{\kappa}{\sigma_{b^*}} \geqslant c_1, \mathbb{E}_b \geqslant A, b \in \bar{\mathcal{B}}\backslash\{b^*\}\right)$$

$$= P\left(c_2 \geqslant Z_{b^*} + \bar{\lambda}_{b^*} + \frac{\kappa}{\sigma_{b^*}} \geqslant c_1\right)$$

$$P\left(\mathbb{E}_b \geqslant c_1 - \bar{\lambda}_b - |\rho_{b^*b}|\left(c_1 - \bar{\lambda}_{b^*}\right) - \left(\frac{\sigma_{b^*}}{\sigma_b} - |\rho_{b^*b}|\right)\frac{\kappa}{\sigma_{b^*}}, b \in \bar{\mathcal{B}}\backslash\{b^*\}\right)$$

where

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

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

$$A = c_1 - \bar{\lambda}_b - |\rho_{b^*b}|\left(c_1 - \bar{\lambda}_{b^*}\right) - \left(\frac{\sigma_{b^*}}{\sigma_b} - |\rho_{b^*b}|\right)\frac{\kappa}{\sigma_{b^*}}$$

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

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

and therefore

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

Part II. Large Bounds. By Lemma 71, there is α_1' such that

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

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 $\kappa_n' \to \infty$ and $\kappa_n' \ll \kappa_n$. Lemma 61 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 (2.36),

$$\begin{aligned}
\mathcal{Z}_{\ell,\hat{b}_\ell} \leqslant \mathcal{Z}_{\ell,b_\ell} &= \frac{\hat{\lambda}_{\ell,b_\ell} - \theta}{\hat{\sigma}_{\ell,b_\ell}/\sqrt{n}} \\
&\leqslant \frac{\hat{\lambda}_{\ell,b_\ell} - \hat{\lambda}_{u,b_u} - \frac{\hat{\sigma}_{u,b_u}}{\sqrt{n}}c_1}{\hat{\sigma}_{\ell,b_\ell}/\sqrt{n}} \\
&= \frac{\hat{\lambda}_{\ell,b_\ell} - \hat{\lambda}_{u,b_u}}{\hat{\sigma}_{\ell,b_\ell}/\sqrt{n}} - \frac{\hat{\sigma}_{u,b_u}}{\hat{\sigma}_{\ell,b_\ell}}c_1 \to -\infty
\end{aligned} \tag{A.23}$$

and

$$\mathcal{Z}_{u,\hat{b}_u} = \frac{\theta - \hat{\lambda}_{u,\hat{b}_u}}{\hat{\sigma}_{u,\hat{b}_u}} \in \left(c_1,\ \kappa_n'\right]. \tag{A.24}$$

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, (A.23) and (A.24) implies that

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

Therefore,

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

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

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

and θ is rejected.

The proof for (2.38) is similar to Part I Step 2. □

Proof of Theorem 22.

Proof. By Lemma 61, it is easy to see that

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n} P\left(\theta_{\ell,n} \notin CI^{\mathrm{m}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) = 1 \tag{A.25}$$

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

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

$$\begin{aligned}
& P\left(\theta_{\ell,n} \notin CI^{\mathrm{YKHS}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) \\
=& P\left(\theta_{\ell,n} < \hat{\lambda}_{m,\min} - \sqrt{\frac{n}{m}} Q^*\left(\hat{\lambda}^*_{n,\min} - \hat{\lambda}_{n,\min}, \hat{p}\right)\right) \\
& + P\left(\theta_{\ell,n} > \hat{\theta}_{m,\max} - \sqrt{\frac{n}{m}} Q^*\left(\hat{\lambda}^*_{n,\max} - \hat{\lambda}_{n,\max}, 1-\hat{p}\right)\right) \\
=& P\left(Q^*\left(\sqrt{n}\left(\hat{\lambda}^*_{n,\min} - \hat{\lambda}_{n,\min}\right), \hat{p}\right) < \sqrt{m}(\hat{\lambda}_{m,\min} - \theta_{\ell,n})\right) \qquad \text{(A.26)} \\
& + P\left(Q^*\left(\sqrt{n}(\hat{\lambda}^*_{n,\max} - \hat{\lambda}_{n,\max}), 1-\hat{p}\right) > \sqrt{m}(\hat{\lambda}_{m,\max} - \theta_{\ell,n})\right). \qquad \text{(A.27)}
\end{aligned}$$

As for (A.26), 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

$$(A.26) = P\left(Q^*\left(\sqrt{n}\left(\hat{\lambda}^*_{n,\min} - \hat{\lambda}_{n,\min}\right), \hat{p}\right) < \sqrt{m}(\hat{\lambda}_{m,\min} - \lambda_{\ell,b_\ell})\right) + o(1).$$

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

$$(A.27) = P\left(Q^*\left(\sqrt{n}(\hat{\lambda}^*_{n,\max} - \hat{\lambda}_{n,\max}), 1 - \hat{p}\right) > \sqrt{m}(\hat{\lambda}_{m,\max} - \lambda_{\ell,b_\ell})\right) + o(1) \tag{A.28}$$

with similar argument. And if $\sqrt{m}(\lambda_{u,b_u} - \theta_{\ell,n}) \to \infty$, (A.28) 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

$$\limsup_n \sup_P P\left(\theta_{\ell,n} \notin CI^{\text{YKHS}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) \leqslant \alpha. \tag{A.29}$$

(A.25) and (A.29) gives (2.39). □

A.1.4 Auxiliary Lemmas

Lemma 61. *($\sqrt{n}$ Convergence Rate) Suppose Assumptions 13, 14, 15, 16, and 17 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 62, 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

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

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

$$P_{a_n}\left(\hat{T}(\theta) > \hat{c}^{\mathrm{m}}(\theta;\alpha) \text{ for all } \theta < \theta_\ell - \frac{\bar{\kappa}}{\sqrt{a_n}}\right) \geqslant 1 - \frac{\varepsilon}{2}, \tag{A.30}$$

$$P_{a_n}\left(\hat{T}(\theta) > \hat{c}^{\mathrm{m}}(\theta;\alpha) \text{ for all } \theta < \theta_\ell - \frac{\bar{\kappa}}{\sqrt{a_n}}\right) \geqslant 1 - \frac{\varepsilon}{2}. \tag{A.31}$$

I will show (A.30), and the proof of (A.31) 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

$$\liminf_n P_n\left(A_n(\bar{\kappa}_1)\right) \geqslant 1 - \frac{\varepsilon}{6}, \tag{A.32}$$

where

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

To see this,

$$
\begin{aligned}
&P_n\left(A_n(\bar{\kappa}_1)\right)\\
=&P_n\left(\mathcal{Z}_{\ell,\hat{b}_\ell} \geqslant \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 \leqslant \theta_\ell - \frac{\bar{\kappa}_1}{\sqrt{n}}\right)\\
\geqslant&P_n\left(\mathcal{Z}_{\ell,\hat{b}_\ell} \geqslant \mathcal{Z}_{u,b_u}, \mathcal{Z}_{\ell,\hat{b}_\ell} > \Phi^{-1}(1-\frac{\alpha-\eta}{2}),\ \text{for all } \theta \leqslant \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}} \geqslant \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.\\
&\qquad \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 \leqslant \theta_\ell - \frac{\bar{\kappa}_1}{\sqrt{n}}\right)\\
\geqslant&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}} \geqslant -\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 \geqslant A, \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}
$$

where

$$
A = \left(\frac{1}{\hat{\sigma}_{\ell,\hat{b}_\ell}} + \frac{1}{\hat{\sigma}_{u,b_u}}\right)^{-1}\left(\frac{\theta_u-\hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}} - \frac{\hat{\lambda}_{\ell,\hat{b}_\ell}-\lambda_{\ell,\hat{b}_\ell}}{\hat{\sigma}_{\ell,\hat{b}_\ell}/\sqrt{n}}\right)
$$

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

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

$$
\hat{\sigma}_{\ell,\hat{b}_\ell}\Phi^{-1}(1-\frac{\alpha-\eta}{2}) - \sqrt{n}\left(\hat{\lambda}_{\ell,\hat{b}_\ell}-\lambda_{\ell,\hat{b}_\ell}\right) = O_P(1).
$$

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 13, 14, 15, and 16. Then, for all $\varepsilon > 0$, there is $\bar{M} \in \mathbb{R}$ such that

$$
\liminf_n P_n\left(B_n\right) \geqslant 1 - \frac{\varepsilon}{6}, \tag{A.33}
$$

where

$$
B_n = B_{1n} \cup B_{2n},
$$

$$B_{1n} = \left\{ \min_{b\in\mathcal{B}} \rho_{\ell u}(\hat{b}_\ell, b) = -1 \right\},$$

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

because

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

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 (B.8), there is $\bar{\kappa}_2$ such that

$$\liminf_n P_n\left(C_n(\bar{\kappa}_2)\right) \geqslant 1 - \frac{\varepsilon}{6}, \tag{A.34}$$

where

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

where $\bar{z}$ is defined in Lemma 64 with $\bar{M}$ given above (A.33).

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

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

we have

$$\liminf_n P_n\left(D_n\right)$$

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

where the equality follows from Lemma 64: the three assumptions in Lemma 64 hold because (i) $\bar{\kappa} \geqslant 0$, (ii) $A_n(\bar{\kappa})$, (iii) $B_n \cap C_n(\bar{\kappa})$. The last inequality follows from (B.8), (A.33) and (A.34). □

Lemma 62. *Under Assumptions 13, 14, 15, 16, 17, to prove that*

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

it suffices to show that we have

$$
\begin{aligned}
&\limsup_{n\to\infty} \max\Big\{ P_n\left(\hat{T}(\theta_{P_n,\ell}) > \tilde{c}^m(\theta_{P_n,\ell}, \hat{c}^t_{P_n}; \alpha) \text{ or } \hat{T}(\theta_{P_n,m}) > \tilde{c}^m(\theta_{P_n,m}, \hat{c}^t_{P_n}; \alpha)\right), \\
&P_n\left(\hat{T}(\theta_{P_n,u}) > \tilde{c}^m(\theta_{P_n,u}, \hat{c}^t_{P_n}; \alpha) \text{ or } \hat{T}(\theta_{P_n,m}) > \tilde{c}^m(\theta_{P_n,m}, \hat{c}^t_{P_n}; \alpha)\right)\Big\} \leqslant \alpha - \eta, \qquad \text{(A.35)}
\end{aligned}
$$

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

1. *The convergence of* Ω*,*

$$
\Omega(P_n) \to \Omega_0 \in \mathcal{S}. \qquad \text{(A.36)}
$$

2. *The convergence of*

$$
(\lambda_{n,k\ell}, \lambda_{n,ku}) = \left(\left(\frac{\lambda_{P_n,\ell,b} - \theta_{P_n,k}}{\sigma_{P_n,\ell,b}/\sqrt{n}}\right)_{b\in\mathcal{B}}, \left(\frac{\theta_{P_n,k} - \lambda_{P_n,u,b}}{\sigma_{P_n,u,b}/\sqrt{n}}\right)_{b\in\mathcal{B}}\right) \to (\lambda_{k\ell}, \lambda_{ku}), \quad \text{(A.37)}
$$

with $\lambda_{\ell\ell} \in \Lambda_0$, $\lambda_{uu} \in \Lambda_0$, $\lambda_{\ell,u}, \lambda_{mu} \in \Lambda_-$ *and* $\lambda_{m\ell}, \lambda_{u\ell} \in \Lambda_-$, $\sigma_{0,\ell,b} = \sqrt{A_{\ell,b}\Omega_0 A'_{\ell,b}}$, $\sigma_{0,u,b} = \sqrt{A_{u,b}\Omega_0 A'_{u,b}}$,

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

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

where

$$\hat{c}^t_{P_n} = \inf_c \left\{ c \geqslant 0 : \bar{p}\left(c, \lambda_{P_n}, \hat{\Sigma}_n/n\right) + \eta \leqslant \alpha \right\}. \tag{A.38}$$

Recall that $\tilde{c}^h(\theta, c; \alpha)$ is defined in (2.27) and $\bar{p}\left(c, \lambda_{P_n}, \hat{\Sigma}_n/n\right)$ is defined in (2.30).

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

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \sup_{\theta\in[\theta_{P,\ell},\theta_{P,u}]} P\left(\theta \notin CI^{\mathrm{m}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right) = \lim_{n_a} P_{n_a}\left(\theta_{n_a} \notin CI^{\mathrm{m}}\left(\hat{\lambda}_n, \hat{\Sigma}_n/n; \alpha\right)\right). \tag{A.39}$$

Since $\mathcal{S}$ defined in Assumption 15 is compact (e.g. in the Frobenius norm), and Assumption 15 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 (A.37) holds. We have found a subsequence n_s such that (A.36) and (A.37) hold. And, by (A.39), 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:

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

$$P_n\left([\theta_{P_n,m},\theta_{P_n,u}]\nsubseteq CI^{\mathrm{m}}\left(\hat{\lambda}_n,\hat{\Sigma}_n/n;\alpha\right),\lambda_{P_n}\in\hat{\Lambda}_n\right)\Big\}+P_{n_s}\left(\lambda_{P_n}\notin\hat{\Lambda}_n\right)$$

$$\leqslant\max\Big\{P_n\left(\hat{T}(\theta_{P_n,\ell})>\hat{c}^{\mathrm{m}}(\theta_{P_n,\ell};\alpha)\text{ or }\hat{T}(\theta_{P_n,m})>\hat{c}^{\mathrm{m}}(\theta_{P_n,m};\alpha),\lambda_{P_n}\in\hat{\Lambda}_n\right),$$

$$P_n\left(\hat{T}(\theta_{P_n,m})>\hat{c}^{\mathrm{m}}(\theta_{P_n,m};\alpha)\text{ or }\hat{T}(\theta_{P_n,u})>\hat{c}^{\mathrm{m}}(\theta_{P_n,u};\alpha),\lambda_{P_n}\in\hat{\Lambda}_n\right)\Big\}+\eta+o(1)$$

$$\leqslant\max\Big\{P_n\left(\hat{T}(\theta_{P_n,\ell})>\tilde{c}^{\mathrm{m}}(\theta_{P_n,\ell},\hat{c}^{\mathrm{t}}_{P_n};\alpha)\text{ or }\hat{T}(\theta_{P_n,m})>\tilde{c}^{\mathrm{m}}(\theta_{P_n,m},\hat{c}^{\mathrm{t}}_{P_n};\alpha)\right),$$

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

Recall that $\hat{c}^{\mathrm{t}}$ is defined in (2.31) and $\hat{c}^{\mathrm{t}}_{P_n}$ is defined in (A.38), thus the last inequality follows from the fact that $\hat{c}^{\mathrm{t}}_{P_n}\leqslant\hat{c}^{\mathrm{t}}$ if $\lambda_{P_n}\in\hat{\Lambda}_n$. Therefore it suffices to show (A.35). □

Lemma 63. *Under Assumptions 13, 14, 15, 16, 17, under sequences (A.36) and (A.37), if*

$$\min_{b\in\mathcal{B}}\sigma_{0,u,b}\lambda_{\ell u,b}\in\mathbb{R},\tag{A.40}$$

it holds that

$$\left(\hat{T}(\theta_{P_n,k}),\hat{c}^{c}(\theta_{P_n,k},\alpha^{c})\right)_{k=\ell,m,u}\xrightarrow{d}(T_k,c^{c}_{k})_{k=\ell,m,u}\,.\tag{A.41}$$

If

$$\min_{b\in\mathcal{B}}\sigma_{0,u,b}\lambda_{\ell u,b}=-\infty,\tag{A.42}$$

it holds that

$$\left(\hat{T}(\theta_{P_n,k}),\hat{c}^{c}(\theta_{P_n,k},\alpha^{c})\right)_{k=\ell,u}\xrightarrow{d}(T_k,c^{c}_{k})_{k=\ell,u}\,,\tag{A.43}$$

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

$$P_n\left(\hat{T}(\theta_{P_n,m})\geqslant c\right)\to 0.\tag{A.44}$$

Proof. Note that

$$\lim_n\sqrt{n}\left(\lambda_{P_n,\ell,b_\ell}-\lambda_{P_n,u,b_u}\right)=\lim_n\sigma_{P_n,u,b}\frac{\lambda_{P_n,\ell,b_\ell}-\max_{b\in\mathcal{B}}\lambda_{P_n,u,b}}{\sigma_{P_n,u,b}/\sqrt{n}}$$

$$=\lim_n\min_{b\in\mathcal{B}}\sigma_{P_n,u,b}\frac{\theta_{P_n,\ell}-\lambda_{P_n,u,b}}{\sigma_{P_n,u,b}/\sqrt{n}}$$

$$= \min_{b\in\mathcal{B}} \lim_n \sigma_{P_n,u,b} \frac{\theta_{P_n,\ell} - \lambda_{P_n,u,b}}{\sigma_{P_n,u,b}/\sqrt{n}} = \min_{b\in\mathcal{B}} \sigma_{0,u,b}\lambda_{\ell u,b}. \tag{A.45}$$

Thus the two cases in (A.40) and (A.42) correspond to whether the length of the identified set of θ is large asymptotically. I will show (A.41) under (A.40 in Step 1 and 2, then show (A.43) and (A.44) under (A.42) in Step 3.

Step 1. Show that under (A.40),

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

$$\xrightarrow{d} \left(b_{k\ell}, b_{ku}, T_k, \Phi\left(t_{k\ell}(\mathcal{B}_{k\ell})\right), \Phi\left(t_{ku}(\mathcal{B}_{ku})\right)\right)_{k=\ell,m,u}.$$

where

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

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

Step 1.1. Note that with probability one

$$\begin{aligned}
\hat{T}(\theta_{P_n,k}) &= \max\left\{\min_{b\in\mathcal{B}} \frac{\hat{\lambda}_{\ell,b} - \theta_{P_n,k}}{\hat{\sigma}_{\ell,b}/\sqrt{n}},\ \min_{b\in\mathcal{B}} \frac{\theta_{P_n,k} - \hat{\lambda}_{u,b}}{\hat{\sigma}_{u,b}/\sqrt{n}}\right\} \\
&= \max\left\{\min_{b\in\mathcal{B}_\ell} \frac{\hat{\lambda}_{\ell,b} - \lambda_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}/\sqrt{n}} + \frac{\sigma_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}}\lambda_{n,k\ell,b},\ \min_{b\in\mathcal{B}_u} \frac{\lambda_{P_n,u,b} - \hat{\lambda}_{u,b}}{\hat{\sigma}_{P_n,u,b}/\sqrt{n}} + \frac{\sigma_{P_n,u,b}}{\hat{\sigma}_{u,b}}\lambda_{n,ku,b}\right\} \\
&= \max\left\{\min_{b\in\mathcal{B}_{k\ell}} \frac{\hat{\lambda}_{\ell,b} - \lambda_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}/\sqrt{n}} + \frac{\sigma_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}}\lambda_{n,k\ell,b},\ \min_{b\in\mathcal{B}_{ku}} \frac{\lambda_{P_n,u,b} - \hat{\lambda}_{u,b}}{\hat{\sigma}_{u,b}/\sqrt{n}} + \frac{\sigma_{P_n,u,b}}{\hat{\sigma}_{u,b}}\lambda_{n,ku,b}\right\} \\
&\xrightarrow{d} \max\left\{\min_{b\in\mathcal{B}_{k\ell}} Z_{\ell,b} + \lambda_{k\ell,b},\ \min_{b\in\mathcal{B}_{ku}} Z_{u,b} + \lambda_{ku,b}\right\}.
\end{aligned} \tag{A.46}$$

The first line is by definition, the second line simply rearranges terms with $\lambda_{n,k\ell,b}$, $\lambda_{n,ku,b}$ defined in (A.37) and $\mathcal{B}_\ell$, $\mathcal{B}_u$ defined in Section B.1.1 first paragraph. To see the third line, note that by Assumption 13, 14, 15, 16, we have

$$\left(\left(\frac{\hat{\lambda}_{\ell,b} - \lambda_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}/\sqrt{n}}\right)_{b\in\mathcal{B}}, \left(\frac{\lambda_{P_n,u,b} - \hat{\lambda}_{u,b}}{\hat{\sigma}_{u,b}/\sqrt{n}}\right)_{b\in\mathcal{B}}, \left(\frac{\sigma_{P_n,\ell,b}}{\hat{\sigma}_{\ell,b}}\right)_{b\in\mathcal{B}}, \left(\frac{\sigma_{P_n,u,b}}{\hat{\sigma}_{u,b}}\right)_{b\in\mathcal{B}}\right) \tag{A.47}$$

$$\xrightarrow{d} \left(Z_\ell, Z_u, \mathbf{1}_{2|\mathcal{B}|}\right).$$

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,

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

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) (A.47), (ii) Slustsky's Lemma and (iii) the limit distribution is well defined because

$$\lambda_{k\ell,b} = \lim_n \frac{\lambda_{P_n,\ell,b} - \theta_{P_n,k}}{\sigma_{P_n,\ell,b}/\sqrt{n}} \geqslant \lim_n \frac{\lambda_{P_n,\ell,b_\ell} - \lambda_{P_n,u,b_u}}{\sigma_{P_n,\ell,b}/\sqrt{n}} = \frac{\min_{b\in\mathcal{B}} \sigma_{0,u,b}\lambda_{\ell u,b}}{\sigma_{0,\ell,b}} \in \mathbb{R}, \tag{A.48}$$

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

Step 1.2. As for $\Phi\left(\hat{t}_{\ell,1}(\theta_k, \mathcal{B}_{k\ell})\right)$, let $b \in \mathcal{B}_{k\ell}$. If $\min_{\tilde{b}\in\mathcal{B}} \rho_{\ell u}(b, \tilde{b}) = -1$, then $\Phi\left(t_{k\ell,1}(b)\right) = 0$ by construction in (A.76). 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 (A.48) 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

$$\Phi\left(t_{\ell,1}(\theta_k, \mathcal{B}_{k\ell})\right) \xrightarrow{d} \Phi\left(t_{k\ell,1}(\mathcal{B}_{k\ell})\right). \tag{A.49}$$

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

Step 1.3. Let

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

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

$$g\left(\mathcal{Z}_{\ell,b_1}, \mathcal{Z}_{\ell,b_2}\right) \xrightarrow{d} g\left(Z_{\ell,b_1} + \lambda_{k\ell,b_1}, Z_{\ell,b_2} + \lambda_{k\ell,b_2}\right). \tag{A.50}$$

Similarly, we have

$$g\left(\mathcal{Z}_{u,b_2}, \mathcal{Z}_{\ell,b_1}\right) \xrightarrow{d} g\left(Z_{u,b_2} + \lambda_{ku,b_2}, Z_{\ell,b_1} + \lambda_{k\ell,b_1}\right). \tag{A.51}$$

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

$$g\left(\mathcal{Z}_{\ell,b_1}, \mathcal{Z}_{u,b_2}\right) \xrightarrow{d} g\left(Z_{\ell,b_1} + \lambda_{k\ell,b_1}, Z_{u,b_2} + \lambda_{ku,b_2}\right). \tag{A.52}$$

(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 (A.46), (A.49), (A.50), (A.51), (A.52) 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

$$\begin{aligned}
&\Phi\left(\hat{c}^c(\theta_k, \alpha)\right) \\
&= \left((1-\alpha)\Phi\left(t_{\ell,2}(\theta_k, \hat{b}_\ell)\right) + \alpha\Phi\left(t_{\ell,1}(\theta_k, \hat{b}_\ell)\right)\right)\mathbf{1}\left[\mathcal{Z}_{\ell,\hat{b}_\ell} \geqslant \mathcal{Z}_{u,\hat{b}_u}\right]
\end{aligned}$$

$$+\left((1-\alpha)\Phi\left(t_{u,2}(\theta_k,\hat{b}_\ell)\right)+\alpha\Phi\left(t_{u,1}(\theta_k,\hat{b}_\ell)\right)\right)\mathbf{1}\left[\mathcal{Z}_{\ell,\hat{b}_\ell}<\mathcal{Z}_{u,\hat{b}_u}\right]$$

$$=\sum_{b_1\in\mathcal{B}_{k\ell}}\sum_{b_2\in\mathcal{B}_{ku}}\mathbf{1}\left[\mathcal{Z}_{\ell,b_1}\geqslant\mathcal{Z}_{u,b_2},\mathcal{Z}_{\ell,b_1}\leqslant\mathcal{Z}_{\ell,\mathcal{B}_{k\ell}\setminus b_1},\mathcal{Z}_{u,b_2}\leqslant\mathcal{Z}_{u,\mathcal{B}_{ku}\setminus b_2}\right]\times$$

$$\left((1-\alpha)\Phi\left(t_{u,2}(\theta_k,b_1)\right)+\alpha\Phi\left(t_{\ell,1}(\theta_k,b_2)\right)\right)$$

$$+\sum_{b_1\in\mathcal{B}_{k\ell}}\sum_{b_2\in\mathcal{B}_{ku}}\mathbf{1}\left[\mathcal{Z}_{\ell,b_1}<\mathcal{Z}_{u,b_2},\mathcal{Z}_{\ell,b_1}\leqslant\mathcal{Z}_{\ell,\mathcal{B}_{k\ell}\setminus b_1},\mathcal{Z}_{u,b_2}\leqslant\mathcal{Z}_{u,\mathcal{B}_{ku}\setminus b_2}\right]\times$$

$$\left((1-\alpha)\Phi\left(t_{u,2}(\theta_k,b_1)\right)+\alpha\Phi\left(t_{u,1}(\theta_k,b_2)\right)\right)\text{ w.p.a. 1}$$

And

$$\mathbf{1}\left[\mathcal{Z}_{\ell,b_1}\geqslant\mathcal{Z}_{u,b_2},\mathcal{Z}_{\ell,b_1}\leqslant\mathcal{Z}_{\ell,\mathcal{B}_{k\ell}\setminus b_1},\mathcal{Z}_{u,b_2}\leqslant\mathcal{Z}_{u,\mathcal{B}_{ku}\setminus b_2}\right]$$

$$=g\left(\mathcal{Z}_{u,b_2},\mathcal{Z}_{\ell,b_1}\right)\prod_{\tilde{b}_1\in\mathcal{B}_{k\ell}\setminus b_1}g\left(\mathcal{Z}_{\ell,b_1},\mathcal{Z}_{\ell,\tilde{b}_1}\right)\prod_{\tilde{b}_2\in\mathcal{B}_{k\ell}\setminus b_2}g\left(\mathcal{Z}_{\ell,b_2},\mathcal{Z}_{\ell,\tilde{b}_2}\right)$$

$$\mathbf{1}\left[\mathcal{Z}_{\ell,b_1}<\mathcal{Z}_{u,b_2},\mathcal{Z}_{\ell,b_1}\leqslant\mathcal{Z}_{\ell,\mathcal{B}_{k\ell}\setminus b_1},\mathcal{Z}_{u,b_2}\leqslant\mathcal{Z}_{u,\mathcal{B}_{ku}\setminus b_2}\right]$$

$$=\left[1-g\left(\mathcal{Z}_{u,b_2},\mathcal{Z}_{\ell,b_1}\right)\right]\prod_{\tilde{b}_1\in\mathcal{B}_{k\ell}\setminus b_1}g\left(\mathcal{Z}_{\ell,b_1},\mathcal{Z}_{\ell,\tilde{b}_1}\right)\prod_{\tilde{b}_2\in\mathcal{B}_{k\ell}\setminus b_2}g\left(\mathcal{Z}_{\ell,b_2},\mathcal{Z}_{\ell,\tilde{b}_2}\right)$$

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 (A.49), (A.50), (A.51), (A.52).

Step 3. Now assume (A.42) holds. We can show that

$$\left(\hat{b}_\ell(\theta_{P_n,k}),\hat{b}_u(\theta_{P_n,k}),\hat{T}(\theta_{P_n,k}),\Phi\left(\hat{t}_\ell(\theta_{P_n,k},\mathcal{B}_{k\ell})\right),\Phi\left(\hat{t}_u(\theta_{P_n,k},\mathcal{B}_{ku})\right)\right)_{k=\ell,u}$$

$$\xrightarrow{d}\left(b_{k\ell},b_{ku},T_k,\Phi\left(t_{k\ell}(\mathcal{B}_{k\ell})\right),\Phi\left(t_{ku}(\mathcal{B}_{ku})\right)\right)_{k=\ell,u}$$

with similar argument as Step 1 and 2. Regarding (A.44), note that

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

$$\leqslant\max\left\{\frac{\hat{\lambda}_{\ell,b_\ell}-\theta_m}{\hat{\sigma}_{\ell,b_\ell}/\sqrt{n}},\ \frac{\theta_m-\hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}}\right\}$$

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

By (A.42), (A.45) 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 (A.44) holds. □

Lemma 64. *Assume that (i) $\theta \leqslant \theta_\ell$; (ii)*

$$\hat{T}(\theta) = \mathcal{Z}_{\ell,\hat{b}_\ell} > \Phi^{-1}(1 - \frac{\alpha - \eta}{2}); \tag{A.53}$$

(iii) either

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

or

$$\left(1 + \hat{\rho}_{\ell u}(\hat{b}_\ell, b_u)\right)^{-1} \frac{\lambda_{u,b_u} - \hat{\lambda}_{u,b_u}}{\hat{\sigma}_{u,b_u}/\sqrt{n}} \leqslant \bar{M}, \tag{A.54}$$

$$\hat{T}(\theta) > \bar{z}, \tag{A.55}$$

where $\bar{M} \in \mathbb{R}$, $\bar{z}$ is defined in Lemma 65 with $\bar{M}$ given in (A.54). Then

$$\hat{T}(\theta) > \hat{c}^m(\theta, 1 - \alpha^c). \tag{A.56}$$

Proof. Note that $c^{\mathrm{t}} \leqslant \Phi^{-1}(1 - \frac{\alpha-\eta}{2})$ by construction, thus under (A.53), $\hat{T}(\theta) > c^{\mathrm{t}}$ and (A.56) is equivalent to

$$\hat{T}(\theta) > \hat{c}^{\mathrm{c}}(\theta, 1 - \alpha^{\mathrm{c}}). \tag{A.57}$$

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

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

In this case, (A.57) holds trivially. If $\min_{\tilde{b}\in\mathcal{B}} \rho_{\ell u}(\hat{b}_\ell, \tilde{b}) > -1$, we have

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

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

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

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

where $H(z,\bar{M})$ is defined in Lemma 65. $H(\mathcal{Z}_{\ell,\hat{b}_\ell},\bar{M})>0$ follows from (A.55) and Lemma 65. □

Lemma 65. *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 $z\geqslant\bar{z}$.

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 $z\geqslant\bar{z}$. Also note that

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

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

Lemma 66. *Let $\alpha \in (0, \frac{1}{2})$, $\alpha^c \in (\frac{\alpha}{2}, \alpha)$, $\eta \in \left[0, \frac{\alpha}{4}\right)$. Recall that $c^{sim} = \Phi^{-1}(1 - \frac{\alpha}{2})$. Let*

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

$$\rho_2^*(\alpha, \eta) = \sup_{\rho} \left\{ \rho : \sup_{\Delta \geqslant 0} H\left(c^{sim}, \Delta, \rho\right) \leqslant \alpha - \eta \right\}. \tag{A.58}$$

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

$$\sup_{\rho \leqslant \rho_2^*(\alpha,\eta) - \xi} \sup_{\Delta \geqslant 0} H(\bar{c}, \Delta, \rho) < \alpha - \eta. \tag{A.59}$$

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

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

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

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

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

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

$$\sup_{\Delta \geqslant 0} H(c, \Delta, \rho) = \sup_{\Delta \in [0, \bar{\Delta}]} H(c, \Delta, \rho) \tag{A.60}$$

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

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

Therefore, (A.60) holds for all $c \in (0, c^{\text{sim}}]$.

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

$$\alpha - \eta \geqslant \sup_{\Delta \in [0, \bar{\Delta}]} H\left(c^{\text{sim}}, \Delta, \rho_2^*(\alpha, \eta)\right)$$

$$= \sup_{\Delta\in[0,\bar{\Delta}]} H\left(c^{\text{sim}},\Delta,\bar{\rho}\right) + \frac{\mathrm{d}H\left(c^{\text{sim}},\Delta,\tilde{\rho}(\Delta)\right)}{\mathrm{d}\rho}\xi$$

$$\geqslant \sup_{\Delta\in[0,\bar{\Delta}]} H\left(c^{\text{sim}},\Delta,\bar{\rho}\right) + a\xi \tag{A.61}$$

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 (A.61) we get

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

Lastly,

$$\frac{\mathrm{d}H\left(c,\Delta,\rho\right)}{\mathrm{d}c} = -\phi(\Delta-c)\Phi\left(\frac{(\rho-1)}{\sqrt{1-\rho^2}}c - \frac{\Delta\rho}{\sqrt{1-\rho^2}}\right) \tag{A.62}$$

$$-\phi(-c)\Phi\left(\frac{c\rho-c+\Delta}{\sqrt{1-\rho^2}}\right) - \phi\left(-\frac{\Delta}{2}-c\right). \tag{A.63}$$

Let

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

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

$$\sup_{\Delta\in[0,\bar{\Delta}]} H\left(c^{\text{sim}},\Delta,\rho\right) = \sup_{\Delta\in[0,\bar{\Delta}]} H\left(\bar{c},\Delta,\rho\right) + \frac{\mathrm{d}H\left(\tilde{c}(\Delta),\Delta,\rho\right)}{\mathrm{d}c}(c^{\text{sim}}-\bar{c})$$

$$\geqslant \sup_{\Delta\in[0,\bar{\Delta}]} H\left(\bar{c},\Delta,\rho\right) - b(c^{\text{sim}}-\bar{c}). \tag{A.64}$$

In sum, for all $\rho \leqslant \rho_2^*(\alpha,\eta)-\xi$, by (A.60), (A.61), (A.64),

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

□

Lemma 67. *Suppose Assumptions 13, 14, 15, 16, and 17 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*

$$\sup_{P\in\mathcal{P}} \rho_\ell(b_\ell, b_u) < \rho_2^*(\alpha, \eta), \tag{A.65}$$

where $\rho_2^*(\alpha, \eta)$ *is defined in Lemma 66 equation (A.58). Then there is* $\alpha' > \alpha$ *such that*

$$\liminf_n \inf_{P\in\mathcal{P}} P\left(\hat{c}^t \leqslant \Phi^{-1}(1 - \frac{\alpha'}{2})\right) = 1. \tag{A.66}$$

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 66, there is $\bar{c} < \Phi^{-1}(1 - \frac{\alpha}{2})$ such that (A.59) holds. To show (A.66), note that

$$\begin{aligned}\liminf_n \inf_{P\in\mathcal{P}} P\left(\hat{c}^{\mathrm{t}} \leqslant \bar{c}\right) &\geqslant \liminf_n \inf_{P\in\mathcal{P}} P\left(\sup_{\lambda\in\hat{\Lambda}} \bar{p}(\bar{c}, \lambda) \leqslant \alpha - \eta\right) \\ &\geqslant \liminf_n \inf_{P\in\mathcal{P}} P\left(\sup_{\lambda\in\Lambda} \bar{p}(\bar{c}, \lambda) \leqslant \alpha - \eta\right)\end{aligned}$$

Recall that

$$\begin{aligned}\bar{p}(\bar{c}, \lambda) = &\max\left\{P^s\left(\hat{T}(\theta_\ell) > \tilde{c}^{\mathrm{m}}(\theta_\ell, \bar{c}) \vee \left\{\hat{T}(\theta_m) > \tilde{c}^{\mathrm{m}}(\theta_m, \bar{c}) \wedge \hat{T}(\theta_u) > \tilde{c}^{\mathrm{m}}(\theta_u, \bar{c})\right\}; (\lambda, \hat{\Sigma})\right)\right. \\ &\left. P^s\left(\hat{T}(\theta_u) > \tilde{c}^{\mathrm{m}}(\theta_u, \bar{c}) \vee \left\{\hat{T}(\theta_m) > \tilde{c}^{\mathrm{m}}(\theta_m, \bar{c}) \wedge \hat{T}(\theta_\ell) > \tilde{c}^{\mathrm{m}}(\theta_\ell, \bar{c})\right\}; (\lambda, \hat{\Sigma})\right)\right\} \\ \leqslant &\max\left\{P^s\left(\hat{T}(\theta_\ell) > \bar{c} \text{ or } \hat{T}(\theta_m) > \bar{c}; (\lambda, \hat{\Sigma})\right),\ P^s\left(\hat{T}(\theta_m) > \bar{c} \text{ or } \hat{T}(\theta_u) > \bar{c}; (\lambda, \hat{\Sigma})\right)\right\}.\end{aligned}$$

I will show that

$$\sup_{\lambda\in\Lambda} P^s\left(\hat{T}(\theta_\ell) > \bar{c} \text{ or } \hat{T}(\theta_m) > \bar{c}; (\lambda, \hat{\Sigma})\right) \leqslant \alpha - \eta \text{ w.p.a. } 1,$$

and similarly we can show that

$$\sup_{\lambda\in\Lambda} P^s\left(\hat{T}(\theta_m) > \bar{c} \text{ or } \hat{T}(\theta_u) > \bar{c}; (\lambda, \hat{\Sigma})\right) \leqslant \alpha - \eta \text{ w.p.a. } 1.$$

To see this,

$$
\begin{aligned}
&P^s\left(\hat{T}(\theta_\ell) > \bar{c} \text{ or } \hat{T}(\theta_m) > \bar{c}; (\lambda, \hat{\Sigma})\right) \\
=&P^s\left(\max\left\{\min_{b\in\mathcal{B}} \frac{\hat{\lambda}_b^s - \theta_\ell}{\hat{\sigma}_b/\sqrt{n}}, \min_{b\in\mathcal{B}} \frac{\theta_\ell - \hat{\lambda}_b^s}{\hat{\sigma}_b/\sqrt{n}}\right\} > \bar{c}\right. \\
&\quad \left.\text{or } \max\left\{\min_{b\in\mathcal{B}} \frac{\hat{\lambda}_b^s - \theta_m}{\hat{\sigma}_b/\sqrt{n}}, \min_{b\in\mathcal{B}} \frac{\theta_m - \hat{\lambda}_b^s}{\hat{\sigma}_b/\sqrt{n}}\right\} > \bar{c}; (\lambda, \hat{\Sigma})\right) \\
\leqslant&P^s\left(\max\left\{\min_{b\in\mathcal{B}} \frac{\hat{\lambda}_b^s - \theta_\ell}{\hat{\sigma}_b/\sqrt{n}}, \min_{b\in\mathcal{B}} \frac{\theta_m - \hat{\lambda}_b^s}{\hat{\sigma}_b/\sqrt{n}}\right\} > \bar{c}; (\lambda, \hat{\Sigma})\right) \\
\leqslant&P^s\left(\min\left\{\frac{\hat{\lambda}_{b_\ell}^s - \theta_\ell}{\hat{\sigma}_{b_\ell}/\sqrt{n}}, \frac{\hat{\lambda}_{b_u}^s - \theta_\ell}{\hat{\sigma}_{b_u}/\sqrt{n}}\right\} > \bar{c}, \text{ or } \frac{\theta_m - \hat{\lambda}_{b_u}^s}{\hat{\sigma}_{b_u}/\sqrt{n}} > \bar{c}; (\lambda, \hat{\Sigma})\right) \\
=&P^s\left(\min\left\{\mathbb{Z}_{b_\ell}^s, \mathbb{Z}_{b_u}^s + \frac{\theta_u - \theta_\ell}{\hat{\sigma}_{b_u}/\sqrt{n}}\right\} > \bar{c}, \text{ or } \frac{\theta_m - \theta_u}{\hat{\sigma}_{b_u}/\sqrt{n}} - \mathbb{Z}_{b_u}^s > \bar{c}; (\lambda, \hat{\Sigma})\right) \\
\leqslant&P^s\left(\mathbb{Z}_{b_\ell}^s > \bar{c}, \ \mathbb{Z}_{b_u}^s + \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}_{b_u}^s > \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) \\
\leqslant&\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)
\end{aligned} \tag{A.67}
$$

where $\Delta = \frac{\theta_u - \theta_\ell}{\hat{\sigma}_{b_u}/\sqrt{n}} \geqslant 0$,

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

Under (A.65) and Assumptions 13, 14, 15, 16, and 17, it holds that

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

Thus (A.67) gives w.p.a. 1,

$$
P\left(\hat{T}(\theta_\ell) > \bar{c} \text{ or } \hat{T}(\theta_m) > \bar{c}; (\lambda, \hat{\Sigma})\right) \leqslant H\left(\bar{c}, \Delta, \hat{\rho}_\ell(b_\ell, b_u)\right)
$$

$$\leqslant \sup_{\rho \leqslant \rho_2^*(\alpha,\eta)-\xi} \sup_{\Delta \geqslant 0} H\left(\bar{c}, \Delta, \rho\right) < \alpha - \eta$$

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

Lemma 68. *$p(c)$ is continuous at $c \geqslant 0$.*

Proof. For $\varepsilon > 0$, let

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

Then

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

for all $c \geqslant 0$ since (i) under (A.40) and $k = \ell, m, u$, or under (A.42) and $k = \ell, u$, T_k is continuously distributed, (ii) under (A.42) and $k = m$,

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

But then

$$\begin{aligned}
&p(c-\varepsilon) - p(c) \\
\leqslant &\max\left\{P\left(T_\ell > c_\ell^{\mathrm{h}}(c-\varepsilon) \text{ or } T_m > c_m^{\mathrm{h}}(c-\varepsilon)\right) - P\left(T_\ell > c_\ell^{\mathrm{h}}(c) \text{ or } T_m > c_m^{\mathrm{h}}(c)\right),\right. \\
&\left. P\left(T_m > c_m^{\mathrm{h}}(c-\varepsilon) \text{ or } T_u > c_u^{\mathrm{h}}(c-\varepsilon)\right) - P\left(T_m > c_m^{\mathrm{h}}(c) \text{ or } T_u > c_u^{\mathrm{h}}(c)\right)\right\} \\
\leqslant &\max\left\{p_\ell(c,\varepsilon) + p_m(c,\varepsilon),\ p_u(c,\varepsilon) + p_m(c,\varepsilon)\right\} \\
\xrightarrow{\varepsilon \to 0} &\to 0
\end{aligned}$$

Thus $p(c)$ is continous at $c \geqslant 0$. □

Lemma 69. *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

$$H\left(\rho^*(\alpha^c, \alpha), \alpha^c\right) = 1 - \frac{\alpha}{2}. \tag{A.68}$$

2. Let $\xi > 0$. For all $\rho \in [0, \rho^*(\alpha^c, \alpha) - \xi]$, there is $\varepsilon > 0$ such that

$$H\left(\rho, \alpha^c\right) \leqslant 1 - \frac{\alpha}{2} - \varepsilon.$$

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

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

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

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

thus

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

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

where the inequality follows from $\alpha^{\mathrm{c}} \in (\frac{\alpha}{2}, \alpha)$. 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 70. *Suppose Assumptions 13, 14, 15, 16, and 17 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

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

The lower bound is

$$t_{u,1} = \min_{\tilde{b}\in\mathcal{B}} \left(1 + \rho(\tilde{b}, 1)\right)^{-1} \left(\mathcal{Z}_{\ell,\tilde{b}} + \rho(\tilde{b}, 1)\mathcal{Z}_{u,1}\right) \leqslant \frac{\rho_{12}\mathcal{Z}_1 - \mathcal{Z}_2}{1 + \rho_{12}}$$

and the upper bound is

$$t_{u,2} = \min_{\tilde{b}\in\mathcal{B}:\rho_u(\tilde{b},1)<1} \left(1 - \rho_u(\tilde{b}, 1)\right)^{-1} \left(\mathcal{Z}_{u,\tilde{b}} - \rho_u(\tilde{b}, 1)\mathcal{Z}_{u,1}\right) \leqslant \frac{\mathcal{Z}_2 - \rho_{12}\mathcal{Z}_1}{1 - \rho_{12}}.$$

This θ is rejected if

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

Since by construction, $\mathcal{Z}_2 \geqslant \mathcal{Z}_1$, it sufficies to show that

$$\Phi(\mathcal{Z}_1) > \sup_{z_2\geqslant\mathcal{Z}_1} (1 - \alpha^{\mathrm{c}})\Phi\left(\frac{z_2 - \rho_{12}\mathcal{Z}_1}{1 - \rho_{12}}\right) + \alpha^{\mathrm{c}}\Phi\left(\frac{\rho_{12}\mathcal{Z}_1 - z_2}{1 + \rho_{12}}\right). \tag{A.69}$$

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

$$\begin{aligned} H(\mathcal{Z}_1) &= (1 - \alpha^{\mathrm{c}})\Phi\left(\mathcal{Z}_1\right) + \alpha^{\mathrm{c}}\Phi\left(\frac{\rho_{12} - 1}{1 + \rho_{12}}\mathcal{Z}_1\right) \\ &< (1 - \alpha^{\mathrm{c}})\Phi\left(\mathcal{Z}_1\right) + \frac{\alpha^{\mathrm{c}}}{2} < \Phi\left(\mathcal{Z}_1\right), \end{aligned}$$

where the second line follows from $\mathcal{Z}_1 \geqslant \bar{c} > 0$. 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 $h(z_2) \geqslant 0$ is equivalent to

$$\begin{aligned}\log\left(\frac{1-\alpha^{\mathrm{c}}}{\alpha^{\mathrm{c}}}\frac{1+\rho_{12}}{1-\rho_{12}}\right) &\geqslant \log\left(\phi\left(\frac{\rho_{12}\mathcal{Z}_1-z_2}{1+\rho_{12}}\right)\right) - \log\left(\phi\left(\frac{z_2-\rho_{12}\mathcal{Z}_1}{1-\rho_{12}}\right)\right) \\ &= \frac{2\rho_{12}\left(z_2-\mathcal{Z}_1\rho_{12}\right)^2}{\left(1-\rho_{12}^2\right)^2}. \end{aligned} \tag{A.70}$$

(i) If $2\alpha^c - 1 < \rho \leqslant 0$, then (A.70) holds trivially, and thus

$$\sup_{z_2 \geqslant \mathcal{Z}_1} H(z_2) = \lim_{z_2 \to \infty} H(z_2) < \Phi\left(\mathcal{Z}_1\right),$$

and (A.69) holds.

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

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

Thus

$$\sup_{z_2 \geqslant \mathcal{Z}_1} H(z_2) \leqslant \max\left\{H(\mathcal{Z}_1), \lim_{z_2 \to \infty} H(z_2)\right\} < \Phi\left(\mathcal{Z}_1\right).$$

(iii) if $\rho \in \left(0, \rho^*(\alpha, \alpha^c) - \frac{\xi}{2}\right)$, straightforward calculation shows that $H(z_2)$ increases in $[\max\{\mathcal{Z}_1, z_2^*\}, z_2^*]$ and decreases in $[z_2^*, +\infty)$, thus

$$\sup_{z_2 \geqslant \mathcal{Z}_1} H(z_2) \leqslant H(z_2^*) = (1-\alpha^{\mathrm{c}})\Phi\left((1+\rho)\Xi\right) + \alpha^{\mathrm{c}}\Phi\left((\rho-1)\Xi\right) < \Phi\left(\mathcal{Z}_1\right),$$

where the last inequality is by Lemma 69. □

Lemma 71. *Suppose Assumptions 13, 14, 15, and 16 hold. And $\hat{\Lambda}_\eta$ is defined as in (A.1). 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 (2.36). It holds that*

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

Proof. If $\eta = 0$, trivial, following from the discussion around (2.17). If $\eta \in (0, \frac{\alpha - \alpha^c}{2})$, trivial too, as $\hat{c}^{\mathrm{t}} \xrightarrow{p} \Phi^{-1}(1 - \alpha + \eta)$. □

A.2 Additional Results for Infinite $\mathcal{B}$

A.2.1 Simulation Details

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

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

where

$$\varphi^{\dagger}(a_1, a_2, a_3, a_4) = \frac{a_1 - a_3 a_4}{a_2 - a_4^2},$$

$$m_{\varphi_f}(W) = \left(dr_h r_{f'}, dr_h^2, dr_f, dr_h\right) \mathbf{1}[f' = f].$$

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

$$m_1(W; \beta, \varphi) = -d\frac{1 - \Phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)}{\Phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)} + (1 - d)$$

$$m_2(W; \beta, \varphi) = -(1 - d)\frac{\Phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)}{1 - \Phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)} + d$$

$$m_3(W; \beta, \varphi) = (1 - d)\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right) - d\frac{\phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)}{\Phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)}$$

$$m_4(W; \beta, \varphi) = -d\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right) - (1 - d)\frac{\phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)}{1 - \Phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)}$$

The moment conditions are given by

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

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

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

$$= \lambda^{\dagger}\left(E\left[\Phi\left(\eta^{-1}\varphi_f r_h \beta_1 - \beta_2\right)\right], E\left[\Phi\left(\eta^{-1}\varphi_f X_1 \beta_1 - \beta_2\right)\right]\right)$$

where

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

A.2.2 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

$$r_{ij} = \varphi_j r_{ih} + e_{ij} \tag{A.71}$$

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

$$E_{jt}\left[e_{ij} \mid \mathcal{J}_{ij}, r_{ih}, f_{ij}\right] = 0 \tag{A.72}$$

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

$$d_{ij} = \mathbf{1}\left\{\eta^{-1} E\left[\varphi_j r_{ih} \mid \mathcal{J}_{ij}\right] \beta_1 - \beta_2 - \beta_3 dist_j \geqslant v_{ij}/\tilde{\beta}_3\right\} \tag{A.73}$$

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

$$E\left[d_{ij} \mid \mathcal{J}_{ij}, dist_j\right] = \Phi\left(\eta^{-1} E\left[\varphi_j r_{ih} \mid \mathcal{J}_{ijt}\right] \beta_1 - \beta_2 - \beta_3 dist_j\right) \quad . \tag{A.74}$$

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

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

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

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

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 $\theta = \frac{E[N_{c2}^{ex}]}{E[N^{ex}]}$.

A.3 Proofs for Section 2.3

A.3.1 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 62 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

$$T_k = \max\left\{\min_{b \in \mathcal{B}} Z_{\ell,b} + \lambda_{k\ell,b},\ \min_{b \in \mathcal{B}} Z_{u,b} + \lambda_{ku,b}\right\} \tag{A.75}$$

be the asymptotic analog of $\hat{T}(\theta_{P_n,k})$, where $(\lambda_{k\ell}, \lambda_{ku})$ are specified in Lemma 62. 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 $\mathcal{B}_{k\ell}$, $\mathcal{B}_{ku}$:

$$b_{k\ell} = \min\left\{\arg\min_{b \in \mathcal{B}_\ell} Z_{\ell,b} + \lambda_{k\ell,b}\right\},$$

$$b_{ku} = \min\left\{\arg\min_{b \in \mathcal{B}_u} Z_{u,b} + \lambda_{ku,b}\right\},$$

$$\mathcal{B}_{k\ell} = \{b \in \mathcal{B}_\ell : \lambda_{k\ell,b} < \infty\},$$

$$\mathcal{B}_{ku} = \{b \in \mathcal{B}_u : \lambda_{ku,b} < \infty\}.$$

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

$$t_{k\ell,1}(b) = \begin{cases} \min\limits_{\tilde{b}\in\mathcal{B}} \left(1+\rho_{\ell u}(b,\tilde{b})\right)^{-1} \left(Z_{u,\tilde{b}} + \rho_{\ell u}(b,\tilde{b}) Z_{\ell,b} + t^{\dagger}_{k\ell,1}(b,\tilde{b})\right), & \text{if } \min\limits_{\tilde{b}\in\mathcal{B}} \rho_{\ell u}(b,\tilde{b}) > -1 \\ -\infty & \text{elsewhere} \end{cases} \tag{A.76}$$

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

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

$$t_{ku,2}(b) = \begin{cases} A & \text{if } \min\limits_{\tilde{b}\in\mathcal{B}} \rho_{u}(b,\tilde{b}) < 1 \\ +\infty & \text{elsewhere} \end{cases}$$

where

$$A = \min_{\tilde{b}\in\mathcal{B}:\rho_u(\tilde{b},b)<1} \left(1-\rho_u(\tilde{b},b)\right)^{-1} \left(Z_{u,\tilde{b}} - \rho_u(\tilde{b},b) Z_{u,b} + t^{\dagger}_{ku,2}(b,\tilde{b})\right)$$

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

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

$$t^{\dagger}_{k\ell,2}(b,\tilde{b}) = \lambda_{k\ell,\tilde{b}} - \rho_{\ell}(b,\tilde{b})\lambda_{k\ell,b}$$

$$t^{\dagger}_{ku,2}(b,\tilde{b}) = \lambda_{ku,\tilde{b}} - \rho_{u}(\tilde{b},b)\lambda_{ku,b}.$$

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

$$c^{\mathrm{c}}_k = \begin{cases} \Phi^{-1}\left(\alpha^{\mathrm{c}}\Phi\left(t_{k\ell,1}(b_{k\ell})\right) + (1-\alpha^{\mathrm{c}})\Phi\left(t_{k\ell,2}(b_{k\ell})\right)\right) & \text{if } Z_{\ell,b_{k\ell}} + \lambda_{k\ell,b_{k\ell}} \geqslant Z_{u,b_{ku}} + \lambda_{ku,b_{ku}} \\ \Phi^{-1}\left(\alpha^{\mathrm{c}}\Phi\left(t_{ku,1}(b_{ku})\right) + (1-\alpha^{\mathrm{c}})\Phi\left(t_{ku,2}(b_{ku})\right)\right) & \text{if } Z_{\ell,b_{k\ell}} + \lambda_{k\ell,b_{k\ell}} < Z_{u,b_{ku}} + \lambda_{ku,b_{ku}} \end{cases} \tag{A.77}$$

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

$$p(c) = \max\{P\left(T_\ell > c_\ell^{\mathrm{m}}(c) \text{ or } \{T_m > c_m^{\mathrm{m}}(c) \text{ and } T_u > c_u^{\mathrm{m}}(c)\}\right), \tag{A.78}$$
$$P\left(T_u > c_u^{\mathrm{m}}(c) \text{ or } \{T_m > c_m^{\mathrm{m}}(c) \text{ and } T_\ell > c_\ell^{\mathrm{m}}(c)\}\right)\},$$

where

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

and T_k, c_k^{c} are defined in (A.75), (A.77). Lastly, let

$$c^t = \inf_c \{c \geqslant 0 : p(c) \leqslant \alpha - \eta\}, \tag{A.79}$$

be the asymptotic analog of $\hat{c}^{\mathrm{t}}$ defined (2.31).

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

A.4 Proof for Section 2.4

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.

A.4.1 Notation

The ε expansion of set A is defined as

$$A^\epsilon = \left\{a \in \mathbb{R}^{d_a} : d_H(a, A) \leqslant \epsilon\right\},$$

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 $\Omega_P\left((\beta,\varphi),(\tilde{\beta},\tilde{\varphi})\right)\in\mathbb{R}^{K\times K}$ denote

$$\Omega_P\left((\beta,\varphi),(\tilde{\beta},\tilde{\varphi})\right)=\operatorname{cov}\left(\ m(X_i;\beta,\varphi)\ ,\ m'(X_i;\tilde{\beta},\tilde{\varphi})\ \right),$$

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

$$\widetilde{\mathbb{G}}_n(\beta,\varphi)=\begin{pmatrix}\widetilde{\mathbb{G}}_{n,\mathcal{J}}(\beta,\varphi)\\ \widetilde{\mathbb{G}}_{n,\ell}(\beta,\varphi)\\ \widetilde{\mathbb{G}}_{n,u}(\beta,\varphi)\\ \widetilde{\mathbb{G}}_{n,\varphi}\end{pmatrix}=\begin{pmatrix}\sqrt{n}\left(\bar{m}_{\mathcal{J}}(\beta,\varphi)-E\left[m_{\mathcal{J}}(X_i;\beta,\varphi)\right]\right)\\ \sqrt{n}\left(\bar{m}_{\ell}(\beta,\varphi)-E\left[m_{\ell}(X_i;\beta,\varphi)\right]\right)\\ \sqrt{n}\left(\bar{m}_{u}(\beta,\varphi)-E\left[m_{u}(X_i;\beta,\varphi)\right]\right)\\ \sqrt{n}\left(\bar{m}_{\varphi}-E\left[m_{\varphi}(X_i)\right]\right)\end{pmatrix}. \tag{A.80}$$

A.4.2 Additional Assumptions

Assumption 72. *All distributions $P\in\mathcal{P}$ satisfy the following:*

1. *$E_P\left[m_j\left(X_i,\beta\right)\right]\leqslant 0, j=1,\ldots,J_1$ and $E_P\left[m_j\left(X_i,\beta\right)\right]=0, j=J_1+1,\ldots,J_1+J_2$ for some $\beta\in\overline{\mathcal{B}}$;*
2. *$\{X_i,i\geqslant 1\}$ are i.i.d.;*
3. *There is $\epsilon>0$ such that $\sigma^2_{P,j}(\beta)\in[\epsilon,1/\epsilon]$ 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*

$$E_P\left[\sup_{\beta\in\mathcal{B}}\left|m_j\left(X_i;\beta,\varphi_P\right)\right|^{2+\delta}\right]\leqslant M,\ \forall j=1,...,K.$$

Assumption 73. *All distributions $P\in\mathcal{P}$ satisfy one of the following two conditions for some constants $\omega>0,\varepsilon>0,M<\infty$. The functions $m_j\left(X_i;\beta,\varphi\right)$, $j=1,...,K$, are defined on $\mathcal{X}\times\overline{\mathcal{B}}^{\varepsilon}\times\Phi^{\varepsilon}$. There exists $R_1\in\mathbb{N}$, $1\leqslant R_1\leqslant J_1/2$ 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 $j\in\mathcal{R}_1$,*

$$m_{j+R_1}\left(X;\beta,\varphi\right)=-m_j\left(X;\beta,\varphi\right)-t_j\left(X;\beta,\varphi\right)$$

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*

$$\inf_{\beta\in\mathcal{B}(P)}\operatorname{eig}\left(\Sigma_{\ddot{J}_k}\right)\geqslant\omega,$$

where

$$\ddot{\mathcal{J}}_k = \left\{\ddot{j} : j \in \mathcal{R}_1 \cap \mathcal{J}_1(P,\beta,\varepsilon)\right\}$$
$$\cup\, \mathcal{J}_1(P,\beta,\varepsilon)\backslash\{1,...,2R_1\}$$
$$\cup\, \mathcal{J}_k(P,\beta;\varepsilon).$$

and $\mathcal{J}_1$ and $\mathcal{J}_k$ are defined in (2.65) and (2.66).

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

$$\inf_{\beta\in\mathcal{B}(P)} \operatorname{eig}\left(\Sigma_{\mathring{\mathcal{J}}_\ell}\right) \geqslant \omega, \text{ and } \inf_{\beta\in\mathcal{B}(P)} \operatorname{eig}\left(\Sigma_{\mathring{\mathcal{J}}_u}\right) \geqslant \omega.$$

Assumption 73 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 74. *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*

$$\lim_{\delta\downarrow 0} \sup_{\left\|\left((\beta_1,\varphi_1),(\tilde{\beta}_1,\tilde{\varphi}_1)\right)-\left((\beta_2,\varphi_2),(\tilde{\beta}_2,\tilde{\varphi}_2)\right)\right\|<\delta} \sup_{P\in\mathcal{P}}$$

$$\left\|\Omega_P\left((\beta_1,\varphi_1),(\tilde{\beta}_1,\tilde{\varphi}_1)\right)-\Omega_P\left((\beta_2,\varphi_2),(\tilde{\beta}_2,\varphi_2)\right)\right\|=0$$

A.4.3 Details of the Inference Procedure

If Assumption 73 is invoked, we make the following adjustment to the inference procedure.

In (2.51), 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

$$\hat{\sigma}_j^M(\beta)=\hat{\sigma}_{[j]}^M(\beta)=\hat{\mu}_{n,j}(\beta)\hat{\sigma}_j(\beta)+\left(1-\hat{\mu}_{n,j}(\beta)\right)\hat{\sigma}_{[j]}(\beta). \tag{A.81}$$

with

$$\begin{aligned}\hat{\mu}_{n,[j]}(\beta)&=\min\left\{\max\left(0,\frac{\frac{\bar{m}_{n,j}(\beta,\hat{\varphi})}{\hat{\sigma}_{n,j}(\beta,\hat{\varphi})}}{\frac{\bar{m}_{n,[j]}(\beta,\hat{\varphi})}{\hat{\sigma}_{n,[j]}(\beta,\hat{\varphi})}+\frac{\bar{m}_{n,j}(\beta,\hat{\varphi})}{\hat{\sigma}_{n,j}(\beta,\hat{\varphi})}}\right),1\right\},\\ \hat{\mu}_{n,j}(\beta)&=1-\hat{\mu}_{n,[j]}(\beta,\hat{\varphi}).\end{aligned} \tag{A.82}$$

In (2.55), 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 73.2 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 (2.50), 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 (2.53), (2.54).

A.4.4 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

$$\gamma_{1,P,j}(\beta)=\sigma_{P,j}^{-1}(\beta)E_P\left[m_j\left(X_i,\beta,\varphi_P\right)\right],j=1,...,J, \tag{A.83}$$

$$\gamma_{1,P,0}(\beta)=\frac{\lambda_{P,\ell}(\beta,\varphi_P)-\lambda_{P,u}(\beta,\varphi_P)}{\max\left\{\sigma_{P,u}(\beta),\sigma_{P,\ell}(\beta),\sigma_{P,u\ell}(\beta)\right\}} \tag{A.84}$$

$\gamma_{2,P}=\left(\text{vech}\left(\Omega_P(\beta)\right),\text{vec}\left(D_P(\beta)\right),\text{vec}\left(G_P(\beta)\right)\right)$, 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

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\beta\in\mathcal{B}(P)} \inf_{\theta\in[\lambda_\ell(\beta,\varphi_P),\lambda_u(\beta,\varphi_P)]} P\left(\theta_n \in CI_n\right) = \liminf_{n\to\infty} P_n\left(\theta_n \in CI_n\right). \tag{A.85}$$

Let $\{l_n\}$ be a subsequence of $\{n\}$ such that

$$\liminf_{n\to\infty} P_n\left(\theta_n \in CI_n\right) = \lim_{n\to\infty} P_{l_n}\left(\theta_{l_n} \in CI_{l_n}\right). \tag{A.86}$$

Then there is a further subsequence $\{a_n\}$ of $\{l_n\}$ such that

$$\lim_{a_n\to\infty} \kappa_{a_n}^{-1}\sqrt{a_n}\gamma_{1,P_{a_n},j}\left(\beta_{a_n}\right) = \pi_{1,j} \in \mathbb{R}_{[-\infty]}, \quad j = 0,1,\ldots,J. \tag{A.87}$$

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

$$\pi_{1,j}^* = \begin{cases} 0 & \text{if } \pi_{1,j} = 0 \\ -\infty & \text{if } \pi_{1,j} < 0 \end{cases}. \tag{A.88}$$

The true value θ_n is covered when

$$\left\{ \begin{array}{c} \inf_{\tilde{\beta}} \hat{\lambda}_\ell(\tilde{\beta}) - \frac{\hat{\sigma}_\ell(\tilde{\beta})}{\sqrt{n}}\hat{c}(\tilde{\beta}) \\ \text{s.t. } \tilde{\beta}\in\overline{\mathcal{B}}, \quad \frac{\sqrt{n}\bar{m}_{n,j}(\tilde{\beta},\hat{\varphi})}{\hat{\sigma}_j(\tilde{\beta})} \leqslant \hat{c}(\tilde{\beta}), \forall j = 1,...,J \end{array} \right\} \leqslant \theta_n$$

$$\leqslant \left\{ \begin{array}{c} \sup_{\tilde{\beta}} \hat{\lambda}_u(\tilde{\beta}) + \frac{\hat{\sigma}_u(\tilde{\beta})}{\sqrt{n}}\hat{c}(\tilde{\beta}) \\ \text{s.t. } \tilde{\beta}\in\overline{\mathcal{B}}, \quad \frac{\sqrt{n}\bar{m}_{n,j}(\tilde{\beta},\hat{\varphi})}{\hat{\sigma}_j(\tilde{\beta})} \leqslant \hat{c}(\tilde{\beta}), \forall j = 1,...,J \end{array} \right\}$$

$$\Leftrightarrow \left\{ \begin{array}{c} \inf_\Delta \frac{\sqrt{n}\left(\hat{\lambda}_\ell\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)-\theta_n\right)}{\hat{\sigma}_\ell\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)} - \hat{c}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right) \\ \text{s.t. } \Delta\in\frac{\sqrt{n}}{\rho}\left(\overline{\mathcal{B}}-\beta_n\right), \frac{\sqrt{n}\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)}{\hat{\sigma}_j\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)} \leqslant \hat{c}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right), \forall j = 1,...,J \end{array} \right\} \leqslant 0 \text{ and}$$

$$\left\{ \begin{array}{c} \inf_\Delta \frac{\sqrt{n}\left(\theta_n-\hat{\lambda}_u\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)\right)}{\hat{\sigma}_u\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)} - \hat{c}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right) \\ \text{s.t. } \Delta\in\frac{\sqrt{n}}{\rho}\left(\overline{\mathcal{B}}-\beta_n\right), \quad \frac{\sqrt{n}\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)}{\hat{\sigma}_j\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)} \leqslant \hat{c}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right), \forall j = 1,...,J \end{array} \right\} \leqslant 0 \tag{A.89}$$

where in (A.89), 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

$$\begin{aligned}\frac{\sqrt{n}\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)}{\hat{\sigma}_j\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)}=&\left(1+\frac{\sigma_{P_n,j}\left(\beta_n\right)}{\hat{\sigma}_j\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)}-1\right)\times\left(\frac{\widetilde{\mathbb{G}}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\varphi_{P_n}\right)}{\sigma_{P_n,j}\left(\beta_n\right)}\right.\\&+\frac{\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)-\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\varphi_{P_n}\right)}{\sigma_{P_n,j}\left(\beta_n\right)/\sqrt{n}}\qquad(A.90)\\&\left.\frac{E_{P_n}\left[m_j\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\varphi_{P_n}\right)\right]-E_{P_n}\left[m_j\left(\beta_n,\varphi_{P_n}\right)\right]}{\sigma_{P_n,j}\left(\beta_n\right)/\sqrt{n}}+\sqrt{n}\gamma_{1,n,j}\left(\beta_n\right)\right).\qquad(A.91)\end{aligned}$$

Then I further simplify each element. For (A.90),

$$\begin{aligned}&\frac{\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)-\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\varphi_{P_n}\right)}{\sigma_{P_n,j}\left(\beta_n\right)/\sqrt{n}}\\=&\frac{E_{P_n}\left[m_j\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\varphi\right)\right]_{\varphi=\hat{\varphi}}-E_{P_n}\left[m_j\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\varphi_{P_n}\right)\right]}{\sigma_{P_n,j}\left(\beta_n\right)/\sqrt{n}}+\zeta_{n,j}\\=&\frac{\nabla_{\varphi'}E_{P_n}\left[m_j\left(\beta_n,\bar{\varphi}\right)\right]\nabla_{m'_\varphi}\bar{\varphi}^\dagger}{\sigma_{P_n,j}\left(\beta_n\right)}\widetilde{\mathbb{G}}_{n,\varphi}+\zeta_{n,j}\end{aligned}$$

where $\widetilde{\mathbb{G}}_n$ is defined in (A.80), $\bar{\varphi}$ 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 $E[m_\varphi(W)]$,

$$\zeta_{n,j}=\frac{\widetilde{\mathbb{G}}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}_n\right)-\widetilde{\mathbb{G}}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\varphi_{P_n}\right)}{\sigma_{P_n,j}\left(\beta_n\right)}.\qquad(A.92)$$

As for (A.91), 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,

$$\frac{\sqrt{n}\bar{m}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi}\right)}{\hat{\sigma}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)}=\bar{u}_{n,j,\beta_n}(\Delta) \tag{A.93}$$

where

$$\bar{u}_{n,j,\beta_n}(\Delta)=\left(1+\hat{\eta}_{n,j,\beta_n}(\Delta)\right)\left(\mathbb{Z}_{n,j,\beta_n}+\frac{\sigma_{P_n,j}\left(\bar{\beta}_n\right)}{\sigma_{P_n,j}\left(\beta_n\right)}D_{P_n,j}\left(\bar{\beta}_n\right)\Delta\rho+\sqrt{n}\gamma_{1,P_n,j}\left(\beta_n,\varphi_{P_n}\right)\right), \tag{A.94}$$

$$\mathbb{Z}_{n,j,\beta_n}(\Delta)=\frac{\widetilde{\mathbb{G}}_{n,j}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\varphi_{P_n}\right)+\nabla_{\varphi'}E_{P_n}\left[m_j\left(\bar{\beta}_n,\varphi_{P_n}\right)\right]\nabla_{m'_\varphi}\bar{\varphi}^{\dagger}\widetilde{\mathbb{G}}_{n,\varphi}}{\sigma_{P_n,j}\left(\beta_n\right)}+\zeta_{n,j}, \tag{A.95}$$

$$\hat{\eta}_{n,j,\beta_n}(\Delta)=\frac{\sigma_{P_n,j}\left(\beta_n\right)}{\hat{\sigma}_j\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)}-1. \tag{A.96}$$

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

$$\bar{u}_{n,k,\beta_n}(\Delta)=\left(1+\hat{\eta}_{n,k,\beta_n}(\Delta)\right)\left(\mathbb{Z}_{n,k,\beta_n}+\frac{\sigma_{P_n,k}\left(\bar{\beta}_n\right)}{\sigma_{P_n,k}\left(\beta_n\right)}D_{P_n,k}\left(\bar{\beta}_n\right)\Delta\rho+\sqrt{n}\gamma_{1,P_n,k}(\beta_n)\right) \tag{A.97}$$

with

$$\hat{\eta}_{n,k,\beta_n}(\Delta)=\frac{\sigma_{P_n,k}\left(\beta_n\right)}{\hat{\sigma}_k\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}}\right)}-1, \tag{A.98}$$

$$\mathbb{Z}_{n,k,\beta_n}=\frac{\widetilde{\mathbb{G}}_{n,k}\left(\beta_n+\frac{\Delta\rho}{\sqrt{n}},\varphi_{P_n}\right)+\nabla_{m'_k}\lambda_{P_n,k}\nabla_{\varphi'}E_{P_n}\left[m_k\left(\bar{\beta}_n,\varphi_{P_n}\right)\right]\nabla_{m'_\varphi}\bar{\varphi}^{\dagger}\widetilde{\mathbb{G}}_{n,\varphi}}{\sigma_{P_n,j}\left(\beta_n\right)}+\zeta_{n,k}, \tag{A.99}$$

$$\zeta_{n,k} = \frac{\widetilde{\mathbb{G}}_{n,k}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \hat{\varphi}_n\right) - \widetilde{\mathbb{G}}_{n,k}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \varphi_{P_n}\right)}{\sigma_{P_n,k}(\beta_n)}, \tag{A.100}$$

$$\gamma_{1,P_n,\ell}(\beta_n) = \frac{\lambda_\ell\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \varphi_P\right) - \theta_n}{\sigma_{P_n,\ell}(\beta_n)}, \quad \gamma_{1,P_n,u}(\beta_n) = \frac{\theta_n - \lambda_u\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \varphi_P\right)}{\sigma_{P_n,u}(\beta_n)}. \tag{A.101}$$

Therefore,

$$(A.89) \Leftrightarrow \left\{ \begin{array}{c} \inf_\Delta \bar{u}_{n,\ell,\beta_n}(\Delta) - \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right) \\ \text{s.t. } \Delta \in \frac{\sqrt{n}}{\rho}\left(\overline{\mathcal{B}} - \beta_n\right), \\ \bar{u}_{n,j,\beta_n}(\Delta) \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j = 1, ..., J \end{array} \right\} \leqslant 0,$$

$$\text{and } \left\{ \begin{array}{c} \inf_\Delta \bar{u}_{n,u,\beta_n}(\Delta) - \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right) \\ \text{s.t. } \Delta \in \frac{\sqrt{n}}{\rho}\left(\overline{\mathcal{B}} - \beta_n\right), \\ \bar{u}_{n,j,\beta_n}(\Delta) \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j = 1, ..., J \end{array} \right\} \leqslant 0$$

Denote

$$\mathbf{\Delta}_{n,\rho} = \frac{\sqrt{n}}{\rho}\left(\overline{\mathcal{B}} - \beta_n\right) \cap \mathbf{\Delta}, \tag{A.102}$$

with $\mathbf{\Delta} = \left\{x \in \mathbb{R}^d : |x_i| \leqslant 1, i = 1, \ldots, d\right\}$. Then the event in (A.89) is implied by

$$\left\{ \begin{array}{c} \inf_\Delta \bar{u}_{n,\ell,\beta_n}(\Delta) - \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right) \\ \text{s.t. } \Delta \in \mathbf{\Delta}_{n,\rho}, \\ \bar{u}_{n,j,\beta_n}(\Delta) \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j = 1, ..., J \end{array} \right\} \leqslant 0, \tag{A.103}$$

$$\left\{ \begin{array}{c} \inf_\Delta \bar{u}_{n,u,\beta_n}(\Delta) - \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right) \\ \text{s.t. } \Delta \in \mathbf{\Delta}_{n,\rho}, \\ \bar{u}_{n,j,\beta_n}(\Delta) \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j = 1, ..., J \end{array} \right\} \leqslant 0. \tag{A.104}$$

Step 3. This step is used only when Assumption 73 is invoked. When this assumption is invoked, recall we use modification in Section A.4.3. For each $j = 1, ..., R_1$ such that

$$\pi^*_{1,j} = \pi^*_{1,j+R_1} = 0, \tag{A.105}$$

let

$$\tilde{\mu}_j = 1 - \tilde{\mu}_{j+R_1}, \tag{A.106}$$

$$\tilde{\mu}_{j+R_1} = \begin{cases} 0 & \text{if } \gamma_{1,P_n,j}(\beta) = \gamma_{1,P_n,j+R_1}(\beta) = 0 \\ \frac{\hat{\tau}_{n,j,\beta_n}(\Delta)}{\hat{\tau}_{n,j,\beta_n}(\Delta)+\hat{\tau}_{n,j+R_1,\beta_n}(\Delta)} & \text{otherwise,} \end{cases} \tag{A.107}$$

with

$$\hat{\tau}_{n,j,\beta}(\Delta) = \gamma_{1,P_n,j}(\beta_n)\left(1+\hat{\eta}_{n,j,\beta_n}(\Delta)\right). \tag{A.108}$$

For each $j = 1, ..., R_1$, replace the constraint indexed by j, that is

$$\frac{\sqrt{n}\bar{m}_{n,j}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \hat{\varphi}\right)}{\hat{\sigma}_j^M\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)} \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \tag{A.109}$$

with the following weighted sum of the paired inequalities

$$\tilde{\mu}_j \frac{\sqrt{n}\bar{m}_{n,j}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \hat{\varphi}\right)}{\hat{\sigma}_j^M\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)} - \tilde{\mu}_{j+R_1} \frac{\sqrt{n}\bar{m}_{j+R_1,n}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \hat{\varphi}\right)}{\hat{\sigma}_{j+R_1}^M\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)} \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \tag{A.110}$$

and replace the constraint indexed by $j + R_1$, that is

$$\frac{\sqrt{n}\bar{m}_{n,j+R_1}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \hat{\varphi}\right)}{\hat{\sigma}_{j+R_1}^M\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)} \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \tag{A.111}$$

with

$$-\tilde{\mu}_j \frac{\sqrt{n}\bar{m}_{n,j}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \hat{\varphi}\right)}{\hat{\sigma}_j^M\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)} + \tilde{\mu}_{j+R_1} \frac{\sqrt{n}\bar{m}_{j+R_1,n}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \hat{\varphi}\right)}{\hat{\sigma}_{j+R_1}^M\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)} \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right). \tag{A.112}$$

It then follows from Assumption 73 that these replacements are conservative because

$$\frac{\bar{m}_{n,j+R_1}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \hat{\varphi}\right)}{\hat{\sigma}_{j+R_1}^M\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)} \leqslant -\frac{\bar{m}_{n,j}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}, \hat{\varphi}\right)}{\hat{\sigma}_j^M\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)},$$

and therefore (A.110) implies (A.109) and (A.112) implies (A.111). Similarly, if Assumption 73.2 is invoked, for $k = \ell, u$, replace

$$\bar{u}_{n,k,\beta_n}(\Delta) \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)$$

with

$$\tilde{\mu}_k \bar{u}_{n,k,\beta_n}(\Delta) - \tilde{\mu}_{[k]} \bar{u}_{n,[k],\beta_n}(\Delta) \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right),$$

where $\tilde{\mu}_u = 1 - \tilde{\mu}_\ell$ and

$$\tilde{\mu}_\ell = \begin{cases} 0 & \text{if } \gamma_{1,P_n,\ell}(\beta) = \gamma_{1,P_n,u}(\beta) = 0 \\ \frac{\hat{\tau}_{n,u,\beta_n}(\Delta)}{\hat{\tau}_{n,\ell,\beta_n}(\Delta)+\hat{\tau}_{n,u,\beta_n}(\Delta)} & \text{otherwise,} \end{cases} \tag{A.113}$$

with

$$\hat{\tau}_{n,k,\beta}(\Delta) = \gamma_{1,P_n,k}(\beta_n)\left(1 + \hat{\eta}_{n,k,\beta_n}(\Delta)\right). \tag{A.114}$$

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 $j = 1, ..., J_1$,

$$\pi^*_{1,j} = 0 \Rightarrow \pi^*_{1,j} \geqslant \sqrt{n}\gamma_{1,P_n,j}(\beta_n), \tag{A.115}$$

$$\pi^*_{1,j} = -\infty \Rightarrow \sqrt{n}\gamma_{1,P_n,j}(\beta_n) \to -\infty. \tag{A.116}$$

For any constraint j for which $\pi^*_{1,j} = 0$, (A.115) yields that replacing $\sqrt{n}\gamma_{1,P_n,j}(\beta_n)$ in (A.104) with $\pi^*_{1,j}$ introduces a conservative distortion. Under Assumption 73, for any j such that (A.105) holds, the substitutions in (A.110) and (A.112) 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 $\pi^*_{1,j} = -\infty$, (A.116) 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 26, Assumption 72.3, Assumption 74.3 and $\Delta \in \mathbf{\Delta}^d_{n,\rho}$, it follows that

$$\frac{\sigma_{F_n,j}\left(\bar{\beta}_n\right)}{\sigma_{F_n,j}(\beta_n)} D_{F_n,j}\left(\bar{\beta}_n\right)\Delta\rho = O(1).$$

Together with Lemma 83.1 and Lemma 84, it holds that

$$P_n\left(\max_{j:\pi^*_{1,j}=-\infty} \bar{u}_{n,j,\beta_n}(\Delta) \leqslant 0, \quad \forall \Delta \in \mathbf{\Delta}^d_{n,\rho}\right) \to 1. \tag{A.117}$$

Since $\hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right) \geqslant 0$ by construction, $\bar{u}_{n,j,\beta_n}(\Delta) \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)$ with $\pi^*_{1,j} = -\infty$ is asmptotically not bingding and thus negligible. We therefore have that for $n \geqslant N$,

$$P_n\left(\left\{\begin{array}{c}\inf_\Delta \bar{u}_{n,\ell,\beta_n}(\Delta) \\ \text{s.t. } \Delta \in \mathbf{\Delta}^d_{n,\rho}, \\ \bar{u}_{n,j,\beta_n}(\Delta) \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j\end{array}\right\} \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \text{ and} \tag{A.118}$$

$$\left\{\begin{array}{c}\inf_\Delta \bar{u}_{n,u,\beta_n}(\Delta) \\ \text{s.t. } \Delta \in \mathbf{\Delta}^d_{n,\rho}, \\ \bar{u}_{n,j,\beta_n}(\Delta) \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j\end{array}\right\} \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)\Bigg) + o(1)$$

$$\geqslant P_n\left(\left\{\begin{array}{c}\inf_\Delta \bar{u}_{n,\ell,\beta_n}(\Delta) \\ \text{s.t. } \Delta \in \mathbf{\Delta}^d_{n,\rho}, \\ u_{n,j,\beta_n}(\Delta) \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j\end{array}\right\} \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \text{ and} \tag{A.119}$$

$$\left\{\begin{array}{c}\inf_\Delta \bar{u}_{n,u,\beta_n}(\Delta) \\ \text{s.t. } \Delta \in \mathbf{\Delta}^d_{n,\rho}, \\ u_{n,j,\beta_n}(\Delta) \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j\end{array}\right\} \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)\Bigg).$$

where

$$u_{n,j,\beta_n}(\Delta) = (1 + \hat{\eta}_{n,j,\beta_n}(\Delta))\left(\mathbb{Z}_{n,j,\beta_n}(\Delta) + \frac{\sigma_{P_n,j}\left(\bar{\beta}_n\right)}{\sigma_{P_n,j}\left(\beta_n\right)} D_{P_n,j}\left(\bar{\beta}_n\right)\Delta\rho + \pi^*_{1,j}\right). \tag{A.120}$$

Hence, I focus on the event in (A.119) 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 $\pi^*_{1,0} = 0$,

$$(A.119) \geqslant P_n\left(\left\{\begin{array}{c}\inf_\Delta u_{n,\ell,\beta_n}(\Delta) \\ \text{s.t. } \Delta \in \mathbf{\Delta}^d_{n,\rho}, \\ u_{n,j,\beta_n}(\Delta) \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j\end{array}\right\} \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \text{ and}$$

$$\left\{\begin{array}{c}\inf_\Delta u_{n,u,\beta_n}(\Delta) \\ \text{s.t. } \Delta \in \mathbf{\Delta}^d_{n,\rho}, \\ u_{n,j,\beta_n}(\Delta) \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j\end{array}\right\} \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)\Bigg)$$

where

$$u_{n,k,\beta_n}(\Delta) = (1 + \hat{\eta}_{n,k,\beta_n}(\Delta))\left(\mathbb{Z}_{n,k,\beta_n}(\Delta) + \frac{\sigma_{P_n,k}\left(\bar{\beta}_n\right)}{\sigma_{P_n,k}\left(\beta_n\right)} D_{P_n,k}\left(\bar{\beta}_n\right)\Delta\rho\right). \tag{A.121}$$

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

$$(A.119) \geqslant \min \left\{ P_n \left(\left\{ \begin{array}{c} \inf_\Delta u_{n,\ell,\beta_n}(\Delta) \\ \text{s.t. } \Delta \in \mathbf{\Delta}^d_{n,\rho}, \\ u_{n,j,\beta_n}(\Delta) \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j \end{array} \right\} \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right) \right), \right.$$

$$\left. P_n \left(\left\{ \begin{array}{c} \inf_\Delta u_{n,u,\beta_n}(\Delta) \\ \text{s.t. } \Delta \in \mathbf{\Delta}^d_{n,\rho}, \\ u_{n,j,\beta_n}(\Delta) \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right), \forall j \end{array} \right\} \leqslant \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right) \right) \right\} + o(1).$$

In sum, let

$$U_{n,\ell}(\beta_n, c) = \Big\{ \Delta \in \mathbf{\Delta}^d_{n,\rho} :$$

$$u_{n,\ell,\beta_n}(\Delta) \leqslant c, u_{n,u,\beta_n}(\Delta) + \pi^*_{1,0} \leqslant c, u_{n,j,\beta_n}(\Delta) \leqslant c, \forall j = 1, \ldots, J \Big\}, \tag{A.122}$$

$$U_{n,u}(\beta_n, c) = \Big\{ \Delta \in \mathbf{\Delta}^d_{n,\rho} :$$

$$u_{n,u,\beta_n}(\Delta) \leqslant c, u_{n,\ell,\beta_n}(\Delta) + \pi^*_{1,0} \leqslant c, u_{n,j,\beta_n}(\Delta) \leqslant c, \forall j = 1, \ldots, J \Big\}. \tag{A.123}$$

It holds that

$$(A.119) \geqslant \min \left\{ P_n \left(U_{n,\ell}\left(\beta_n, \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)\right) \neq \varnothing \right), \right.$$

$$\left. P_n \left(U_{n,u}\left(\beta_n, \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)\right) \neq \varnothing \right) \right\}.$$

Step 6. Simplify $\hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right)$. By definition $\hat{c}(\cdot) \geqslant 0$ and thus $\hat{c}_{n,\rho}$ defined by

$$\hat{c}_{n,\rho} = \inf_{\Delta \in \mathbf{\Delta}^d_{n,\rho}} \hat{c}\left(\beta_n + \frac{\Delta\rho}{\sqrt{n}}\right) \tag{A.124}$$

exists. Therefore, the event whose probability is evaluated in (A.119) is implied by the event

$$\left\{ \begin{array}{c} \inf_\Delta u_{n,\ell,\beta_n}(\Delta) \\ \text{s.t. } \Delta \in \mathbf{\Delta}^d_{n,\rho}, \\ u_{n,j,\beta_n}(\Delta) \leqslant \hat{c}_{n,\rho}, \forall j = 1, ..., J \end{array} \right\} \leqslant \hat{c}_{n,\rho},$$

$$\text{and } \left\{ \begin{array}{c} \inf_\Delta u_{n,u,\beta_n}(\Delta) \\ \text{s.t. } \Delta \in \mathbf{\Delta}^d_{n,\rho}, \\ u_{n,j,\beta_n}(\Delta) \leqslant \hat{c}_{n,\rho}, \forall j = 1, ..., J \end{array} \right\} \leqslant \hat{c}_{n,\rho} \tag{A.125}$$

Then by (A.125) and the definition of $U_{n,\ell}$ and $U_{n,u}$, we obtain

$$P_n\left(\theta_n \in CI_n\right) \geqslant \min\left\{P_n\left(U_{n,\ell}\left(\beta_n, \hat{c}_{n,\rho}\right) \neq \varnothing\right), P_n\left(U_{n,u}\left(\beta_n, \hat{c}_{n,\rho}\right) \neq \varnothing\right)\right\}. \tag{A.126}$$

By passing to a further subsequence, we may assume that

$$D_{P_n}\left(\beta_n\right) \to D, G_{P_n}\left(\beta_n\right) \to G,$$

for some $(J+2) \times d$ matrix D such that $\|D\| \leqslant M$, and some $(J+2) \times K$ matrix G such that $\|G\| \leqslant M$. 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 75,

$$\liminf_{n\to\infty} \min\left\{P_n\left(U_{n,\ell}\left(\beta_n, \hat{c}_{n,\rho}\right) \neq \varnothing\right), P_n\left(U_{n,u}\left(\beta_n, \hat{c}_{n,\rho}\right) \neq \varnothing\right)\right\} \geqslant 1-\alpha. \tag{A.127}$$

The conclusion of the theorem then follows from (A.85), (A.86), (A.126), and (A.127). □

A.4.5 Lemmas

In the proof, I focus on

$$\liminf_{n\to\infty} P_n\left(U_{n,\ell}\left(\beta_n, \hat{c}_{n,\rho}\right) \neq \varnothing\right) \geqslant 1-\alpha,$$

and the proof for $P_n\left(U_{n,u}\left(\beta_n, \hat{c}_{n,\rho}\right) \neq \varnothing\right)$ 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 29. 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

$$\kappa_n^{-1}\sqrt{n}\gamma_{1,P_n,j}\left(\beta_n\right) \to \pi_{1j} \in \mathbb{R}_{[-\infty]}, j = 0, \ldots, J_1 \tag{A.128}$$

$$\Omega_{P_n} \xrightarrow{u} \overline{\Omega},$$

$$D_{P_n}\left(\beta_n\right) \to D, \tag{A.129}$$

$$G_{P_n}\left(\beta_n\right) \to G. \tag{A.130}$$

When Assumption 73 is invoked, I use modification in Section A.4.3. And if

$$\pi^*_{1,j} = 0 = \pi^*_{1,j+R_1},$$

replace the constraints

$$\mathbb{Z}_j + \rho D_j \Delta \leqslant c,$$

$$\mathbb{Z}_{j+R_1} + \rho D_{j+R_1} \Delta \leqslant c,$$

with

$$\mu_j(\beta) \left\{ \mathbb{Z}_j + \rho D_j \Delta \right\} - \mu_{j+R_1}(\beta) \left\{ \mathbb{Z}_{j+R_1} + \rho D_{j+R_1} \Delta \right\} \leqslant c$$

$$-\mu_j(\beta) \left\{ \mathbb{Z}_j + \rho D_j \Delta \right\} + \mu_{j+R_1}(\beta) \left\{ \mathbb{Z}_{j+R_1} + \rho D_{j+R_1} \Delta \right\} \leqslant c$$

where

$$\mu_j(\beta) = \begin{cases} 1 & \text{if } \gamma_{1,P_n,j}(\beta, \varphi_{P_n}) = 0 = \gamma_{1,P_n,j+R_1}(\beta, \varphi_{P_n}), \\ \frac{\gamma_{1,P_n,j+R_1}(\beta,\varphi_{P_n})}{\gamma_{1,P_n,j+R_1}(\beta,\varphi_{P_n}) + \gamma_{1,P_n,j}(\beta,\varphi_{P_n})} & \text{otherwise,} \end{cases} \tag{A.131}$$

$$\mu_{j+R_1}(\beta) = \begin{cases} 0 & \text{if } \gamma_{1,P_n,j}(\beta) = 0 = \gamma_{1,P_n,j+R_1}(\beta), \\ \frac{\gamma_{1,P_n,j}(\beta,\varphi_{P_n})}{\gamma_{1,P_n,j+R_1}(\beta,\varphi_{P_n}) + \gamma_{1,P_n,j}(\beta,\varphi_{P_n})} & \text{otherwise,} \end{cases} . \tag{A.132}$$

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 (A.110) and (A.112), 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 28,

$$\mathbb{J}_1 = \{1, ..., J_1, \ell, u\},$$

$$\mathbb{J}_2 = \{J_1 + 1, ..., J_1 + 2J_2\}.$$

Under Assumption 73.1,

$$\mathbb{J}_1 = \{2R_1 + 1, ..., J_1, \ell, u\},$$

$$\mathbb{J}_2 = \{1, ..., R_1, J_1 + 1, ..., J_1 + J_2\}.$$

Under Assumption 73.2,

$$\mathbb{J}_1 = \{2R_1 + 1, ..., J_1\},$$

$$\mathbb{J}_2 = \{1, ..., R_1, J_1 + 1, ..., J_1 + J_2, \ell\}.$$

In all three cases,

$$[\mathbb{J}_2] = \{[j] : j \in \mathbb{J}_2\},$$

$$\mathbb{J} = \mathbb{J}_1 \cup \mathbb{J}_2 \cup [\mathbb{J}_2].$$

Fix $c \geqslant 0$. For each $\Delta \in \mathbb{R}^d$ and $\beta \in (\beta_n + \rho/\sqrt{n}\boldsymbol{\Delta}) \cap \mathcal{B}$, let

$$\mathfrak{w}_j(\Delta) = \mathbb{Z}_j + \rho D_j \Delta + \pi^*_{1,j}, \tag{A.133}$$

$$\mathfrak{w}_\ell(\Delta) = \mathbb{Z}_\ell + \rho D_\ell \Delta,$$

$$\mathfrak{w}_u(\Delta) = \mathbb{Z}_u + \rho D_u \Delta + \pi^*_{1,0}.$$

Let $\boldsymbol{\Delta}^d_{\infty,\rho} = \lim_{n\to\infty} \boldsymbol{\Delta}^d_{n,\rho}$. Let

$$\mathfrak{W}(c) = \left\{\Delta \in \boldsymbol{\Delta}^d_{\infty,\rho} : \mathfrak{w}_j(\Delta) \leqslant c, \forall j \in \mathbb{J}\right\}. \tag{A.134}$$

With that convention, for given $\delta \in \mathbb{R}$, define

$$\mathfrak{W}^\delta(c) = \left\{\Delta \in \boldsymbol{\Delta}^d_{\infty,\rho} : \mathfrak{w}_j(\Delta) \leqslant c + \delta, \forall j \in \mathbb{J}_1, \mathfrak{w}_j(\Delta) \leqslant c, \forall j \in \mathbb{J}_2 \cup \left[\bar{\mathbb{J}}_2\right]\right\}. \tag{A.135}$$

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 75. *Suppose Assumptions 26, 27, 25, 72 hold. In addition, suppose Assumption 28 or 73 hold. Then,*

$$\liminf_{n\to\infty} P_n\left(U_{\ell,n}\left(\beta_n, \hat{c}_{n,\rho}\right) \neq \varnothing\right) \geqslant 1 - \alpha. \tag{A.136}$$

Proof. Consider a subsequence along which

$$\liminf_{n\to\infty} P_n\left(U_{\ell,n}\left(\beta_n, \hat{c}_{n,\rho}\right) \neq \varnothing\right)$$

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

$$\hat{c}_n\left(\beta_n'\right) \geqslant c_n^I\left(\beta_n'\right) + o_p(1) \tag{A.137}$$

and $c_n^I\left(\beta_n'\right) \xrightarrow{p} c_{\pi *}$ for any $\beta_n' \in \left(\beta_n + \rho/\sqrt{n}\mathbf{\Delta}\right) \cap \overline{\mathcal{B}}$, where

$$c_{\pi *} = \inf\left\{c \in \mathbb{R}_+ : P\left(\mathfrak{W}(c) \neq \varnothing\right) \geqslant 1-\alpha\right\}. \tag{A.138}$$

The construction is as follows. When $c_{\pi *} = 0$, let $c_n^I\left(\beta_n'\right) = 0$ for all $\beta_n' \in \left(\beta_n + \rho/\sqrt{n}\mathbf{\Delta}\right) \cap \overline{\mathcal{B}}$, and hence $c_n^I\left(\beta_n'\right) \xrightarrow{p} c_{\pi *}$. If $c_{\pi *} > 0$, let

$$c_n^I\left(\beta_n\right) = \inf\left\{c \in \mathbb{R}_+ : P^s\left(V_n^I\left(\beta_n, c\right) \neq \varnothing\right) \geqslant 1-\alpha\right\}, \tag{A.139}$$

where

$$V_n^I\left(\beta_n', c\right) = \left\{\Delta \in \mathbf{\Delta}_{n,\rho}^d : v_{n,j,\beta_n'}^I(\Delta) \leqslant c, j \in \mathbb{J}\right\}, \tag{A.140}$$

$$v_{n,j,\beta_n'}^I(\Delta) = \mathbb{Z}_{n,j}^s\left(\beta_n'\right) + \rho\hat{D}_{n,j}\left(\beta_n'\right)\Delta + \phi_j^*\left(\hat{\xi}_{n,j}\left(\beta_n'\right)\right), \tag{A.141}$$

$$v_{n,\ell,\beta_n'}^I(\Delta) = \mathbb{Z}_{n,\ell}^s\left(\beta_n\right) + \rho\hat{D}_{n,\ell}\left(\beta_n'\right)\Delta$$

$$v_{n,u,\beta_n'}^I(\Delta) = \mathbb{Z}_{n,u}^s\left(\beta_n\right) + \rho\hat{D}_{n,u}\left(\beta_n'\right)\Delta + \phi_0^*\left(\hat{\xi}_{n,0}\left(\beta_n'\right)\right)$$

$$\phi_j^*(\xi) = \begin{cases} \xi & \pi_{1,j} = 0 \\ -\infty & \pi_{1,j} < 0 \end{cases}, \qquad \phi_0^*(\xi) = \begin{cases} \xi & \pi_{1,0} = 0 \\ -\infty & \pi_{1,0} < 0 \end{cases}. \tag{A.142}$$

and P^s is from the conditional distribution of $\mathbb{Z}_n^s$ conditional on the estimators. By Lemma 77.3, this critical value sequence satisfies (A.137). Further, by Lemma 77.2, $c_n^I\left(\beta_n'\right) \xrightarrow{p} c_{\pi *}$ for any $\beta_n' \in \left(\beta_n + \rho/\sqrt{n}\mathbf{\Delta}\right) \cap \overline{\mathcal{B}}$. For each $\beta \in \overline{\mathcal{B}}$, let

$$c_{n,\rho}^I(\beta) = \inf_{\Delta \in \mathbf{\Delta}_{n,\rho}^d} c_n^I\left(\beta + \frac{\Delta\rho}{\sqrt{n}}\right). \tag{A.143}$$

Since the $o_p(1)$ term in (A.137) does not affect the argument below, I redefine $c^I_{n,\rho}(\beta_n)$ as $c^I_{n,\rho}(\beta_n) + o_p(1)$. By (A.137) and simple addition and subtraction,

$$\begin{aligned}
&P_n\left(U_{n,\ell}\left(\beta_n, \hat{c}_{n,\rho}(\beta_n)\right) \neq \varnothing\right) \\
\geqslant &P_n\left(U_{n,\ell}\left(\beta_n, c^I_{n,\rho}(\beta_n)\right) \neq \varnothing\right) \\
= &P\left(\mathfrak{W}\left(c_{\pi *}\right) \neq \varnothing\right) + \left[P_n\left(U_{n,\ell}\left(\beta_n, c^I_{n,\rho}(\beta_n)\right) \neq \varnothing\right) - P\left(\mathfrak{W}\left(c_{\pi *}\right) \neq \varnothing\right)\right].
\end{aligned} \tag{A.144}$$

By Lemma 84,

$$\mathbb{Z}_{n,j,\beta_n}(\Delta) \xrightarrow{d} \mathbb{Z} \sim \mathcal{N}(0, \Sigma).$$

And by Assumption 28 and Assumption 74.3,

$$\sup_{\beta \in \mathcal{B}} \sup_{\Delta \in \boldsymbol{\Delta}} \frac{\sigma_{P_n,k}\left(\bar{\beta}_n\right)}{\sigma_{P_n,k}\left(\beta_n\right)} \to 1$$

Moreover, by Lemma 83,

$$\sup_{\beta \in \mathcal{B}} \sup_{\Delta \in \boldsymbol{\Delta}} \|\hat{\eta}_{n,j,\beta}(\Delta)\| \xrightarrow{p} 0$$

uniformly in $\mathcal{P}$, and by Lemma 77, $c^I_{n,\rho}(\beta_n) \xrightarrow{p} c_{\pi *}$. Therefore, uniformly in $\Delta \in \boldsymbol{\Delta}$, it holds that

$$\left(\mathbb{Z}_{n,\beta_n}(\Delta),\ \left\{\hat{\eta}_{n,\beta}(\Delta),\ c^I_{n,\rho}(\beta_n)\right) \xrightarrow{d} (\mathbb{Z},\ 0,\ c_{\pi *}\right). \tag{A.145}$$

In what follows, using Lemma 1.10.4 in Van Der Vaart and Wellner (1996) I take

$$\left(\mathbb{Z}^*_n(\Delta), \eta^*_n(\Delta), \left\{D^*_{n,j}\right\}_{j \leqslant J}, c^*_n\right)$$

to be the almost sure representation of the left hand side of (A.145), 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

$$u^*_{n,j,\beta_n}(\Delta) = (1 + \eta^*_n(\Delta)) \left\{\mathbb{Z}^*_{n,j}(\Delta) + \rho \frac{\sigma_{P_n,j}\left(\bar{\beta}_n\right)}{\sigma_{P_n,j}\left(\beta_n\right)} D_{P_n,j}\left(\bar{\beta}_n\right) \Delta + \pi^*_{1,j}\right\},$$

$$\mathfrak{w}^*_j(\Delta) = \mathbb{Z}^*_j + \rho D_j \Delta + \pi^*_{1,j},$$

where we used that by Lemma 85.1, $\kappa_n^{-1}\sqrt{n}\gamma_{1,P,j}(\beta_n) - \kappa_n^{-1}\sqrt{n}\gamma_{1,P,j}(\beta_n') = o(1)$ 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

$$U_n^*(\beta_n, c_n^*) = \left\{\Delta \in \mathbf{\Delta}_{n,\rho}^d : u_{n,j,\beta_n}^*(\Delta) \leqslant c_n^*,\ \forall j \in \mathbb{J}\right\}, \tag{A.146}$$

$$\mathfrak{W}^*(c_{\pi*}) = \left\{\Delta \in \mathbf{\Delta}_{\infty,\rho}^d : \mathfrak{w}_j^*(\Delta) \leqslant c_{\pi*},\ \forall j \in \mathbb{J}\right\}. \tag{A.147}$$

It then follows that equation (A.144) can be rewritten as

$$\begin{aligned}
&P_n\left(\{U_n(\beta_n, \hat{c}_{n,\rho}(\beta_n)) \neq \varnothing\}\right) \\
\geqslant &P\left(\mathfrak{W}^*(c_{\pi*}) \neq \varnothing\right) + \left[P_n\left(U_n^*(\beta_n, c_n^*) \neq \varnothing\right) - P\left(\mathfrak{W}^*(c_{\pi*}) \neq \varnothing\right)\right].
\end{aligned} \tag{A.148}$$

By the definition of $c_{\pi*}$, we have

$$P\left(\mathfrak{W}^*(c_{\pi*}) \neq \varnothing\right) \geqslant 1 - \alpha.$$

Therefore, we are left to show that the second term on the right hand side of (A.148) tends to 0 as $n \to \infty$. Note that

$$\begin{aligned}
&\left|P\left(U_n^*(\beta_n, c_n^*) \neq \varnothing\right) - P\left(\mathfrak{W}^*(c_{\pi*}) \neq \varnothing\right)\right| \\
\leqslant &P_n\left(\{U_n^*(\beta_n, c_n^*) = \varnothing\} \cap \{\mathfrak{W}^*(c_{\pi*}) \neq \varnothing\}\right) + P_n\left(\{U_n^*(\beta_n, c_n^*) \neq \varnothing\} \cap \{\mathfrak{W}^*(c_{\pi*}) = \varnothing\}\right).
\end{aligned} \tag{A.149}$$

The conclusion holds because (A.149) converges to zero by Lemma 76. □

Lemma 76. *Suppose Assumptions 26, 27, 25, 72 hold. In addition, suppose Assumption 28 or 73 hold. Let (P_n, β_n, θ_n) have the almost sure representations given in Lemma 75. For any $\eta > 0$, there exists $N \in \mathbb{N}$ such that*

$$P_n\left(\{U_n^*(\beta_n, c_n^*) \neq \varnothing\} \cap \{\mathfrak{W}^*(c_{\pi*}) = \varnothing\}\right) \leqslant \eta/2 \tag{A.150}$$

$$P_n\left(\{U_n^*(\beta_n, c_n^*) = \varnothing\} \cap \{\mathfrak{W}^*(c_{\pi*}) \neq \varnothing\}\right) \leqslant \eta/2, \tag{A.151}$$

for all $n \geqslant N$, where the sets in the above expressions are defined in equations (A.146) and (A.147).

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

$$u^*_{n,j,\beta_n}(\Delta) = (1 + \eta^*_n(\Delta)) \left\{ \mathbb{Z}^*_{n,j}(\Delta) + \rho \frac{\sigma_{P_n,j}\left(\bar{\beta}_n\right)}{\sigma_{P_n,j}\left(\beta_n\right)} D_{P_n,j}\left(\bar{\beta}_n\right) \Delta + \pi^*_{1,j} \right\} \leqslant c^*_n$$

$$\mathfrak{w}^*_j(\Delta) = \mathbb{Z}^*_j + \rho D_j \Delta + \pi^*_{1,j} \leqslant c_{\pi *}$$

are satisfied with probability approaching one by similar arguments as in (A.117). Hence, we can redefine the sets of interest as

$$U^*_n\left(\beta_n, c^*_n\right) = \left\{ \Delta \in \mathbf{\Delta}^d_{n,\rho} : u^*_{n,j,\beta_n}(\Delta) \leqslant c^*_n, \forall j \in \mathbb{J}^* \right\} \tag{A.152}$$

$$\mathfrak{W}^*\left(c_{\pi *}\right) = \left\{ \Delta \in \mathbf{\Delta}^d_{\infty,\rho} : \mathfrak{w}^*_j(\Delta) \leqslant c_{\pi *}, \forall j \in \mathbb{J}^* \right\}. \tag{A.153}$$

I first show (A.150). I bound the left hand side of (A.150) as

$$\begin{aligned} & P\left(\left\{U^*_n\left(\beta_n, c^*_n\right) \neq \varnothing\right\} \cap \left\{\mathfrak{W}^*\left(c_{\pi *}\right) = \varnothing\right\}\right) \\ \leqslant & P\left(\left\{U^*_n\left(\beta_n, c^*_n\right) \neq \varnothing\right\} \cap \left\{\mathfrak{W}^{*,+\delta}\left(c_{\pi *}\right) = \varnothing\right\}\right) && \text{(A.154)} \\ & + P\left(\left\{\mathfrak{W}^{*,+\delta}\left(c_{\pi *}\right) \neq \varnothing\right\} \cap \left\{\mathfrak{W}^*\left(c_{\pi *}\right) = \varnothing\right\}\right), && \text{(A.155)} \end{aligned}$$

following from $P(A \cap B) \leqslant P(A \cap C) + P\left(B \cap C^c\right)$ for any events A, B, and C. I will then specify δ such that (A.154)$\leqslant \frac{\eta}{2}$ and (A.155)$\leqslant \frac{\eta}{2}$. Lemma 79 (A.197) implies that there is some $\delta > 0$ such that

$$\begin{aligned} & P\left(\left\{\mathfrak{W}^{*,+\delta}\left(c_{\pi *}\right) \neq \varnothing\right\} \cap \left\{\mathfrak{W}^*\left(c_{\pi *}\right) = \varnothing\right\}\right) \\ \leqslant & \sup_{c \geqslant 0} P\left(\left\{\mathfrak{W}^*\left(c\right) \neq \varnothing\right\} \cap \left\{\mathfrak{W}^{*,-\delta}\left(c\right) = \varnothing\right\}\right) \leqslant \eta/2. && \text{(A.156)} \end{aligned}$$

Next, define the events

$$A_n(\delta) = \left\{ \sup_{\Delta \in \mathbf{\Delta}} \max_{j \in \mathbb{J}^*} \left| \left(u^*_{n,j,\beta_n}(\Delta) - c^*_n\right) - \left(\mathfrak{w}^*_j(\Delta) - c_{\pi *}\right) \right| \geqslant \delta \right\},$$

with δ specified in (A.156). By Lemma 78 (A.195), for any $\eta > 0$ there exists $N \in \mathbb{N}$ such that

$$P\left(A_n\right) < \eta/2, \forall n \geqslant N. \tag{A.157}$$

Define the event $L_n = \{U_n^*(\beta_n, c_n^*) \subseteq \mathfrak{W}^*(c_{\pi*} + \delta)\}$ and note that $A_n^c \subseteq L_n$. The right hand side of (A.154) can further be bounded as

$$P(\{U_n^*(\beta_n, c_n^*) \neq \varnothing\} \cap \{\mathfrak{W}^*(c_{\pi*} + \delta) = \varnothing\}) \leqslant P(U_n^*(\beta_n, c_n^*) \nsubseteq \mathfrak{W}^*(c_{\pi*} + \delta))$$
$$= P(L_n^c) \leqslant P(A_n) < \eta/2 \qquad \forall n \geqslant N, \tag{A.158}$$

where the penultimate inequality follows from $A_n^c \subseteq L_n$ as argued above, and the last inequality follows from (A.157). Hence, (A.150) follows from (A.154), (A.155), (A.156), and (A.158).

To establish (A.151), I distinguish three cases.

Case 1. Suppose first that $\mathbb{J}_2 = \varnothing$, and hence one has only moment inequalities. Define the event

$$\widetilde{R}_{2n} = \{\mathfrak{W}^*(c_{\pi*} - \delta) \subseteq U_n^*(\beta_n, c_n^*)\}. \tag{A.159}$$

and note that $A_n^c \subseteq \widetilde{R}_{2n}$. The result in equation (A.151) then follows by Lemma 79 (A.197) using again similar steps to (A.154)-(A.158).

Case 2. Next suppose that $|\mathbb{J}_2| > d$. Note that, by Lemma 81, there exists $N \in \mathbb{N}$ such that for all $n \geqslant N$, $c_n^I(\theta)$ 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 $c_{\pi*}$ is bounded from below by $\underline{c} > 0$. Note that $A_n^c \subseteq \widetilde{R}_{2n}$, where $\widetilde{R}_{2n}$ is defined as in (A.159), and therefore the same argument as in the previous case applies using Lemma 79(A.199).

Case 3. Finally, suppose that $1 \leqslant |\mathbb{J}_2| \leqslant d$. Recall that, with probability 1,

$$c_{\pi*} = \lim_{n\to\infty} c_n^*, \tag{A.160}$$

and note that by construction $c_{\pi*} \geqslant 0$. 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

$$P(\{U_n^*(\beta_n, c_n^*) = \varnothing\} \cap \{\mathfrak{W}^*(c_{\pi*}) \neq \varnothing\}),$$
$$\leqslant P\left(\{U_n^*(\beta_n, c_n^*) = \varnothing\} \cap \left\{\mathfrak{W}^{*,-\delta}(0) \neq \varnothing\right\}\right) + P\left(\left\{\mathfrak{W}^{*,-\delta}(0) = \varnothing\right\} \cap \{\mathfrak{W}^*(0) \neq \varnothing\}\right). \tag{A.161}$$

By Lemma 79, for any $\eta > 0$ there exists $\delta > 0$ and $N \in \mathbb{N}$ such that

$$P\left(\left\{\mathfrak{W}^{*,-\delta}(0) = \varnothing\right\} \cap \left\{\mathfrak{W}^{*}(0) \neq \varnothing\right\}\right) < \eta/3$$

for all $n \geqslant N$. Therefore, the second term of (A.161) can be made arbitrarily small.

I now consider the first term of (A.161), and it contains the following four steps.

Step 1. More notation. Let g be a $|\mathbb{J}| + 2d$ vector with

$$g_j = \begin{cases} -\mathbb{Z}_j, & j \in \mathbb{J}, \\ 1, & j = |\mathbb{J}| + 1, \ldots, |\mathbb{J}| + 2d, \end{cases} \tag{A.162}$$

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 (A.152), (A.153). 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

$$K = \begin{bmatrix} [\rho D_j]_{j \in \mathbb{J}_1 \cup \mathbb{J}_2} \\ [-\rho D_j]_{j \in \mathbb{J}_2} \\ I_d \\ -I_d \end{bmatrix}. \tag{A.163}$$

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 $\mathfrak{W}^{*,-\delta}(0) \neq \varnothing$, 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, $\mathfrak{W}^{*,-\delta}(0) \neq \varnothing$ 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 $\mathfrak{W}^{*,-\delta}(0) \neq \varnothing$ and equals $\mathbf{\Delta}$ otherwise, see, e.g., Molchanov and Molchanov (2005) (Definition 1.2.2).

Let g_n be a $|\mathbb{J}| + 2d$ vector with

$$g_{n,j}(\Delta) = \begin{cases} c_n^* / \left(1 + \eta_{n,j}^*(\Delta)\right) - \mathbb{Z}_{n,j}^*(\Delta) & \text{if } j \in \mathbb{J} \\ 1, & \text{if } j = |\mathbb{J}| + 1, \ldots, |\mathbb{J}| + 2d \end{cases} \tag{A.164}$$

using again that $\pi_{1,j}^* = 0$ for $j \in \mathbb{J}^*$. For each $P \in \mathcal{P}$, let

$$K_P(\beta, \bar{\beta}, \rho) = \begin{bmatrix} \left[\rho \frac{\sigma_{P_n,j}(\bar{\beta})}{\sigma_{P_n,j}(\beta)} D_{P,j}(\bar{\beta})\right]_{j \in \mathbb{J}_1 \cup \mathbb{J}_2} \\ \left[-\rho \frac{\sigma_{P_n,j}(\bar{\beta})}{\sigma_{P_n,j}(\beta)} D_{P,j}(\bar{\beta})\right]_{j \in \mathbb{J}_2} \\ I_d \\ -I_d \end{bmatrix}. \tag{A.165}$$

For each n and $\Delta \in \mathbf{\Delta}$, define the mapping $\phi_n : \mathbf{\Delta} \to \mathbb{R}_{[\pm\infty]}^d$ by

$$\phi_n(\Delta) = \left(K_{P_n}^{\mathcal{C}}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta\right), \rho\right)\right)^{-1} g_n^{\mathcal{C}}\left(\Delta\right), \tag{A.166}$$

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 $\Delta, \Delta' \in \mathbf{\Delta}$ write

$$\begin{aligned} &\left\|\phi_n(\Delta) - \phi_n\left(\Delta'\right)\right\| \\ =&\left\|\left(K_{P_n}^{\mathcal{C}}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta\right), \rho\right)\right)^{-1} g_n^{\mathcal{C}}\left(\Delta\right) - \left(K_{P_n}^{\mathcal{C}}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta'\right), \rho\right)\right)^{-1} g_n^{\mathcal{C}}\left(\Delta'\right)\right\| \\ \leqslant&\left\|\left(K_{P_n}^{\mathcal{C}}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta\right), \rho\right)\right)^{-1}\right\|_2 \left\|g_n^{\mathcal{C}}\left(\Delta\right) - g_n^{\mathcal{C}}\left(\Delta'\right)\right\| \\ &+\left\|\left(K_{P_n}^{\mathcal{C}}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta\right), \rho\right)\right)^{-1} - \left(K_{P_n}^{\mathcal{C}}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta'\right), \rho\right)\right)^{-1}\right\|_2 \left\|g_n^{\mathcal{C}}\left(\Delta'\right)\right\|, \end{aligned} \tag{A.167}$$

where $\|\cdot\|_2$ denotes the spectral norm (induced by the Euclidean norm). By Assumption 74.2, for any $\eta > 0, k > 0$, there is $N \in \mathbb{N}$ such that

$$\begin{aligned} &P_n\left(\left\|g_n^{\mathcal{C}}(\Delta) - g_n^{\mathcal{C}}(\Delta')\right\| \leqslant k\left\|\Delta - \Delta'\right\|\right) \\ =&P_n\left(\left\|\mathbb{Z}_n^{*,\mathcal{C}}(\Delta) - \mathbb{Z}_n^{*,\mathcal{C}}(\Delta')\right\| \leqslant k\left\|\Delta - \Delta'\right\|\right) \geqslant 1 - \eta, \forall n \geqslant N. \end{aligned}$$

Moreover, for any η there exist $0 < L < \infty$ and $N \in \mathbb{N}$ such that $\forall n \geqslant N$

$$\mathrm{P}\left(\sup_{\Delta' \in \boldsymbol{\Delta}} \left\| g_n^{\mathcal{C}}\left(\Delta'\right)\right\| \leqslant L\right) \geqslant 1 - \eta. \tag{A.168}$$

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 80 and the definition of $\mathcal{C}$, for any $\eta > 0$, there exists $0 < L < \infty$ and $N \in \mathbb{N}$ such that

$$\mathrm{P}\left(\left\|\left(K^{\mathcal{C}}\right)^{-1}\right\|_2 \leqslant L\right) \geqslant 1 - \eta, \forall n \geqslant N.$$

By Johnson and Horn (1985) (ch. 5.8), for any invertible matrices $K, \tilde{K}$ such that

$$\left\|\tilde{K}^{-1}(K - \tilde{K})\right\|_2 < 1,$$

$$\left\|K^{-1} - \tilde{K}^{-1}\right\|_2 \leqslant \frac{\left\|\tilde{K}^{-1}(K - \tilde{K})\right\|_2}{1 - \left\|\tilde{K}^{-1}(K - \tilde{K})\right\|_2} \left\|\tilde{K}^{-1}\right\|_2. \tag{A.169}$$

By the assumption that $D_{P_n}\left(\beta_n\right) \to D$ and $\frac{\sigma_{P_n,j}\left(\bar{\beta}\right)}{\sigma_{P_n,j}(\beta)} \to 1$ and Assumption 26, for any $\eta > 0$, there exists $N \in \mathbb{N}$ such that

$$\sup_{\Delta \in \boldsymbol{\Delta}} \left\|K_{P_n}^{\mathcal{C}}\left(\bar{\beta}\left(\beta_n, \Delta\right), \rho\right) - K^{\mathcal{C}}\right\|_2 \leqslant \eta, \forall n \geqslant N.$$

By (A.169), 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

$$\begin{aligned}
&P_n\left(\sup_{\Delta \in \boldsymbol{\Delta}} \left\|\left(K_{P_n}^{\mathcal{C}}\left(\bar{\beta}\left(\beta_n, \Delta\right), \rho\right)\right)^{-1}\right\|_2 \leqslant 2L_1\right) \\
\geqslant &P_n\left(\left\|\left(K^{\mathcal{C}}\right)^{-1}\right\|_2 + \sup_{\Delta \in \boldsymbol{\Delta}} \left\|K_{P_n}^{\mathcal{C}}\left(\bar{\beta}\left(\beta_n, \Delta\right), \rho\right)^{-1} - \left(K^{\mathcal{C}}\right)^{-1}\right\|_2 \leqslant 2L_1\right) \\
\geqslant &P_n\left(\left\|\left(K^{\mathcal{C}}\right)^{-1}\right\|_2 \leqslant L_1, \frac{\left\|\left(K^{\mathcal{C}}\right)^{-1}\right\|_2^2}{1 - \left\|\left(K^{\mathcal{C}}\right)^{-1}\left(K_{P_n}^{\mathcal{C}}\left(\bar{\beta}\left(\beta_n, \Delta\right), \rho\right) - K^{\mathcal{C}}\right)\right\|_2} \leqslant L_2,\right. \\
&\left.\sup_{\Delta \in \boldsymbol{\Delta}} \left\|K_{P_n}^{\mathcal{C}}\left(\bar{\beta}\left(\beta_n, \Delta\right), \rho\right) - K^{\mathcal{C}}\right\|_2 \leqslant \frac{L_1}{L_2}\right)
\end{aligned}$$

$$\geqslant 1-2\eta, \forall n \geqslant N, \tag{A.170}$$

Again by applying (A.169), for any $k>0$, there exists $N \in \mathbb{N}$ such that

$$\begin{aligned} &P_n\left(\left\|\left(K_{P_n}^{\mathcal{C}}\left(\bar{\beta}\left(\beta_n, \Delta\right)\right)\right)^{-1}-\left(K_{P_n}^{\mathcal{C}}\left(\bar{\beta}\left(\beta_n, \Delta'\right)\right)\right)^{-1}\right\|_2 \leqslant k\left\|\Delta-\Delta'\right\|\right) \\ \geqslant & P_n\left(\sup_{\Delta \in \boldsymbol{\Delta}}\left\|\left(K_{P_n}^{\mathcal{C}}\left(\bar{\beta}\left(\beta_n, \Delta\right)\right)\right)^{-1}\right\|_2^2 \frac{M \rho^2}{\sqrt{n}}\left\|\Delta-\Delta'\right\| \leqslant k\left\|\Delta-\Delta'\right\|\right) \geqslant 1-\eta, \quad \forall n \geqslant N, \end{aligned} \tag{A.171}$$

where the first inequality follows from

$$\left\|K_{P_n}^{\mathcal{C}}\left(\bar{\beta}\left(\beta_n, \Delta\right)\right)-K_{P_n}^{\mathcal{C}}\left(\bar{\beta}\left(\beta_n, \Delta'\right)\right)\right\|_2 \leqslant M \rho\left\|\bar{\beta}\left(\beta_n, \Delta\right)-\bar{\beta}\left(\beta_n, \Delta'\right)\right\| \leqslant M \rho^2 / \sqrt{n}\left\|\Delta-\Delta'\right\|$$

by Assumption 26, and the last inequality follows from (A.170).

By (A.167)-(A.168) and (A.170)-(A.171), it then follows that there exists $a \in [0,1)$ such that for any $\eta>0$, there exists $N \in \mathbb{N}$ such that

$$\mathrm{P}\left(\left|\phi_n(\Delta)-\phi_n\left(\Delta'\right)\right| \leqslant a\left\|\Delta-\Delta'\right\|, \quad \forall \Delta, \Delta' \in \boldsymbol{\Delta}\right) \geqslant 1-\eta, \quad \forall n \geqslant N. \tag{A.172}$$

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 et al. (2019) (Theorem 1.1)). This in turn implies that for any $\eta>0$ there exists a $N \in \mathbb{N}$ such that

$$P\left(\exists \Delta_n^f: \Delta_n^f=\phi_n\left(\Delta_n^f\right)\right) \geqslant 1-\eta, \forall n \geqslant N. \tag{A.173}$$

Step 3. I show that $\Delta_n^{\mathcal{C}}$ and Δ_n^f are close enough. Next, define the mapping

$$\psi_n(\Delta)=\left(K^{\mathcal{C}}\right)^{-1} g^{\mathcal{C}} \tag{A.174}$$

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 (A.166) and (A.174) arguing as in (A.167), it follows that for any $\Delta \in \boldsymbol{\Delta}$,

$$\begin{aligned} \left\|\psi_n(\Delta)-\phi_n(\Delta)\right\| \leqslant & \left\|\left(K_{P_n}^{\mathcal{C}}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta\right), \rho\right)\right)^{-1}\right\|_2\left\|g^{\mathcal{C}}-g_n^{\mathcal{C}}(\Delta)\right\| \\ & +\left\|\left(K^{\mathcal{C}}\right)^{-1}-\left(K_{P_n}^{\mathcal{C}}\left(\beta_n, \bar{\beta}\left(\beta_n, \Delta\right), \rho\right)\right)^{-1}\right\|_2\left\|g^{\mathcal{C}}\right\|. \end{aligned}$$

By (A.162) and (A.164)

$$\begin{aligned}\left\|g^{\mathcal{C}}-g_n^{\mathcal{C}}(\Delta)\right\| &\leqslant \max_{j\in\mathbb{J}^*}\left|-\mathbb{Z}_j^*-c_n^*/\left(1+\eta_{n,j}^*(\Delta)\right)+\mathbb{Z}_{n,j}^*(\Delta)\right| \\ &\leqslant \max_{j\in\mathbb{J}^*}\left|\mathbb{Z}_j^*-\mathbb{Z}_{n,j}^*\left(\Delta\right)\right|+\max_{j\in\mathbb{J}^*}\left|c_n^*/\left(1+\eta_{n,j}^*(\Delta)\right)\right|. \end{aligned} \tag{A.175}$$

Moreover, $\mathbb{Z}_n^*\left(\Delta\right)\overset{a.s.}{\to}\mathbb{Z}^*$ and (A.160) implies $c_n^*\to 0$ so that we have

$$\sup_{\Delta\in\boldsymbol{\Delta}}\left\|g^{\mathcal{C}}-g_n^{\mathcal{C}}\left(\Delta\right)\right\|\overset{a.s.}{\to}0.$$

Further, by (A.164), $D_{P_n}\to D$ and, Assumption 26, for any $\eta>0$, there exists $N\in\mathbb{N}$ such that

$$\sup_{\Delta\in\boldsymbol{\Delta}}\left\|\left(K^{\mathcal{C}}\right)^{-1}-\left(K_{P_n}^{\mathcal{C}}\left(\beta_n,\bar{\beta}\left(\beta_n,\Delta\right),\rho\right)\right)^{-1}\right\|_2\leqslant\eta,\forall n\geqslant N. \tag{A.176}$$

In sum, by (A.168), (A.170), and (A.175), (A.176), for any $\eta,\nu>0$, there exists $N\geqslant\mathbb{N}$ such that

$$P\left(\sup_{\Delta\in\boldsymbol{\Delta}}\left\|\psi_n(\Delta)-\phi_n(\Delta)\right\|<\nu\right)\geqslant 1-\eta,\forall n\geqslant\mathbb{N}. \tag{A.177}$$

Hence, for a specific choice of $\nu=\kappa(1-a)$, where a is defined in equation (A.172), we have that $\sup_{\Delta\in\boldsymbol{\Delta}}\left\|\psi_n(\Delta)-\phi_n(\Delta)\right\|<\kappa(1-a)$ implies

$$\begin{aligned}\left\|\Delta_n^{\mathcal{C}}-\Delta_n^f\right\| &= \left\|\psi_n\left(\Delta_n^{\mathcal{C}}\right)-\phi_n\left(\Delta_n^f\right)\right\| \\ &\leqslant \left\|\psi_n\left(\Delta_n^{\mathcal{C}}\right)-\phi_n\left(\Delta_n^{\mathcal{C}}\right)\right\|+\left\|\phi_n\left(\Delta_n^{\mathcal{C}}\right)-\phi_n\left(\Delta_n^f\right)\right\| \\ &\leqslant \kappa(1-a)+a\left\|\Delta_n^{\mathcal{C}}-\Delta_n^f\right\|\end{aligned}$$

Rearranging terms, we obtain $\left\|\Delta_n^{\mathcal{C}}-\Delta_n^f\right\|\leqslant\kappa$.

Step 4. I complete the proof. Note that by Assumptions 26 and 74.1, for any $\delta>0$, there exists $\kappa_\delta>0$ and $N\in\mathbb{N}$ such that

$$P\left(\sup_{\left\|\Delta-\Delta'\right\|\leqslant\kappa_\delta}\left|u_{n,j,\beta_n}^*(\Delta)-u_{n,j,\beta_n}^*\left(\Delta'\right)\right|<\delta\right)\geqslant 1-\eta,\forall n\geqslant N. \tag{A.178}$$

For $\Delta_n^{\mathcal{C}} \in \mathfrak{W}^{*,-\delta}(0)$, one has

$$\mathfrak{w}_j^*\left(\Delta_n^{\mathcal{C}}\right) + \delta \leqslant 0, j \in \mathbb{J}_1. \tag{A.179}$$

Hence, by (A.157), (A.160), and (A.178)-(A.179), $\left\|\Delta_n^{\mathcal{C}} - \Delta_n^f\right\| \leqslant \kappa_{\delta/4}$, for each $j \in \mathbb{J}_1$ we have

$$u_{n,j,\beta_n}^*\left(\Delta_n^f\right) - c_n^*\left(\beta_n\right) \leqslant u_{n,j,\beta_n}^*\left(\Delta_n^{\mathcal{C}}\right) - c_n^*\left(\beta_n\right) + \delta/4 \leqslant \mathfrak{w}_j^*\left(\Delta_n^{\mathcal{C}}\right) + \delta/2 \leqslant 0.$$

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 $n \geqslant N$ we have

$$P\left(\{U_n^*\left(\beta_n, c_n^*\right) = \varnothing\} \cap \left\{\mathfrak{W}^{*,-\delta}(0) \neq \varnothing\right\}\right) \leqslant P\left(\nexists \Delta_n^f \in U_n^*\left(\beta_n, c_n^*\right), \exists \Delta_n^{\mathcal{C}} \in \mathfrak{W}^{*,-\delta}(0)\right)$$

$$\leqslant P\left(\left\{\sup_{\Delta \in \mathbf{\Delta}} \|\psi_n(\Delta) - \phi_n(\Delta)\| < \kappa_{\delta/4}(1-a) \cap A_n(\frac{\delta}{2})\right\}^c\right) \leqslant \eta/3,$$

where A^c denotes the complement of the set A, and the last inequality follows from (A.157) and (A.177). □

Lemma 77. *Suppose Assumptions 26, 27, 25, 72 hold. In addition, suppose Assumption 28 or 73 hold. Let $\{P_n, \beta_n, \theta_n\}$ be a sequence satisfying (A.128)-(A.129). 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 $c > 0$,*

$$P_n^b\left(V_n^I\left(\beta_n', c\right) \neq \varnothing\right) - P(\mathfrak{W}(c) \neq \varnothing) \to 0,$$

   with probability approaching 1 ;

2. *If $c_{\pi^*} > 0, c_n^I\left(\beta_n'\right) \xrightarrow{P} c_{\pi^*}$;*

3.

$$\hat{c}_n\left(\beta_n'\right) \geqslant c_n^I\left(\beta_n'\right) + o_P(1).$$

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 26, $\left\|\hat{D}\left(\beta_n'\right) - D_{P_n}\left(\beta_n\right)\right\| \xrightarrow{p} 0$. Further, by Lemma 85, $\hat{\xi}_{n,j}\left(\beta_n'\right) \xrightarrow{P} \pi_{1,j}$. Therefore,

$$\left(\mathbb{Z}_n^b\left(\beta_n'\right), \hat{D}\left(\beta_n'\right), \hat{\xi}_n\left(\beta_n'\right)\right) \mid \{X_i\}_{i=1}^{\infty} \xrightarrow{d} (\mathbb{Z}, D, \pi_1) \tag{A.180}$$

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 $\left(\mathbb{Z}_n^b\left(\beta_n'\right), \hat{D}\left(\beta_n'\right), \hat{\xi}_n\left(\beta_n'\right)\right)$ defined on another probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ such that

$$\left(\tilde{\mathbb{Z}}_n^b, \tilde{D}_n, \tilde{\xi}_n\right) \stackrel{d}{=} \left(\mathbb{Z}_n^b\left(\beta_n'\right), \hat{D}\left(\beta_n'\right), \hat{\xi}_n\left(\beta_n'\right)\right)$$

conditional on the sample path. In particular, conditional on the sample, $\left(\hat{D}\left(\beta_n'\right), \hat{\xi}_n\left(\beta_n'\right)\right)$ are non-stochastic. Therefore, we set $\left(\tilde{D}_n, \tilde{\xi}_n\right) = \left(\hat{D}\left(\beta_n'\right), \hat{\xi}_n\left(\beta_n'\right)\right)$, $\tilde{P} - a.s.$ The almost sure representation satisfies $\left(\tilde{\mathbb{Z}}_n^b, \tilde{D}_n, \tilde{\xi}_{n,j}\right) \stackrel{\text{a.s.}}{\to} \left(\tilde{\mathbb{Z}}, D, \pi_1\right)$ 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

$$\begin{aligned}
\tilde{v}_{n,j,\beta_n'}^I(\Delta) &= \tilde{\mathbb{Z}}_{n,j}^s + \rho\tilde{D}_{n,j}\Delta + \varphi_j^*\left(\tilde{\xi}_{n,j}\right), \qquad \text{(A.181)}\\
\tilde{v}_{n,\ell,\beta_n'}^I(\Delta) &= \tilde{\mathbb{Z}}_{n,\ell}^s + \rho\tilde{D}_{n,\ell}\Delta,\\
\tilde{v}_{n,u,\beta_n'}^I(\Delta) &= \tilde{\mathbb{Z}}_{n,u}^s + \rho\tilde{D}_{n,u}\Delta + \varphi_0^*\left(\tilde{\xi}_{n,0}\right),\\
\tilde{\mathfrak{w}}_j(\Delta) &= \tilde{\mathbb{Z}}_j + \rho D_j\Delta + \pi_{1,j}^*\\
\tilde{\mathfrak{w}}_\ell(\Delta) &= \tilde{\mathbb{Z}}_\ell + \rho D_\ell\Delta\\
\tilde{\mathfrak{w}}_u(\Delta) &= \tilde{\mathbb{Z}}_u + \rho D_u\Delta + \pi_{1,0}^*
\end{aligned}$$

where $\tilde{\mathbb{Z}} \stackrel{d}{=} \mathbb{Z}$, and $\tilde{\mathbb{Z}}_n^b \to \tilde{\mathbb{Z}}, \tilde{P} - a.s.$ conditional on $\{X_i\}_{i=1}^{\infty}$. With this construction, one may write

$$\begin{aligned}
&\left|P_n^s\left(V_n^I\left(\beta_n', c\right) \neq \varnothing\right) - P(\mathfrak{W}(c) \neq \varnothing)\right|\\
=&\left|\tilde{P}\left(\tilde{V}_n^I\left(\beta_n', c\right) \neq \varnothing\right) - \tilde{P}(\tilde{\mathfrak{W}}(c) \neq \varnothing)\right|\\
\leqslant&\tilde{P}\left(\tilde{V}_n^I\left(\beta_n', c\right) = \varnothing \cap \tilde{\mathfrak{W}}(c) \neq \varnothing\right) + \tilde{P}\left(\tilde{V}_n^I\left(\beta_n', c\right) \neq \varnothing \cap \tilde{\mathfrak{W}}(c) = \varnothing\right). \qquad \text{(A.182)}
\end{aligned}$$

First, we bound the first term of (A.182). Note that

$$\begin{aligned}&\tilde{P}\left(\tilde{V}_n^I\left(\beta_n', c\right) = \varnothing \cap \tilde{\mathfrak{W}}(c) \neq \varnothing\right)\\ \leqslant &\tilde{P}\left(\tilde{V}_n^I\left(\beta_n', c+\delta\right) = \varnothing \cap \tilde{\mathfrak{W}}(c) \neq \varnothing\right) + \tilde{P}\left(\tilde{V}_n^I\left(\beta_n', c+\delta\right) \neq \varnothing \cap \tilde{V}_n^I\left(\beta_n', c\right) = \varnothing\right).\end{aligned} \tag{A.183}$$

Let

$$A_n = \left\{\tilde{\omega} \in \tilde{\Omega} : \sup_{\Delta \in \boldsymbol{\Delta}} \max_{j \in \mathbb{J}^*} \left|\tilde{v}_{n,j,\beta_n'}^I(\Delta) - \tilde{\mathfrak{w}}_j(\Delta)\right| \geqslant \delta\right\}.$$

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, $P_n(E) \geqslant 1 - \eta$ for all n sufficiently large by Assumption 26 and Lemma 85. 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 $j \in \mathbb{J}^*$,

$$\left|\tilde{v}_{n,j,\beta_n'}^I(\Delta) - \tilde{\mathfrak{w}}_j(\Delta)\right| \leqslant \left|\tilde{\mathbb{Z}}_{n,j}^s - \tilde{\mathbb{Z}}_j\right| + \rho\left\|\tilde{D}_{j,n} - D_j\right\| \|\Delta\| + \left|\phi_j^*\left(\tilde{\xi}_{n,j}\right) - \pi_{1,j}^*\right| \leqslant (2+\rho)\eta,$$

uniformly in $\Delta \in \boldsymbol{\Delta}$, where we used $\tilde{\mathbb{Z}}_n^s \to \tilde{\mathbb{Z}}, \tilde{P}$ - 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 $\sup_{\Delta \in \boldsymbol{\Delta}} \max_{j \in \mathbb{J}} \left|\tilde{v}_{n,j,\beta_n'}^I(\Delta) - \tilde{\mathfrak{w}}_j(\Delta)\right| < \delta$ implies $\tilde{\mathfrak{W}}(c) \subseteq \tilde{V}_n^I\left(\beta_n', c+\delta\right)$, and hence A_n^c is a subset of

$$L_n = \left\{\tilde{\omega} \in \tilde{\Omega} : \tilde{\mathfrak{W}}(c) \subseteq \tilde{V}_n^I\left(\beta_n', c+\delta\right)\right\}.$$

Using this,

$$\begin{aligned}\tilde{P}\left(\tilde{V}_n^I\left(\beta_n', c+\delta\right) = \varnothing \cap \tilde{\mathfrak{W}}(c) \neq \varnothing\right) &\leqslant \tilde{P}\left(\tilde{\mathfrak{W}}(c) \nsubseteq \tilde{V}_n^I\left(\beta_n', c+\delta\right)\right)\\ &= \tilde{P}\left(L_n^c\right) \leqslant \tilde{P}\left(A_n\right) < \eta/2,\end{aligned} \tag{A.184}$$

for all n sufficiently large. Also, by Lemma 79,

$$\tilde{P}\left(\tilde{V}_n^I\left(\beta_n', c+\delta\right) \neq \varnothing \cap \tilde{V}_n^I\left(\beta_n', c\right) = \varnothing\right) < \eta/2, \tag{A.185}$$

for all n sufficiently large. Combining (A.183), (A.184), (A.185), and using $P_n(E) \geqslant 1-\eta$ for all n, we have

$$\int_E \tilde{P}\left(\tilde{V}_n^I\left(\beta_n', c\right) = \varnothing \cap \tilde{\mathfrak{W}}(c) \neq \varnothing\right) dP_n + \int_{E^c} \tilde{P}\left(\tilde{V}_n^I\left(\beta_n', c\right) = \varnothing \cap \tilde{\mathfrak{W}}(c) \neq \varnothing\right) dP_n$$

$$\leqslant \eta(1-\eta) + \eta \leqslant 2\eta.$$

The second term of the right hand side of (A.182) can be bounded similarly. Therefore,

$$\left|P^s\left(V_n^I\left(\beta_n', c\right) \neq \varnothing\right) - P\left(\mathfrak{W}(c) \neq \varnothing\right)\right| \to 0$$

with probability (under P_n) approaching 1. This establishes the first claim.

(ii) By Part (i), for $c > 0$, we have

$$P_n^s\left(V_n^I\left(\beta_n', c\right) \neq \varnothing\right) - P\left(\mathfrak{W}(c) \neq \varnothing\right) \to 0$$

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 (A.185) and apply Lemma 79. Then, there is $\delta > 0$ such that

$$|P(\mathfrak{W}(c) \neq \varnothing) - P(\mathfrak{W}(c-\delta) \neq \varnothing)| = P\left(\{\mathfrak{W}(c) \neq \varnothing\} \cap \{\mathfrak{W}(c-\delta) = \varnothing\}\right) \leqslant \eta$$

$$|P\left(\mathfrak{W}(c+\delta) \neq \varnothing\right) - P\left(\mathfrak{W}(c) \neq \varnothing\right)| = P\left(\{\mathfrak{W}(c+\delta) \neq \varnothing\} \cap \{\mathfrak{W}(c) = \varnothing\}\right) \leqslant \eta$$

which therefore ensures that $c \mapsto P(\mathfrak{W}(c) \neq \varnothing)$ is continuous at $c > 0$.

Next, we show $c \mapsto P(\mathfrak{W}(c) \neq \varnothing)$ 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

$$K = \begin{bmatrix} D^{\mathbb{J}^*:} \\ I_d \\ -I_d \end{bmatrix}, g = \begin{bmatrix} e^{\mathbb{J}^*} \\ \rho \cdot 1_d \\ \rho \cdot 1_d \end{bmatrix}, \tau = \begin{bmatrix} 1_{J*} \\ 0_d \\ 0_d \end{bmatrix}. \tag{A.186}$$

By Farkas' lemma (R. Rockafellar (1970) Theorem 22.1) and arguing as in (A.206),

$$\begin{aligned} & P\left(\{\mathfrak{W}(c) \neq \varnothing\} \cap \{\mathfrak{W}(c-\delta) = \varnothing\}\right) \\ =& P\left(\left\{\mu' g \geqslant 0, \forall \mu \in \mathcal{M}\right\} \cap \left\{\mu'(g - \delta\tau) < 0, \exists \mu \in \mathcal{M}\right\}\right), \end{aligned} \tag{A.187}$$

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

$$\mathcal{M} = \left\{\mu : \mu = b \sum_{t=1}^{T} a_t \nu^t, b > 0, a_t \geqslant 0, \sum_{t=1}^{T} a_t = 1\right\}.$$

This implies

$$\mu' g \geqslant 0, \forall \mu \in \mathcal{M} \Leftrightarrow \nu^{t\prime} g \geqslant 0, \forall t \in \{1, \ldots, T\},$$

$$\mu'(g - \delta\tau) < 0, \exists \mu \in \mathcal{M} \Leftrightarrow \nu^{t\prime} g < \delta \nu^{t\prime} \tau, \exists t \in \{1, \ldots, T\}.$$

Hence,

$$(A.187) = P\left(0 \leqslant \nu^{s\prime} g, 0 \leqslant \nu^{t\prime} g < \delta \nu^{t\prime} \tau, \forall s, \exists t\right) \tag{A.188}$$

Note that by (A.186), for each $s \in \{1, \ldots, T\}$,

$$\nu^{s\prime} g = \left(\nu^{s,\mathbb{J}^*}\right)' \left(c 1_{\mathbb{J}^*} - \mathbb{Z}_{\mathbb{J}^*}\right) + \rho \sum_{j=J^*+1}^{J^*+2d} \nu^{s,j},$$

$$\nu^{s\prime} \tau = \sum_{j=1}^{J^*} \nu^{s,j}.$$

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 $0 \leqslant h_s^L < h_s^U$ 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 (A.188) as

$$\begin{aligned} &P\left(0 \leqslant \nu^{s\prime} g, 0 \leqslant \nu^{t\prime} g < \delta \nu^{t\prime}\tau, \forall s, \exists t\right) \\ =&P\left(\nu^{s,\mathbb{J}^{*\prime}}\mathbb{Z}_{\mathbb{J}^*} \leqslant h_s^U, h_t^L < \nu^{t,\mathbb{J}^{*\prime}}\mathbb{Z}_{\mathbb{J}^*} \leqslant h_t^U,\ \forall s,\ \exists t\right) > 0, \end{aligned} \tag{A.189}$$

where the last inequality follows because $\mathbb{Z}_{\mathbb{J}^*}$'s correlation matrix Ω has an eigenvalue bounded away from 0 by Assumption 28 and 73. By (A.187), (A.188), and (A.189), $c \mapsto P(\mathfrak{W}(c) \neq \varnothing)$ 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 $c_n^I(\beta_n') \xrightarrow{P_n} c_{\pi *}$

(iii) Begin with observing that one can equivalently express $\hat{c}_\ell$ (originally defined in (2.53)) as $\hat{c}_\ell(\beta) = \inf\{c \in \mathbb{R}_+ : P_n^s(V_n^s(\beta,c) \neq \varnothing) \geqslant 1-\alpha\}$ where

$$V_n^s\left(\beta_n', c\right) = \left\{\Delta \in \mathbf{\Delta}_{n,\rho}^d : v_{n,j,\beta_n'}^s(\Delta) \leqslant c, j \in \mathbb{J}\right\}, \tag{A.190}$$

$$v_{n,j,\beta_n'}^s(\Delta) = \mathbb{Z}_{n,j}^s\left(\beta_n'\right) + \rho\hat{D}_{n,j}\left(\beta_n'\right)\Delta + \hat{\xi}_{n,j}\left(\beta_n'\right). \tag{A.191}$$

Suppose first that Assumption 28 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, $\phi_j^*(\xi) \leqslant \xi$ for any j and ξ, and therefore $\hat{c}_\ell(\beta_n) \geqslant c_n^I(\beta_n)$ by construction.

Next, suppose that Assumption 73 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 $V_n^I\left(\beta_n', c\right)$, inequality $j + R_1$, which is

$$\mathbb{Z}_{n,j+R_1}^s\left(\beta_n'\right) + \rho\hat{D}_{n,j+R_1}\left(\beta_n'\right)\Delta \leqslant c$$

is replaced with inequality

$$-\mathbb{Z}_{n,j}^s\left(\beta_n'\right) - \rho\hat{D}_{n,j}\left(\beta_n'\right)\Delta \leqslant c,$$

as explained in Section A.4.3. In this case, $\hat{c}_n\left(\beta_n\right) \geqslant c_n^I\left(\beta_n\right)$ is not guaranteed in finite sample. However, let v_n^{IP} be as in (A.141) 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 (A.140) but replacing v_n^I with v_n^{IP}. Define

$$c_n^{IP}\left(\beta_n\right) = \inf\left\{c \in \mathbb{R}_+ : P^*\left(V_n^{IP}\left(\beta_n, c\right)\right) \geqslant 1 - \alpha\right\}.$$

By construction, $\hat{c}_n\left(\beta_n'\right) \geqslant c_n^{IP}\left(\beta_n'\right)$ for any $\beta_n' \in \left(\beta_n + \rho/\sqrt{n}B^d\right) \cap \overline{\mathcal{B}}$. Therefore, it suffices to show that $c_n^{IP}\left(\beta_n'\right) - c_n^I\left(\beta_n'\right) \xrightarrow{p} 0$. For this, note that Lemma 82.3 and 82.4 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 $c_n^{IP}\left(\beta_n'\right) - c_n^I\left(\beta_n'\right) \xrightarrow{p} 0$. □

Lemma 78. *Suppose Assumptions 26, 27, 25, 72 hold. In addition, suppose Assumption 28 or 73 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 $n \geqslant N'$,*

$$P\left(\sup_{\Delta \in \boldsymbol{\Delta}} \left| \max_{j \in \mathbb{J}} \left(u_{n,j,\beta_n}^*(\Delta) - c_n^*\right) - \max_{j \in \mathbb{J}} \left(\mathfrak{w}_j^*(\Delta) - c_{\pi *}\right) \right| \geqslant \varepsilon \right) < \eta, \tag{A.192}$$

$$\tilde{P}\left(\sup_{\Delta \in \boldsymbol{\Delta}} \left| \max_{j \in \mathbb{J}} \tilde{\mathfrak{w}}_j(\Delta) - \max_{j \in \mathbb{J}} \tilde{v}_{n,j,\beta_n'}^I(\Delta) \right| \geqslant \varepsilon \right) < \eta, \text{ w.p. 1.} \tag{A.193}$$

Proof. We first establish (A.192). By definition, $\pi^*_{1,j} = -\infty$ for all $j \notin \mathbb{J}^*$ and therefore

$$\begin{aligned} & P\left(\sup_{\Delta\in\mathbf{\Delta}}\left|\max_{j\in\mathbb{J}}\left(u^*_{n,j,\beta_n}(\Delta)-c^*_n\right)-\max_{j\in\mathbb{J}}\left(\mathfrak{w}^*_j(\Delta)-c_{\pi*}\right)\right|\geqslant\varepsilon\right) \\ =& P\left(\sup_{\Delta\in\mathbf{\Delta}}\left|\max_{j\in\mathbb{J}^*}\left(u^*_{n,j,\beta_n}(\Delta)-c^*_n\right)-\max_{j\in\mathbb{J}^*}\left(\mathfrak{w}^*_j(\Delta)-c_{\pi*}\right)\right|\geqslant\varepsilon\right). \end{aligned} \tag{A.194}$$

Hence, for the conclusion of the lemma, it suffices to show, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\sup_{\Delta\in\mathbf{\Delta}}\left|\max_{j\in\mathbb{J}^*}\left(u^*_{n,j,\beta_n}(\Delta)-c^*_n\right)-\max_{j\in\mathbb{J}^*}\left(\mathfrak{w}^*_j(\Delta)-c_{\pi*}\right)\right|\geqslant\varepsilon\right)=0.$$

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 $\Delta \in \mathbf{\Delta} \cap \frac{\sqrt{n}}{\rho}\left(\overline{\mathcal{B}}-\beta_n\right)$ and $j \in \mathbb{J}^*$, we have

$$\begin{aligned} |r_{n,j,\beta_n}(\Delta)| \leqslant & \left|\mathbb{Z}^*_{n,j,\beta_n}-\mathbb{Z}^*_j\right| + \rho\left\|\frac{\sigma_{F,j}\left(\bar{\beta}_n\right)}{\sigma_{F,j}\left(\beta_n\right)}D_{P_n,j}\left(\bar{\beta}_n\right)-D_j\right\|\|\Delta\| + \left|c^*_n-c_{\pi*}\right| \\ & + \left|\hat{\eta}_{n,j,\beta_n}(\Delta)\right|\left\|\mathbb{Z}_{n,j,\beta_n}(\Delta)+\frac{\sigma_{P_n,j}\left(\bar{\beta}_n\right)}{\sigma_{P_n,j}\left(\beta_n\right)}D_{P_n,j}\left(\bar{\beta}_n\right)\Delta\rho\right\| \\ =& o_P(1) \end{aligned} \tag{A.195}$$

where the first equality follows from $\|\Delta\| \leqslant \sqrt{d}, D_{P_n}\left(\bar{\beta}_n\right) \to D$ due to $D_{P_n}\left(\beta_n\right) \to D$, Assumption 26, Assumption 27, 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 27.1), and Lemma 83.

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 (A.106)-(A.107) and (A.131)-(A.132) respectively. By (A.195) and the fact that $j \in \mathbb{J}^*$, we have

$$\sup_{\Delta\in\mathbf{\Delta}}\left|\max_{j\in\mathbb{J}^*}\left(u^*_{n,j,\beta_n}(\Delta)-c^*_n\right)-\max_{j\in\mathbb{J}^*}\left(\mathfrak{w}^*_j(\Delta)-c_{\pi}\right)\right| \leqslant \sup_{\Delta\in\mathbf{\Delta}}\max_{j\in\mathbb{J}^*}\left|r_{n,j,\beta_n}(\Delta)\right| = o_P(1). \tag{A.196}$$

The conclusion of the lemma then follows from (A.194) and (A.196). The result in (A.193) follows from similar arguments. □

Lemma 79. *Let Assumptions 26, 27, 25, 72 hold. And suppose Assumption 28 or 73 hold. For any* $\beta_n' \in (\beta_n + \rho/\sqrt{n}\mathbf{\Delta}) \cap \overline{\mathcal{B}}$,

1. *For any* $\eta > 0$, *there exist* $\delta > 0$ *such that*

$$\sup_{c \geqslant 0} P\left(\{\mathfrak{W}(c) \neq \varnothing\} \cap \left\{\mathfrak{W}^{-\delta}(c) = \varnothing\right\}\right) < \eta. \tag{A.197}$$

Moreover, for any $\eta > 0$, *there exist* $\delta > 0$ *and* $N \in \mathbb{N}$ *such that*

$$\sup_{c \geqslant 0} P_n^s\left(\left\{V_n^I\left(\beta_n', c\right) \neq \varnothing\right\} \cap \left\{V_n^{I,-\delta}\left(\beta_n', c\right) = \varnothing\right\}\right) < \eta, \quad \forall n \geqslant N. \tag{A.198}$$

2. *Fix* $\underline{c} > 0$. *Then for any* $\eta > 0$, *there exists* $\delta > 0$ *such that*

$$\sup_{c \geqslant \underline{c}} P\left(\{\mathfrak{W}(c) \neq \varnothing\} \cap \{\mathfrak{W}(c - \delta) = \varnothing\}\right) < \eta. \tag{A.199}$$

Moreover, for any $\eta > 0$, *there exist* $\delta > 0$ *and* $N \in \mathbb{N}$ *such that*

$$\sup_{c \geqslant \underline{c}} P_n^s\left(\left\{V_n^I\left(\beta_n', c\right) \neq \varnothing\right\} \cap \left\{V_n^I\left(\beta_n', c - \delta\right) = \varnothing\right\}\right) < \eta, \quad \forall n \geqslant N. \tag{A.200}$$

Proof. We first show (A.197). Any inequality indexed by $j \notin \mathbb{J}^*$ is satisfied with probability approaching one by similar arguments as in (A.117) (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 $c \geqslant 0$ be given. Let g be a $|\mathbb{J}| + 2d$ vector with entries

$$g_j = \begin{cases} c - \mathbb{Z}_j, & j \in \mathbb{J}, \\ 1, & j = |\mathbb{J}| + 1, \ldots, |\mathbb{J}| + 2d, \end{cases} \tag{A.201}$$

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

$$\tau_j = \begin{cases} 1, & j = 1, \ldots, |\mathbb{J}_1| \\ 0, & j = |\mathbb{J}_1| + 1, \ldots, |\mathbb{J}| + 2d. \end{cases} \tag{A.202}$$

Then we can express the sets of interest as

$$\mathfrak{W}(c) = \{\Delta : K\Delta \leqslant g\}, \tag{A.203}$$

$$\mathfrak{W}^{-\delta}(c) = \{\Delta : K\Delta \leqslant g - \delta\tau\} \tag{A.204}$$

By Farkas' Lemma, e.g. R. Rockafellar (1970) (Theorem 22.1), a solution to the system of linear inequalities in (A.203) exists if and only if for all $\mu \in \mathbb{R}_+^{J+2d}$ such that $\mu' K = 0$, one has $\mu' g \geqslant 0$. Similarly, a solution to the system of linear inequalities in (A.204) exists if and only if for all $\mu \in \mathbb{R}^{|\mathbb{J}|+2d}$ such that $\mu' K = 0$, one has $\mu'(g - \delta\tau) \geqslant 0$. Define

$$\mathcal{M} \equiv \left\{\mu \in \mathbb{R}_+^{|\mathbb{J}|+2d} : \mu' K = 0\right\}. \tag{A.205}$$

Then, one may write

$$\begin{aligned} & P\left(\{\mathfrak{W}(c) \neq \varnothing\} \cap \left\{\mathfrak{W}^{-\delta}(c) = \varnothing\right\}\right) \\ =& P\left(\left\{\mu' g \geqslant 0, \forall \mu \in \mathcal{M}\right\} \cap \left\{\mu'(g - \delta\tau) < 0, \exists \mu \in \mathcal{M}\right\}\right) \\ =& P\left(\left\{\mu' g \geqslant 0, \forall \mu \in \mathcal{M}\right\} \cap \left\{\mu' g < \delta\mu'\tau, \exists \mu \in \mathcal{M}\right\}\right). \end{aligned} \tag{A.206}$$

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

$$\left\{\nu^t \in \mathcal{M}, t = 1, \ldots, T\right\},$$

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 $a_t \geqslant 0$, $t = 1, \ldots, T$, $\sum_{t=1}^{T} a_t = 1$. 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 $a_t \geqslant 0, \forall t$ 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 $0 \leqslant \nu^{t\prime} g$. Therefore,

$$P\left(\left\{\mu' g \geqslant 0, \forall \mu \in \mathcal{M}\right\} \cap \left\{\mu' g < \delta \mu' \tau, \exists \mu \in \mathcal{M}\right\}\right) \tag{A.207}$$

$$\leqslant P\left(0 \leqslant \nu^{t\prime} g < \delta \nu^{t\prime} \tau, \exists t \in \{1, \ldots, T\}\right) \leqslant \sum_{t=1}^{T} P\left(0 \leqslant \nu^{t\prime} g < \delta \nu^{t\prime} \tau\right).$$

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 (A.202). Therefore $P\left(0 \leqslant \nu^{t\prime} g < \delta \nu^{t\prime} \tau\right) = 0$.

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:

$$g = \begin{cases} c - \mathbb{Z}_j & j \in \mathbb{J}_2, \\ c + \mathbb{Z}_{[j]} & j \in [\mathbb{J}_2]. \end{cases} \tag{A.208}$$

For each ν^t, (A.208) 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

$$\tilde{\nu}_j^t \equiv \begin{cases} -\nu_j^t & j \in \mathbb{J}_1 \\ -\nu_j^t + \nu_{[j]}^t & j \in \mathbb{J}_2. \end{cases} \tag{A.209}$$

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

$$\nu^{t\prime} g = \sum_{j \in \mathbb{J}_1 \cup \mathbb{J}_2} \tilde{\nu}^t_j \mathbb{Z}_j + c \sum_{j \in \mathbb{J}} \nu^t_j + \sum_{j=|\mathbb{J}|+1}^{|\mathbb{J}|+2d} \nu^t_j \tag{A.210}$$

Case 2-a. Suppose $\tilde{\nu}^t \neq 0$. Then, by (A.210), $\frac{\nu^{t\prime} g}{\nu^{t\prime} \tau}$ 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 28 and Assumption 73, 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

$$\frac{\tilde{\nu}'^t \Omega \tilde{\nu}^t}{\left(\tilde{\nu}^{t\prime}\tau\right)^2} \geqslant \frac{\omega \left\|\tilde{\nu}^t\right\|_2^2}{(|\mathbb{J}|+2d)^2 \left\|\tilde{\nu}^t\right\|_2^2} \geqslant \frac{\omega}{(|\mathbb{J}|+2d)^2}.$$

Therefore, the variance of the normal random variable in (A.207) is uniformly bounded away from 0, which in turn allows one to find $\delta > 0$ such that $P\left(0 \leqslant \frac{\nu^t g}{\nu^t \tau} < \delta\right) \leqslant \eta/T$.

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 $\sum_{j \in \mathbb{J}} \nu^t_j g \geqslant 0$, 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 $P\left(0 \leqslant \nu^{t\prime} g < \delta \nu^{t\prime} \tau\right) = 0$. This establishes equation (A.197).

As for (A.198), 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 (A.165) 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 (A.205) 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 (A.180), $\mathbb{Z}^s_n \xrightarrow{d} \mathbb{Z}$ in $l^\infty(\Theta)$ uniformly in $\mathcal{P}$ conditional on $\{X_1, \ldots, X_n\}$, and by Assumption 26 $\hat{D}_n\left(\beta'_n\right) \xrightarrow{P_n} D$, 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 (A.201) to conclude (A.198).

The results in (A.199)-(A.200) follow by similar arguments, with proper redefinition of τ in equation (A.202). □

Lemma 80. *Let Assumptions 26, 27, 25, 72 hold. And suppose Assumption 28 or 73 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 76. Then, for any $\eta > 0$, there exist $0 < \varepsilon_\eta < \infty$ and $N \in \mathbb{N}$ s.t. $n \geqslant N$ implies*

$$P\left(\mathfrak{W}^*(-\delta) \neq \varnothing, \left|\det K^{\mathcal{C}}\right| \leqslant \varepsilon_\eta\right) \leqslant \eta \tag{A.211}$$

Proof. (A.211) can be bounded as follows:

$$\begin{aligned} P\left(\mathfrak{W}^{*,-\delta}(0) \neq \varnothing, \left|\det K^{\mathcal{C}}\right| \leqslant \varepsilon_\eta\right) &\leqslant P\left(\exists C \in \widetilde{\mathcal{C}} : \Delta^C \in \mathbf{\Delta}, \left|\det K^C\right| \leqslant \varepsilon_\eta\right) \\ &\leqslant \sum_{C \in \widetilde{\mathcal{C}} : |\det K^C| \leqslant \varepsilon_\eta} \mathrm{P}\left(\Delta^C \in \mathbf{\Delta}\right) \\ &\leqslant \sum_{C \in \widetilde{\mathcal{C}} : |\alpha^C| \leqslant \varepsilon_\eta^{2/d}} P\left(\Delta^C \in \mathbf{\Delta}\right) \end{aligned}$$

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 $\left|\det K^C\right| \geqslant \left|\alpha^C\right|^{d/2}$.

Noting that $\widetilde{\mathcal{C}}$ has $\begin{pmatrix} |\mathbb{J}|+2d \\ d \end{pmatrix}$ elements, it suffices to show that

$$\left|\alpha^C\right| \leqslant \varepsilon_\eta^{2/d} \Longrightarrow P\left(\Delta^C \in \mathbf{\Delta}\right) \leqslant \bar{\eta} \equiv \frac{\eta}{\binom{|\mathbb{J}|+2d}{d}}.$$

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:

$$\left|q^{C\prime} K^C K^{C\prime} q^C\right| \leqslant \varepsilon_\eta^{2/d} \Longrightarrow P\left(\left(K^C\right)^{-1} g^C \in \mathbf{\Delta}\right) \leqslant \bar{\eta}.$$

Now, if $\left|q^{C\prime} K^C K^{C\prime} q^C\right| \leqslant \varepsilon_\eta^{2/d}$ 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

$$P\left(\left(K^C\right)^{-1} g^C \in \mathbf{\Delta}\right) \leqslant P\left(\left|q^{C\prime} g^C\right| < \sqrt{d}\varepsilon_\eta^{1/d}\right).$$

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 28 (or 73) and Assumption 74 yield

$$\begin{aligned} & \operatorname{eig}(\tilde{\Omega}) \geqslant \omega \\ \Longrightarrow & \operatorname{Var}_P\left(q^{C\prime} g^C\right) \geqslant \omega \\ \Longrightarrow P\left(\left|q^{C\prime} g^C\right| < \sqrt{d}\varepsilon_\eta^{1/d}\right) &= P\left(-\sqrt{d}\varepsilon_\eta^{1/d} < q^{C\prime} g^C < \sqrt{d}\varepsilon_\eta^{1/d}\right) \\ & < \frac{2\sqrt{d}\varepsilon_\eta^{1/d}}{\sqrt{2\omega\pi}}, \end{aligned} \tag{A.212}$$

where the result in (A.212) 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 81. *Assumptions 26, 27, 25, 72 hold. In addition, suppose Assumption 28 or 73 hold. If* $|\mathbb{J}_1| \geqslant 1, |\mathbb{J}_2| \geqslant d$, *or if* $|\mathbb{J}_2| > d$, *then* $\exists \underline{c} > 0$ *s.t.*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\beta\in\mathcal{B}(P)} P\left(c_n^I(\beta) \geqslant \underline{c}\right) = 1.$$

Proof. I first consider the case where $\mathbb{J}_1 \neq \varnothing$ and $|\mathbb{J}_2| \geqslant d$. Fix any $c \geqslant 0$ 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 $\mathbb{Z}_{n,j}^s(\beta)$ 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 (A.213) to (A.215) below.) Then the coverage probability $P^s\left(V_n^I(\beta, c) \neq \varnothing\right)$ is asymptotically bounded above by

$$\begin{aligned}
&P^s\left(\left\{\Delta \in \rho\mathbf{\Delta}_n : \hat{D}_{n,j}\Delta \leqslant c - \mathbb{Z}_{n,j}^s, j \in \mathbb{J}^*\right\} \neq \varnothing\right) \\
\leqslant &P^s\left(\left\{\Delta \in \mathbb{R}^d : \hat{D}_{n,j}\Delta \leqslant c - \mathbb{Z}_{n,j}^s, j \in \bar{C}\right\} \neq \varnothing\right) && \text{(A.213)} \\
= &P^s\left(\hat{D}_{n,1}'\left(\hat{D}_n^{\bar{C}}\right)^{-1}\left(c\mathbf{1}_d - \mathbb{Z}_n^{s,\bar{C}}\right) \leqslant c - \mathbb{Z}_{n,1}^b\right) && \text{(A.214)}
\end{aligned}$$

$$
\begin{aligned}
=&P^s\left(\left(1-\hat{D}'_{n,1}\left(\hat{D}_n^{\bar{C}}\right)^{-1}\mathbf{1}_d\right)c \geqslant \mathbb{Z}^s_{n,1}-c\hat{D}'_{n,1}\left(\hat{D}_n^{\bar{C}}\right)^{-1}\mathbb{Z}_n^{s,\bar{C}}\right)\\
=&P^s\left(\frac{\left(1-\hat{D}'_{n,1}\left(\hat{D}_n^{\bar{C}}\right)^{-1}\mathbf{1}_d\right)c}{\sqrt{\Omega_P^C}} \geqslant \frac{\mathbb{Z}^s_{n,1}-c\hat{D}'_{n,1}\left(\hat{D}_n^{\bar{C}}\right)^{-1}\mathbb{Z}_n^{s,\bar{C}}}{\sqrt{\Omega_P^C}}\right) \qquad \text{(A.215)}\\
\leqslant&\Phi\left(\frac{\left(1+\left|\hat{D}'_{n,1}\left(\hat{D}_n^{\bar{C}}\right)^{-1}\mathbf{1}_d\right|\right)c}{\sqrt{\Omega_P^C}}\right)+o_p(1) \qquad \text{(A.216)}
\end{aligned}
$$

Here, (A.213) removes constraints and hence enlarges the feasible set; (A.214) solves in closed form; and (A.216) uses that by Assumption 28 or 73.

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

$$
\begin{aligned}
&P^s\left(\left\{\Delta\in\rho\boldsymbol{\Delta}_n:\hat{D}_{n,j}\Delta\leqslant c-\mathbb{Z}^s_{n,j},j\in\mathbb{J}^*\right\}\neq\varnothing\right) \qquad \text{(A.217)}\\
\leqslant&P^s\left(\exists\Delta:\hat{D}_{n,j}\Delta\leqslant c-\mathbb{Z}^s_{n,j},j\in\bar{C}\cup\{|\mathbb{J}_1|+|\mathbb{J}_2|+1\}\right)\\
\leqslant&P^s\left(\sup_{\Delta}\left\{\hat{D}_{n,|\mathbb{J}_1|+1}\Delta:\hat{D}_{n,j}\Delta\leqslant c-\mathbb{Z}^s_{n,j},j\in\bar{C}\right\}\right.\\
&\qquad\left.\geqslant\inf_{\Delta}\left\{\hat{D}_{n,|\mathbb{J}_1|+1}\Delta:\hat{D}_{n,|\mathbb{J}_1|+|\mathbb{J}_2|+1}\Delta\leqslant c-\mathbb{Z}^s_{n,|\mathbb{J}_1|+|\mathbb{J}_2|+1}\right\}\right) \qquad \text{(A.218)}\\
=&P^s\left(\hat{D}_{n,|\mathbb{J}_1|+1}\hat{D}_n^{\bar{C}\prime}\left(\hat{D}_n^{\bar{C}}\hat{D}_n^{\bar{C}\prime}\right)^{-1}\left(c\mathbf{1}_{\bar{d}}-\mathbb{Z}_n^{s,\bar{C}}\right)\geqslant -c+\mathbb{Z}^s_{n,|\mathbb{J}_1|+|\mathbb{J}_2|+1}\right) \qquad \text{(A.219)}
\end{aligned}
$$

Here, the reasoning from (A.217) to (A.218) holds because we evaluate the probability of increasingly larger events; in particular, if the event in (A.218) 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 (A.219) resembles (A.215), and the argument can be concluded in analogy to Case (ii). □

Lemma 82. *Assumptions 26, 27, 25, 72 hold. In addition, suppose Assumption 28 or 73 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 77. 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. $\left|\mathbb{Z}_{n,j}(\beta_n') + \mathbb{Z}_{n,[j]}(\beta_n')\right| \xrightarrow{P_n} 0$, *and* $\left|\mathbb{Z}^b_{n,j}(\beta_n') + \mathbb{Z}^b_{n,[j]}(\beta_n')\right| \xrightarrow{P^*} 0$ *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 85, 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 $\frac{\sqrt{n}E_{P_n}\left[m_{[j]}(X_i;\beta_n',\varphi_{P_n})\right]}{\kappa_n\sigma_{P_n,[j]}(\beta_n')}$.

(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 73, $E_{P_n}(t_j(X_i,\beta_n')) \to 0$. We then have, using Assumption 73 again,

$$\begin{aligned}\operatorname{Var}\left(t_j\left(X_i,\beta_n'\right)\right) &= \int t_j\left(x,\beta_n'\right)^2 dP_n(x) - E_{P_n}\left[t_j\left(X_i,\beta_n'\right)\right]^2 \\ &\leqslant M\int t_j\left(x,\beta_n'\right) dP_n(x) - E_{P_n}\left[t_j\left(X_i,\beta_n'\right)\right]^2 \to 0. \end{aligned} \tag{A.220}$$

Hence,

$$\begin{aligned}&\Omega_{P_n,[j]}\left(\beta_n',\varphi_{P_n}\right) - \Omega_{P_n,j}\left(\beta_n',\varphi_{P_n}\right) \\ &= \operatorname{var}_{P_n}\left(t_j\left(X_i;\beta_n',\varphi_{P_n}\right)\right) + 2\operatorname{cov}_{P_n}\left(m_j\left(X_i;\beta_n',\varphi_{P_n}\right), t_j\left(X_i;\beta_n',\varphi_{P_n}\right)\right) \\ &\leqslant \operatorname{var}_{P_n}\left(t_j\left(X_i,\beta_n'\right)\right) + 2\left(\operatorname{var}_{P_n}\left(t_j\left(X_i,\beta_n'\right)\right)\right)^{1/2}\left(\operatorname{var}_{P_n}\left(m_j\left(X_i;\beta_n',\varphi_{P_n}\right)\right)\right)^{1/2} \to 0,\end{aligned} \tag{A.221}$$

And similarly

$$\Omega_{P_n,[j],\varphi}\left(\beta_n',\varphi_{P_n}\right) - \Omega_{P_n,j,\varphi}\left(\beta_n',\varphi_{P_n}\right) = \operatorname{cov}\left(t_j\left(X_i;\beta_n',\varphi_{P_n}\right), m_\varphi(X_i)\right)$$

$$\leqslant \sqrt{\operatorname{var}_{P_n}\left(t_j\left(X_i, \beta_n'\right)\right) \operatorname{var}_{P_n}\left(m_\varphi(X_i)\right)} \to 0.$$

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

$$\begin{aligned} E_{P_n}\left[t_j\left(X_i; \beta_n', \varphi_{P_n} + r_n\right)\right] =& E_{P_n}\left[t_j\left(X_i; \beta_n', \varphi_{P_n}\right)\right] + \nabla_\varphi E_{P_n}\left[t_j\left(X_i; \beta_n', \varphi_{P_n}\right)\right] r_n \\ &+ \left(\nabla_\varphi E_{P_n}\left[t_j\left(X_i; \beta_n', \bar{\varphi}\right)\right] - \nabla_\varphi E_{P_n}\left[t_j\left(X_i; \beta_n', \varphi_{P_n}\right)\right]\right) r_n \\ =& O(\frac{\kappa_n}{\sqrt{n}}) - \|q_n\| \frac{\kappa_n^2}{\sqrt{n}} + O\left(\frac{\kappa_n^4}{n}\right) \end{aligned}$$

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 73 that $t_j\left(X_i; \beta_n', \varphi_{P_n} + r_n\right) \geqslant 0$.

Then by

$$\begin{aligned} \sigma_{P_n,j}^2\left(\beta_n'\right) =& \left[\begin{array}{cc} 1 & \nabla_{\varphi'} E\left[m_j\left(X_i; \beta_n', \varphi_{P_n}\right)\right] \end{array}\right] \left[\begin{array}{cc} \Omega_{P_n,j}\left(\beta_n'\right) & \Omega_{P_n,j,\varphi}\left(\beta_n'\right) \\ \Omega_{P_n,j,\varphi}\left(\beta_n'\right) & \Omega_{P_n\varphi}\left(\beta_n'\right) \end{array}\right] \\ & \times \nabla_{\varphi'} E\left[m_j\left(X_i; \beta_n', \varphi_{P_n}\right)\right] \end{aligned}$$

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 (A.220) and (A.221).

(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 83. *Assumptions 26, 27, 25, 72 hold. And suppose Assumption 28 or 73 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 $n \geqslant N$.

Proof. (1) First, for any $\epsilon > 0$ and for any $j = 1, ..., J_1 + J_2$, by Assumption 72, 74, 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

$$\sup_{P \in \mathcal{P}} P\left(\sup_{m \geqslant n_1} \sup_{(\beta,\varphi) \in \overline{\mathcal{B}} \times \Phi} \left\|\hat{\Omega}(\beta, \varphi) - \Omega_P(\beta, \varphi)\right\| \leqslant \epsilon\right) \to 0.$$

Next, note that by Assumption 72,

$$\sup_{P\in\mathcal{P}} P\left(\sup_{m\geqslant n}\|\hat{\varphi}-\varphi_P\|\leqslant\epsilon\right)\to 0.$$

Therefore, we have

$$\begin{aligned}
&\sup_{P\in\mathcal{P}} P\left(\sup_{m\geqslant n}\sup_{\beta\in\overline{\mathcal{B}}}\left\|\hat{\Omega}(\beta)-\Omega_P(\beta)\right\|>\epsilon\right)\\
\leqslant&\sup_{P\in\mathcal{P}} P\left(\sup_{m\geqslant n}\sup_{\beta\in\overline{\mathcal{B}}}\left\|\hat{\Omega}(\beta,\hat{\varphi})-\Omega_P(\beta,\hat{\varphi})\right\|>\frac{\epsilon}{2}\right)\\
&+\sup_{P\in\mathcal{P}} P\left(\sup_{m\geqslant n}\sup_{\beta\in\mathcal{B}}\|\Omega_P(\beta,\hat{\varphi})-\Omega_P(\beta,\varphi_P)\|>\frac{\epsilon}{2}\right)\\
\leqslant&\sup_{P\in\mathcal{P}} P\left(\sup_{m\geqslant n}\sup_{(\beta,\varphi)\in\overline{\mathcal{B}}\times\Phi}\left\|\hat{\Omega}(\beta,\varphi)-\Omega_P(\beta,\varphi)\right\|>\frac{\epsilon}{2}\right)\\
&+\sup_{P\in\mathcal{P}} P\left(\sup_{m\geqslant n}M\|\hat{\varphi}-\varphi_P\|>\frac{\epsilon}{2}\right)\to 0.
\end{aligned}$$

Together with Assumption 2 that $\hat{G}(\beta)-G_P(\beta)$ converges almost sure to 0, and Assumption 72.3, Assumption 26 that Σ_P and G_P are bounded, we can get

$$\inf_{P\in\mathcal{P}} P\left(\sup_{m\geqslant n}\sup_{(\beta,\varphi)\in\overline{\mathcal{B}}\times\Phi}\left\|\hat{\Sigma}(\beta)-\Sigma_P(\beta)\right\|\leqslant\epsilon\right)\to 1.$$

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

$$\eta_{1,n,j,\beta}(\Delta)=\frac{\sigma_{P,j}(\beta+\frac{\Delta\rho}{\sqrt{n}},\varphi_P)}{\hat{\sigma}_{n,j}(\beta+\frac{\Delta\rho}{\sqrt{n}},\hat{\varphi})}-1,$$

$$\eta_{2,n,j,\beta}(\Delta) = \frac{\sigma_{P,j}(\beta, \varphi_P)}{\sigma_{P,j}(\beta + \frac{\Delta\rho}{\sqrt{n}}, \varphi_P)} - 1.$$

We can conclude that

$$\inf_{P\in\mathcal{P}} P\left(\sup_{m\geqslant n} \sup_{(\beta,\varphi)\in\overline{\mathcal{B}}\times\Phi} |\hat{\eta}_{n,j,\beta}(\Delta) - 1| \leqslant \epsilon\right) \to 1.$$

Finally, note that for any $\epsilon > 0$,

$$\begin{aligned}
1 &= \lim_{n\to\infty} \inf_{P\in\mathcal{P}} P\left(\sup_{m\geqslant n} \sup_{(\beta,\varphi)\in\overline{\mathcal{B}}\times\Phi} |\hat{\eta}_{n,j,\beta}(\Delta)| \leqslant \epsilon\right) \\
&\leqslant \inf_{P\in\mathcal{P}} \lim_{n\to\infty} P\left(\bigcap_{m\geqslant n} \left\{\sup_{(\beta,\varphi)\in\overline{\mathcal{B}}\times\Phi} |\hat{\eta}_{n,j,\beta}(\Delta)| \leqslant \epsilon\right\}\right) \\
&= \inf_{P\in\mathcal{P}} P\left(\lim_{n\to\infty} \bigcap_{m\geqslant n} \left\{\sup_{(\beta,\varphi)\in\overline{\mathcal{B}}\times\Phi} |\hat{\eta}_{n,j,\beta}(\Delta)| \leqslant \epsilon\right\}\right) \\
&= \inf_{P\in\mathcal{P}} P\left(\sup_{(\beta,\varphi)\in\overline{\mathcal{B}}\times\Phi} |\hat{\eta}_{n,j,\beta}(\Delta)| \leqslant \epsilon, \text{ for all but finite } n\right),
\end{aligned}$$

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 $\hat{\mu}_{n,j}$:

$$\begin{aligned}
\hat{\mu}_{n,[j]}(\beta) &= \min\left\{\max\left(0, \frac{\frac{\bar{m}_{n,j}(\beta,\hat{\varphi})}{\hat{\sigma}_{n,j}(\beta)}}{\frac{\bar{m}_{n,[j]}(\beta,\hat{\varphi})}{\hat{\sigma}_{n,[j]}(\beta)} + \frac{\bar{m}_{n,j}(\beta,\hat{\varphi})}{\hat{\sigma}_{n,j}(\beta)}}\right), 1\right\}, \\
\hat{\mu}_{n,j}(\beta) &= 1 - \hat{\mu}_{n,[j]}(\beta).
\end{aligned}$$

Note that

$$\begin{aligned}
&\sup_{\beta_n'\in\beta_n+\rho/\sqrt{n}\boldsymbol{\Delta}} \left|\kappa_n^{-1} \frac{\sqrt{n}\bar{m}_{n,j}(\beta_n', \hat{\varphi})}{\hat{\sigma}_{n,j}(\beta_n')} - \kappa_n^{-1} \frac{\sqrt{n} E_{P_n}[m_j(X_i; \beta_n', \varphi_{P_n})]}{\sigma_{P_n,j}(\beta_n')}\right| \\
\leqslant &\sup_{\beta_n'\in\beta_n+\rho/\sqrt{n}\boldsymbol{\Delta}} \left|\frac{\mathbb{G}_n(\beta_n', \hat{\varphi})}{\kappa_n \hat{\sigma}_{n,j}(\beta_n')} + \frac{\sqrt{n}}{\kappa_n}\gamma_{1,P,j}(\beta_n')\hat{\eta}_{n,j,\beta_n'} + \frac{\nabla_\varphi E_{P_n}[m_j(X_i; \beta_n', \bar{\varphi})]\sqrt{n}(\hat{\varphi} - \varphi)}{\kappa_n \hat{\sigma}_{n,j}(\beta_n')}\right| \\
= &o_P(1),
\end{aligned}$$

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

$$\begin{aligned}\frac{\sigma_{P_n,j}(\beta_n)}{\hat{\sigma}^M_{n,j}(\beta_n)} - 1 =& \frac{\sigma_{P_n,j}(\beta_n)}{\hat{\sigma}_{n,j}(\beta_n)}\left(\frac{\hat{\sigma}_{n,j}(\beta_n)}{\hat{\sigma}^M_{n,j}(\beta_n)} - 1\right) + \left(\frac{\sigma_{P_n,j}(\beta_n)}{\hat{\sigma}_{n,j}(\beta_n)} - 1\right) \\ =& O_{P_n}(1)\left(\frac{\hat{\sigma}_{n,j}(\beta_n,\hat{\varphi})}{\hat{\sigma}^M_{n,j}(\beta_n,\hat{\varphi})} - 1\right) + o_p(1),\end{aligned} \tag{A.222}$$

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

$$\begin{aligned}\frac{\hat{\sigma}_{n,j}(\beta_n)}{\hat{\sigma}^M_{n,j}(\beta_n)} - 1 &= \frac{\hat{\sigma}_{n,j}(\beta_n) - \hat{\sigma}^M_{n,j}(\beta_n)}{\hat{\sigma}^M_{n,j}(\beta_n)} \\ &= \frac{(1-\hat{\mu}_{n,j}(\beta_n))\,O_{P_n}(\hat{\sigma}_{n,j}(\beta_n))}{(1+o_{P_n}(1))\,\hat{\sigma}_{n,j}(\beta_n) + (1-\hat{\mu}_{n,j}(\beta_n))\,O_{P_n}(\hat{\sigma}_{n,j}(\beta_n))} = o_p(1),\end{aligned} \tag{A.223}$$

where we used $\hat{\sigma}^{-1}_{n,j}(\beta_n) = O_p(1)$ by Assumption 28 or 73 and part (i) of the lemma. By (A.222) and (A.223), $\sigma_{P_n,j}(\beta_n)/\hat{\sigma}^M_{n,j}(\beta_n) - 1 = o_p(1)$.

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 82 (1),

$$\left|\frac{\sigma_{P_n,j}(\beta_n)}{\sigma_{P_n,[j]}(\beta_n)} - 1\right| = o_p(1), \tag{A.224}$$

for $j = \mathbb{J}_2 \cup [\mathbb{J}_2]$. Therefore,

$$\frac{\sigma_{P_n,j}(\beta_n)}{\hat{\sigma}^M_{n,j}(\beta_n)} - 1 = \frac{\sigma_{P_n,j}(\beta_n) - \hat{\sigma}^M_{n,j}(\beta_n)}{\hat{\sigma}^M_{n,j}(\beta_n)}$$

$$= \frac{\left[\hat{\mu}_{n,j}(\beta_n) + (1 - \hat{\mu}_{n,j}(\beta_n))\right]\sigma_{P_n,j}(\beta_n) - \left[\hat{\mu}_{n,j}(\beta_n)\hat{\sigma}_{n,j}(\beta_n) + (1 - \hat{\mu}_{n,j}(\beta_n))\hat{\sigma}_{n,[j]}(\beta_n)\right]}{\hat{\sigma}^M_{n,j}(\beta_n)}$$

$$= \frac{\hat{\mu}_{n,j}(\beta_n)\left[\sigma_{P_n,j}(\beta_n) - \hat{\sigma}_{n,j}(\beta_n)\right]}{\hat{\sigma}^M_{n,j}(\beta_n)} + \frac{(1 - \hat{\mu}_{n,j}(\beta_n))\left[\sigma_{P_n,[j]}(\beta_n) - \hat{\sigma}_{n,[j]}(\beta_n) + o_{P_n}(1)\right]}{\hat{\sigma}^M_{n,j}(\beta_n)}, \tag{A.225}$$

where the second equality follows from the definition of $\hat{\sigma}^M_{n,j}(\beta_n)$, and the third equality follows from (A.224) and $\sigma_{P_n,[j]}$ bounded away from 0 due to Assumption 28 or 73. 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,

$$\frac{(1 - \hat{\mu}_{n,j}(\beta_n))\left[\sigma_{P_n,[j]}(\beta_n) - \hat{\sigma}_{n,[j]}(\beta_n) + o_{P_n}(1)\right]}{\hat{\sigma}^M_{n,j}(\beta_n)} \tag{A.226}$$

$$= (1 - \hat{\mu}_{n,j}(\beta_n))\frac{\hat{\sigma}_{n,[j]}(\beta_n)}{\hat{\sigma}^M_{n,j}(\beta_n)}\left(\frac{\sigma_{P_n,[j]}(\beta_n)}{\hat{\sigma}_{n,[j]}(\beta_n)} - 1 + o_{P_n}(1)\right) = o_{P_n}(1).$$

By (A.225)-(A.226), it follows that $\sigma_{P_n,j}(\beta_n)/\hat{\sigma}^M_{n,j}(\beta_n) - 1 = o_{P_n}(1)$. □

Lemma 84. $\mathbb{Z}_{n\beta}(\Delta) \xrightarrow{d} \mathbb{Z}$.

Lemma 85. *Suppose Assumptions 26, 27, 25, 72 hold. In addition, suppose Assumption 28 or 73 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* $j = 0, 1, \ldots J$, *let* $\chi_j(\{Q_n, \beta_n\})$ *be a function of the sequence* $\{Q_n, \beta_n\}$ *defined as*

$$\chi_j(\{Q_n, \beta_n\}) = \begin{cases} 0, & \text{if } \lim_{n\to\infty} \kappa_n^{-1}\sqrt{n}\gamma_{1,Q_n,j}(\beta_n) = 0 \\ -\infty, & \text{if } \lim_{n\to\infty} \kappa_n^{-1}\sqrt{n}\gamma_{1,Q_n,j}(\beta_n) < 0 \end{cases} \tag{A.227}$$

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. $\frac{\sqrt{n}\bar{m}_{n,j}(\beta_n',\hat{\varphi})}{\kappa_n\hat{\sigma}_{n,j}(\beta_n')} - \frac{\sqrt{n}E_{Q_n}\left[m_j\left(X_i,\beta_n',\varphi_{Q_n}\right)\right]}{\kappa_n\sigma_{P_n,j}(\beta_n')} = o_P(1)$.

Proof. For (i), the mean value theorem yields

$$\begin{aligned}
&\sup_{P\in\mathcal{P}}\sup_{\beta\in\mathcal{B}(P),\tilde{\beta}\in\beta+\frac{\rho}{\sqrt{n}}\boldsymbol{\Delta}}\left|\frac{\sqrt{n}E_P\left[m_j(X;\beta,\varphi_P)\right]}{\kappa_n\sigma_{P,j}(\beta)} - \frac{\sqrt{n}E_P\left[m_j(X;\tilde{\beta},\varphi_P)\right]}{\kappa_n\sigma_{P,j}(\tilde{\beta})}\right| \\
\leqslant &\sup_{P\in\mathcal{P}}\sup_{\beta\in\mathcal{B}(P),\tilde{\beta}\in\beta+\frac{\rho}{\sqrt{n}}\boldsymbol{\Delta}}\left|\frac{\sqrt{n}\left(E_P\left[m_j(X;\beta,\varphi_P)\right] - E_P\left[m_j(X;\tilde{\beta},\varphi_P)\right]\right)}{\kappa_n\sigma_{P,j}(\beta)}\right| \\
&+\sup_{P\in\mathcal{P}}\sup_{\beta\in\mathcal{B}(P),\tilde{\beta}\in\beta+\frac{\rho}{\sqrt{n}}\boldsymbol{\Delta}}\left|E\left[m_j(X;\tilde{\beta},\varphi_P)\right]\frac{\sqrt{n}\left(\sigma_{P,j}(\beta)-\sigma_{P,j}(\tilde{\beta})\right)}{\kappa_n\sigma_{P,j}(\beta)\sigma_{P,j}(\tilde{\beta})}\right| \\
\leqslant &\sup_{P\in\mathcal{P}}\sup_{\beta\in\mathcal{B}(P),\tilde{\beta}\in\beta+\frac{\rho}{\sqrt{n}}\boldsymbol{\Delta}}\frac{\left\|D_{P,j}(\bar{\beta})\right\|\sqrt{n}\left\|\tilde{\beta}-\beta\right\|}{\kappa_n} + \sup_{P\in\mathcal{P}}\sup_{\beta\in\mathcal{B}(P),\tilde{\beta}\in\beta+\frac{\rho}{\sqrt{n}}\boldsymbol{\Delta}}\frac{M^2\sqrt{n}\left\|\tilde{\beta}-\beta\right\|}{\kappa_n\sigma_{P,j}(\beta)\sigma_{P,j}(\tilde{\beta})} \\
= &o(1),
\end{aligned}$$

where $\bar{\beta}$ represents a mean value that lies componentwise between β and $\tilde{\beta}$ and where we used the fact that $\sup_{P\in\mathcal{P}}\sup_{\beta\in\mathcal{B}(P)}\|D_{P,j}(\beta)\| \leqslant \bar{M}$, $\sqrt{n}\left\|\tilde{\beta}-\beta\right\| \leqslant \rho$ 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 (A.227).

For (iii), note that by (A.93), we have

$$\begin{aligned}
&\sup_{\tilde{\beta}\in\beta_n+\rho/\sqrt{n}\boldsymbol{\Delta}}\left|\kappa_n^{-1}\frac{\sqrt{n}\bar{m}_{n,j}\left(\tilde{\beta},\hat{\varphi}\right)}{\hat{\sigma}_{n,j}\left(\tilde{\beta}\right)} - \kappa_n^{-1}\frac{\sqrt{n}E_{Q_n}\left[m_j\left(X_i,\tilde{\beta},\varphi_{Q_n}\right)\right]}{\sigma_{Q_n,j}\left(\tilde{\beta}\right)}\right| \\
\leqslant &\sup_{\tilde{\beta}\in\beta_n+\rho/\sqrt{n}\boldsymbol{\Delta}}\frac{\sqrt{n}\gamma_{1,P_n,j}(\beta_n)}{\kappa_n}\hat{\eta}_{n,j,\beta_n}(\Delta) \\
&+\left(1+\hat{\eta}_{n,j,\beta_n}(\Delta)\right)\left(\frac{\mathbb{Z}_{n,j,\beta_n}}{\kappa_n} + \frac{\sigma_{Q_n,j}\left(\bar{\beta}_n\right)}{\sigma_{Q_n,j}(\beta_n)}\frac{D_{P_n,j}\left(\bar{\beta}_n\right)\Delta\rho}{\kappa_n}\right)
\end{aligned}$$

$$= o_P(1),$$

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 83(i).

□

Appendix B. Additional Results for Chapter 3

B.1 Proofs

B.1.1 Notation

Table B.1: Important notation

Θ	the parameter space
$\Theta_I(F)$	the identified set
$\Theta(r)$	$\{\theta \in \Theta : c'\theta = r\}$
$\Theta_I(r, F)$	$\Theta(r) \cap \Theta_I(F)$
$\mathcal{P}$	the baseline distribution space
$\mathcal{L}$	the parameter space for (r, F),
$\mathcal{L}_0$	$\{(r, F) : F \in \mathcal{P}, r \in \{c'\theta : \theta \in \Theta_I(F)\}\}$
$\mathbb{R}_{[\pm\infty]}$	$[-\infty, +\infty]$
$\Sigma_F(\theta_1, \theta_2)$	$E_F\left[\Big(m(W,\theta_1) - E_F\big[m(W,\theta_1)\big]\Big)\Big(m(W,\theta_2) - E_F\big[m(W,\theta_2)\big]\Big)'\right]$
$D_F(\theta)$	$\text{diag}\big(\Sigma_F(\theta,\theta)\big)$
$\Omega_F(\theta_1, \theta_2)$	$D_F(\theta_1)^{-\frac{1}{2}}\Sigma_F(\theta_1,\theta_2)D_F(\theta_2)^{-\frac{1}{2}}$
Ψ	$\{\Omega_F(\theta_1, \theta_2) : \theta_1, \theta_2 \in \Theta\}$
Ψ_1	$\{\Omega_F(\theta, \theta) : \theta \in \Theta\}$
$G_F(\theta)$	$\nabla_{\theta'}\Big(E_F\big[D_F(\theta)^{-\frac{1}{2}}m(W,\theta)\big]\Big)$
$\nu_n(\theta)$	$D_F(\theta)^{-\frac{1}{2}}\sqrt{n}\Big(\bar{m}_n(\theta) - E_F\big(m(W,\theta)\big)\Big)$
$h_{n,F}(\theta)$	$\sqrt{n}D_F(\theta)^{-\frac{1}{2}}E_F\left[m(W,\theta)\right]$
$\Gamma_{n,F}(r)$	$\Big\{(\theta,\ell) \in \Theta(r) \times \mathbb{R}^k : \ell = \sqrt{n}D_F^{-1/2}(\theta)E_F\left[m(W,\theta)\right]\Big\}$

Let $\ell^\infty(\Theta)$ be the set of all uniformly bounded functions that map $\Theta \to \mathbb{R}^u$, $u \in \mathbb{N}$, equipped with the supremum norm. The sequence of probability spaces is denoted as $\{(\mathcal{W}, \mathcal{A}, F_n)\}_{n\geqslant 1}$. Stochastic processes W maps $\mathcal{W} \to \ell^\infty(\Theta)$. In this context, I use $\xrightarrow{d}$, $\xrightarrow{p}$, and $\xrightarrow{\text{a.s.}}$ to denote weak convergence, convergence in probability, and convergence almost surely in the $\ell^\infty(\Theta)$ metric, respectively, in the sense of Van Der Vaart and Wellner (1996). For any $u \in \mathbb{N}$, for any $x_1, x_2 \in \mathbb{R}^u_{[\pm\infty]}$, we define matric

$$\tilde{d}(x_1, x_2) = \left(\sum_{i=1}^{u}(\vartheta(x_{1,i}) - \vartheta(x_{2,i}))^2\right)^{1/2},$$

where $\vartheta : \mathbb{R}_{[\pm\infty]} \to [0, 1]$ is such that $\vartheta(-\infty) = 0$, $\vartheta(\infty) = 1$, and $\vartheta(y) = \Phi(y)$ for $y \in \mathbb{R}$, where Φ is the standard normal CDF. Let d_H denote the Hausdorff metric associated with

$\tilde{d}$. Specifically, for any sets $A, B \in \Theta \times \mathbb{R}^k_{[\pm\infty]}$,

$$d_H(A,B) = \max\left\{ \sup_{(\theta_1,h_1)\in A} \inf_{(\theta_2,h_2)\in B} \tilde{d}\left((\theta_1,h_1),(\theta_2,h_2)\right), \right. \tag{B.1}$$

$$\left. \sup_{(\theta_2,h_2)\in B} \inf_{(\theta_1,h_1)\in A} \tilde{d}\left((\theta_1,h_1),(\theta_2,h_2)\right) \right\}.$$

I use $\xrightarrow{H}$ to denote convergence in the Hausdorff metric, that is, $A_n \xrightarrow{H} B \iff d_H(A_n, B) \to 0$. Let $\mathcal{C}(\Theta^2)$ denote the space of continuous functions that map Θ^2 to Ψ, and $\mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})$ denote the space of compact subsets of the metric space $(\Theta \times \mathbb{R}^k_{[\pm\infty]}, \tilde{d})$. For nonstochastic functions of $\theta \in \Theta$, I use $\xrightarrow{u}$ to denote uniform in θ convergence, for example, $\Omega_{F_n} \xrightarrow{u} \Omega \iff \sup_{\theta_1,\theta_2\in\Theta} \tilde{d}\big(\Omega_{F_n}(\theta_1,\theta_2), \Omega(\theta_1,\theta_2)\big) \to 0$. Let d denote the Euclidean distance.

I use $\Omega(\theta)$ and $\Omega(\theta,\theta)$ equivalently. And for a set Θ_1 and a scalar ϵ, the ϵ-expansion of $\Theta_1 \subset \Theta$ is defined as $\Theta_1^\varepsilon := \{\theta \in \Theta : d(\theta, \Theta_1) \leqslant \varepsilon\}$. For a vector a, $\|a\| = \sqrt{a'a}$ and for a matrix A, $\|A\| = \|\text{vec}(A)\|$. For a scalar x, $[x]_+ = \max(x, 0)$ and $[x]_- = \max(-x, 0)$. For a k dimensional vector $x = (x_1, ..., x_k)$, let $\lfloor x \rfloor_0 = \big([x_1]_+, ..., [x_p]_+, 0, ..., 0\big)$, $\lfloor x \rfloor_a = \big(-[x_1]_-, ..., -[x_p]_-, x_{p+1}, ..., x_k\big)$. For a set A, an element a and a scalar α, denote $A + a = \{\tilde{a} + a : \tilde{a} \in A\}$ and $\alpha A = \{\alpha a : a \in A\}$. The projection $T^r : \Theta \to \Theta$ is defined by

$$T^r\theta \in \underset{\tilde{\theta}\in\Theta_I(r,F)}{\text{argmin}}\ d(\tilde{\theta}, \theta).$$

For positive sequences $\{a_n\}$, $\{b_n\}$, I write $a_n \ll b_n$ iff $\frac{a_n}{b_n} \to 0$. The convergence rate of λ_n and κ_n are specified in Assumption 89.

B.1.2 Assumptions

Assumption 86. *The baseline space of distributions $\mathcal{P}$ is the set of distributions F satisfying the following properties:*

1. *The data $\{W_i\}_{i=1}^n$ are i.i.d. under F.*
2. *We have $\sigma^2_{F,j}(\theta) = Var_F(m_j(W_i, \theta)) \in (0, \infty)$ for $j = 1, ..., k$ and all $\theta \in \Theta$.*
3. *For every $F \in \mathcal{F}$ and $j = 1, \ldots, k$, $\left\{\sigma^{-1}_{F,j}(\theta) m_j(\cdot, \theta) : \mathcal{W} \to \mathbb{R}\right\}$ is a measurable class of functions indexed by $\theta \in \Theta$.*

4. *The empirical process* $\nu_n(\theta)$ *with* j*-component*

$$\nu_{n,j}(\theta) = n^{-1/2}\sigma_{F,j}^{-1}(\theta)\sum_{i=1}^{n}\big(m_j(W_i,\theta) - E_F\big[m_j(W_i,\theta)\big]\big) \tag{B.2}$$

for $j = 1,\ldots,k$ *is asymptotically* ρ_F *-equicontinuous uniformly in* $F \in \mathcal{P}$ *in the sense of Van Der Vaart and Wellner (1996) page 169. This is, for any* $\varepsilon > 0$

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\sup_{F\in\mathcal{P}} P_F^*\left(\sup_{\rho_F(\theta_1,\theta_2)<\delta}\|\nu_n(\theta_1) - \nu_n(\theta_2)\| > \varepsilon\right) = 0$$

where P_F^* *denotes outer probability and* ρ_F *denotes the coordinate-wise version of the intrinsic variance semimetric defined as*

$$\rho_F(\theta_1,\theta_2) \equiv \left\|\left\{Var_F\big[\sigma_{F,j}^{-1}(\theta_1)m_j(W,\theta_1) - \sigma_{F,j}^{-1}(\theta_2)m_j(W,\theta_2)\big]^{1/2}\right\}_{j=1}^{k}\right\|.$$

5. *For some constant* $c > 0$ *and all* $j = 1,\ldots,k$,

$$\sup_{F\in\mathcal{F}} E_F\left[\sup_{\theta\in\Theta}\left|\frac{m_j(W,\theta)}{\sigma_{F,j}(\theta)}\right|^{2+c}\right] < \infty.$$

6. *The matrix* Ω_F *satisfies*

$$\lim_{\delta\downarrow 0}\sup_{\|(\theta_1,\tilde{\theta}_1)-(\theta_2,\tilde{\theta}_2)\|<\delta}\sup_{F\in\mathcal{P}}\left\|\Omega_F(\theta_1,\tilde{\theta}_1) - \Omega_F(\theta_2,\tilde{\theta}_2)\right\| = 0.$$

Assumption 87. *For any* $\{(r_n,F_n)\in\mathcal{L}_0\}_{n\geqslant 1}$, *let* (Γ,Ω) *be such that* $\Omega_{F_n} \xrightarrow{u} \Omega$ *and* $\Gamma_{n,F_n}(r_n) \xrightarrow{H} \Gamma$ *with* $(\Gamma,\Omega)\in\mathcal{S}(\Theta\times\mathbb{R}^k_{[\pm\infty]})\times\mathcal{C}(\Theta^2)$ *and* $\Gamma_{n,F_n}(r_n)$ *as in Table B.1. Let* $cv^S(\Gamma,\Omega,1-\alpha)$ *be the* $(1-\alpha)$ *quantile of*

$$J(\Gamma,\Omega) \equiv \inf_{(\theta,\ell)\in\Gamma} S\left(\nu_\Omega(\theta) + \ell, \Omega(\theta)\right) \tag{B.3}$$

where $\nu_\Omega : \Theta \to \mathbb{R}^k$ *is a tight Gaussian process with zero-mean and covariance (correlation) kernel* Ω. *Then the following statements hold:*

1. *If* $cv^S(\Gamma,\Omega,1-\alpha) > 0$, *the distribution of* $J(\Gamma,\Omega)$ *is continuous at* $cv^S(\Gamma,\Omega,1-\alpha)$.

2. *If* $cv^S(\Gamma, \Omega, 1-\alpha) = 0, \liminf_{n\to\infty} P_{F_n}\left(J_n(r_n) = 0\right) \geqslant 1-\alpha$, *where* $J_n(r_n)$ *is as in (3.4).*

Assumption 88. *The test function* S *satisfies the following properties:*

1. $S\left((m_1, m_2), \Sigma\right)$ *is nonincreasing in* m_1 *for all* $m_1 \in \mathbb{R}^p$, $m_2 \in \mathbb{R}^{k-p}$, *and variance matrices* $\Sigma \in \mathbb{R}^{k\times k}$.
2. $S(m, \Sigma) = S(Dm, D\Sigma D)$ *for all* $m \in \mathbb{R}^k$, *variance matrices* $\Sigma \in \mathbb{R}^{k\times k}$,*and pd diagonal* $D \in \mathbb{R}^{k\times k}$.
3. $S(m, \Omega) \geqslant 0$ *for all* $m \in \mathbb{R}^k$ *and* $\Omega \in \Psi_1$.
4. $S(m, \Omega)$ *is continuous at all* $m \in \mathbb{R}^p_{[\pm\infty]} \times \mathbb{R}^{k-p}$ *and* $\Omega \in \Psi_1$.
5. *For some* $\chi > 0$. $S(cm, \Omega) = c^\chi S(m, \Omega)$ *for all scalar* $c > 0$, $m \in \mathbb{R}^k$ *and* $\Omega \in \Psi_1$.
6. *We have* $S(m, \Omega) > 0$ *if and only if* $m_j < 0$ *for some* $j = 1, \ldots, p$ *or* $m_j \neq 0$ *for some* $j = p+1, \ldots, k$, *where* $m = (m_1, \ldots, m_k)'$ *and* $\Omega \in \Psi_1$. *Equivalently,* $S(m, \Omega) = 0$ *if and only if* $m_j \geqslant 0$ *for all* $j = 1, \ldots, p$ *and* $m_j = 0$ *for all* $j = p+1, \ldots, k$, *where* $m = (m_1, \ldots, m_k)'$ *and* $\Omega \in \Psi_1$.
7. *For all* $n \geqslant 1, S\left(\sqrt{n}\bar{m}_n(\theta), \hat{\Sigma}_n(\theta)\right)$ *is a lower semicontinuous function of* $\theta \in \Theta$.

Assumption 89. *Rate of sequences*

1. $\kappa_n \to \infty$ *and* $\kappa_n = O\left(\frac{n^{1/4}}{(\ln n)^{1/\chi}}\right)$.
2. $\frac{\lambda_n}{\sqrt{n}} \to \infty$ *and* $\lambda_n(\ln \kappa_n^2)^{\frac{2}{\chi}} = O(n)$ *where* χ *is defined in Assumption 88.5.*

B.1.3 Proof of Theorem 36

B.1.3.1 Part 1

Define $cv_n\big(\breve{\Theta}, \breve{\Omega}, \breve{\varphi}, \breve{G}, 1-\alpha\big)$ with $\breve{\Theta} \subset \Theta$, $\breve{\Omega} \in \mathcal{C}(\Theta^2)$, $\breve{\varphi} : \Theta \to \mathbb{R}, \breve{G} : \Theta \to \mathbb{R}^{k\times d_\theta}$ to be the $1-\alpha$ quantile of

$$\inf_{\theta\in\breve{\Theta}} S\big(\upsilon(\theta) + \breve{\varphi}(\theta) + \breve{G}(\theta)\Delta(\theta), \breve{\Omega}(\theta)\big)$$

with υ being a Gaussian process with covariance kernel $\breve{\Omega}$, and

$$\Delta(\theta) \in \underset{\Delta\in\mathbf{\Delta}_n^{c\prime\theta}(\theta)}{\arg\min} \left(S\big(\upsilon(\theta) + \breve{\varphi}(\theta) + \breve{G}(\theta)\Delta, \breve{\Omega}(\theta)\big) + \frac{\lambda_n}{n}\Delta'\Delta \right).$$

Note that

$$cv_n^{\mathrm{LL}}(r, 1-\alpha) = cv_n\big(\hat{\Theta}_I(r), \hat{\Omega}_n, \varphi_n, \hat{G}_n, 1-\alpha\big), \tag{B.4}$$

thus we are interested in

$$\limsup_{n\to\infty} \sup_{(r,F)\in\mathcal{L}_0} P_F\left(J_n(r) > cv_n\big(\hat{\Theta}_I(r), \hat{\Omega}_n, \varphi_n, \hat{G}_n, 1-\alpha\big)\right). \tag{B.5}$$

There is always a subsequence $\{a_n\}$, $\{r_{a_n}, F_{a_n}\}$ s.t.

$$(B.5) = \lim_{n\to\infty} P_{F_{a_n}}\left(J_{a_n}(r_{a_n}) > cv_{a_n}\big(\hat{\Theta}_I(r_{a_n}), \hat{\Omega}_{a_n}, \varphi_{a_n}, \hat{G}_{a_n}, 1-\alpha\big)\right).$$

By Lemma S.3.2 in BCS17, there is a further sequence $\{u_n\}_{n\geqslant 1}$ of $\{a_n\}_{n\geqslant 1}$ s.t. $\Omega_{F_{u_n}} \xrightarrow{u} \Omega$, $\Gamma_{u_n,F_{u_n}}(r_{u_n}) \xrightarrow{H} \Gamma$, for some $(\Gamma,\Omega) \in \mathcal{S}(\Theta\times\mathbb{R}^k_{[\pm\infty]})\times\mathcal{C}(\Theta^2)$. Here $\Gamma_{n,F}(r)$ is defined in Table B.1. Then, since $\Omega_{F_{u_n}} \xrightarrow{u} \Omega$ and $\Gamma_{u_n,F_{u_n}}(r_{u_n}) \xrightarrow{H} \Gamma$, Theorem S.2.4 in BCS17 implies that

$$J_{u_n}(r_{u_n}) \xrightarrow{d} J(\Gamma,\Omega) \tag{B.6}$$

where $J(\Gamma,\Omega)$ is defined in (B.3).

B.1.3.2 Part 2

In this part, I show that for a sequence $(r_n, F_n) \in \mathcal{F}_0$ with $\Omega_{F_n} \xrightarrow{u} \Omega$, $\Gamma_{n,F_n}(r_n) \xrightarrow{H} \Gamma$, for some $(\Gamma,\Omega) \in \mathcal{S}(\Theta\times\mathbb{R}^k_{[\pm\infty]})\times\mathcal{C}(\Theta^2)$, for all $\epsilon > 0$

$$\limsup_{n\to\infty} P_{F_n}\left(cv_n\big(\hat{\Theta}_I(r_n), \hat{\Omega}_n, \varphi_n, \hat{G}_n, 1-\alpha\big) \leqslant cv^S(\Gamma,\Omega,1-\alpha) - \epsilon\right) = 0$$

where $cv^S(\Gamma,\Omega,1-\alpha)$ is the $1-\alpha$ quantile of $J(\Gamma,\Omega)$ defined in (B.3). And it also holds if we change n to a subsequence u_n.

Let $\tau_{1n} = a\varepsilon_n$ where $\varepsilon_n = n^{-1/2}\left(\ln\kappa_n^2\right)^{\frac{1}{\chi}}$ and $a > 0$ are specified in Assumption 35. By Assumption 89.2,

$$\tau_{1n} = O\left(\lambda_n^{-\frac{1}{2}}\right). \tag{B.7}$$

Let τ_{2n} satisfy $\frac{\kappa_n}{\sqrt{n}} \ll \tau_{2n} \ll \sqrt{\frac{\lambda_n}{n}}$. By BCS17 Lemma S.3.1.2, we have $\hat{\Omega}_n \xrightarrow{p} \Omega$ in $l^\infty(\Theta)$. Thus there is a sequence $\tau_{3n} = o(1)$ such that $\left\|\hat{\Omega}_n - \Omega\right\| = o_p(\tau_{3n})$. Choose τ_{4n} s.t. $\frac{(\ln \kappa_n^2)^{\frac{1}{\chi}}}{\kappa_n} \ll \tau_{4n} \ll 1$. Define a sequence of sets A_n

$$A_n = \Big\{(\breve{\Theta}, \breve{\Omega}, \breve{\varphi}, \breve{G}) : \breve{\Theta} \subseteq \Theta(r_n) \cap \Theta_I(r_n, F_n)^{\tau_{1n}}, \sup_{\theta\in\Theta} \left\|\breve{G}(\theta) - G_{F_n}(\theta)\right\| \leqslant \tau_{2n},$$

$$\sup_{\theta_1,\theta_2\in\Theta} \left\|\breve{\Omega}(\theta_1,\theta_2) - \Omega_{F_n}(\theta_1,\theta_2)\right\| \leqslant \tau_{3n}, \sup_{\theta\in\breve{\Theta}} \left\|\left\lfloor h_{F_n}(T^{r_n}\theta) - \breve{\varphi}(\theta)\right\rfloor_a\right\| \leqslant \tau_{4n}\Big\}. \tag{B.8}$$

Step 1. In this step, I consider sequences θ_n and $\tilde{\theta}_n$ such that

$$\theta_n \in \Theta(r_n) \cap \Theta_I(r_n, F_n)^{\tau_{1n}}, \tag{B.9}$$

$$\tilde{\theta}_n = T^{r_n}\theta_n + O_p(\lambda_n^{-\frac{1}{2}}), \tag{B.10}$$

and

$$\left\lfloor \nu_{F_n}(\theta_n) + h_{n,F_n}(T^{r_n}\theta_n) + G_{F_n}(\theta_n)\sqrt{n}\big(\tilde{\theta}_n - T^{r_n}\theta_n\big)\right\rfloor_a = O_p(1). \tag{B.11}$$

I show

$$S\big(\hat{D}_n(\tilde{\theta}_n)^{-\frac{1}{2}}\sqrt{n}\bar{m}_n(\tilde{\theta}_n), \hat{\Omega}_n(\tilde{\theta}_n)\big)$$

$$= S\Big(\nu_{F_n}(\theta_n) + h_{n,F_n}(T^{r_n}\theta_n) + G_{F_n}(\theta_n)\sqrt{n}\big(\tilde{\theta}_n - T^{r_n}\theta_n\big), \Omega_{F_n}(\theta_n)\Big) + o_p(1). \tag{B.12}$$

First, simplify $D_{F_n}(\tilde{\theta}_n)^{-\frac{1}{2}}\sqrt{n}\bar{m}_n(\tilde{\theta}_n)$, Note that

$$\begin{aligned}
&D_{F_n}(\tilde{\theta}_n)^{-\frac{1}{2}}\sqrt{n}\bar{m}_n(\tilde{\theta}_n)\\
=&D_{F_n}(\tilde{\theta}_n)^{-\frac{1}{2}}\sqrt{n}\Big(\bar{m}_n(\tilde{\theta}_n) - E_{F_n}\big(m(W,\tilde{\theta}_n)\big)\Big) + \sqrt{n}D_{F_n}(\tilde{\theta}_n)^{-\frac{1}{2}}E_{F_n}\big(m(W,\tilde{\theta}_n)\big)\\
=&\nu_{F_n}(\tilde{\theta}_n) + \sqrt{n}D_{F_n}(T^{r_n}\theta_n)^{-\frac{1}{2}}E_{F_n}\big(m(W,T^{r_n}\theta_n)\big) + G_{F_n}(\bar{\theta}_n)\sqrt{n}(\tilde{\theta}_n - T^{r_n}\theta_n)\\
=&\nu_{F_n}(\tilde{\theta}_n) + h_{n,F_n}(T^{r_n}\theta_n) + G_{F_n}(\theta_n)\sqrt{n}\big(\tilde{\theta}_n - T^{r_n}\theta_n\big)\\
&+ \big(G_{F_n}(\bar{\theta}_n) - G_{F_n}(\theta_n)\big)\sqrt{n}\big(\tilde{\theta}_n - T^{r_n}\theta_n\big)\\
=&\nu_{F_n}(\theta_n) + h_{n,F_n}(T^{r_n}\theta_n) + G_{F_n}(\theta_n)\sqrt{n}\big(\tilde{\theta}_n - T^{r_n}\theta_n\big) + o_p(1).
\end{aligned} \tag{B.13}$$

The first and third equalities are just rearranging terms; the second one is by the mean value theorem where $\bar{\theta}_n$ is between $\tilde{\theta}_n$ and $T^{r_n}\theta_n$. Note that the Taylor expansion is for

k rows and $\bar{\theta}_n$ may be different for each row, but I use $\bar{\theta}_n$ for simplicity. To get the last equality, first, note that

$$\begin{aligned}\left\|\tilde{\theta}_n - \theta_n\right\| &\leqslant \left\|\tilde{\theta}_n - T^{r_n}\theta_n\right\| + \|T^{r_n}\theta_n - \theta_n\| \\ &\leqslant O_p(\lambda_n^{-1/2}) + \tau_{1n} \\ &= O_p(\lambda_n^{-1/2}) \end{aligned} \tag{B.14}$$

where the second line follows from (B.9) and (B.10), and the third one follows from (B.7). And thus by Assumption 86.4

$$\nu_{F_n}(\theta_n) = \nu_{F_n}(\tilde{\theta}_n) + o_p(1). \tag{B.15}$$

And second, by Assumption 34.2, it holds that

$$\begin{aligned}\left\|G_{F_n,j}(\bar{\theta}_n) - G_{F_n,j}(\theta_n)\right\| &\leqslant M_G\left\|\bar{\theta}_n - \theta_n\right\| \\ &\leqslant M_G\left\|\tilde{\theta}_n - \theta_n\right\| \\ &= O_p(\lambda_n^{-\frac{1}{2}}). \end{aligned} \tag{B.16}$$

Combine (B.14), (B.16) and Assumption 89.2, and we get

$$\big(G_{F_n}(\bar{\theta}_n) - G_{F_n}(\theta_n)\big)\sqrt{n}\big(\tilde{\theta}_n - T^{r_n}\theta_n\big) = O_p(\lambda_n^{-\frac{1}{2}})\sqrt{n}O_p(\lambda_n^{-\frac{1}{2}}) = o_p(1). \tag{B.17}$$

(B.15) and (B.17) imply (B.13).

Thus, for all $\tilde{\theta}_n = T^{r_n}\theta_n + O_p(\lambda_n^{-\frac{1}{2}})$, it holds that

$$\begin{aligned}&S\big(\hat{D}_n(\tilde{\theta}_n)^{-\frac{1}{2}}\sqrt{n}\bar{m}_n(\tilde{\theta}_n), \hat{\Omega}_n(\tilde{\theta}_n)\big) \\ =&S\Big(D_{F_n}(\tilde{\theta}_n)^{-\frac{1}{2}}\sqrt{n}\bar{m}_n(\tilde{\theta}_n), D_{F_n}(\tilde{\theta}_n)^{-\frac{1}{2}}\hat{D}_n(\tilde{\theta}_n)^{\frac{1}{2}}\hat{\Omega}_n(\tilde{\theta}_n)\hat{D}_n(\tilde{\theta}_n)^{\frac{1}{2}}D_{F_n}(\tilde{\theta}_n)^{-\frac{1}{2}}\Big) \\ =&S\Big(\nu_{F_n}(\theta_n) + h_{n,F_n}(T^{r_n}\theta_n) + G_{F_n}(\theta_n)\sqrt{n}\big(\tilde{\theta}_n - T^{r_n}\theta_n\big) + o_p(1), \Omega_{F_n}(\theta_n) + o_p(1)\Big) \end{aligned} \tag{B.18}$$

where the first line follows from Assumption 88.2, and the second line from (B.13), and

$$D_{F_n}(\theta_n)^{-\frac{1}{2}}\hat{D}_n(\theta_n)^{\frac{1}{2}} = I_k + o_p(1)$$

$$\hat{\Omega}_n(\theta_n) = \hat{\Omega}_n(\theta_n) - \Omega_{F_n}(\theta_n) + \Omega_{F_n}(\theta_n) = \Omega_{F_n}(\theta_n) + o_p(1)$$

by BCS17 Lemma S.3.1.3 and BCS17 Lemma S.3.1.2, respectively.

Lastly, by (B.11) and Lemma 90, we have

$$(B.18) = S\Big(\nu_{F_n}(\theta_n) + h_{n,F_n}(T^{r_n}\theta_n) + G_{F_n}(\theta_n)\sqrt{n}\big(\tilde{\theta}_n - T^{r_n}\theta_n\big), \Omega_{F_n}(\theta_n)\Big) + o_p(1),$$

which completes the proof of this step.

Step 2. Then I show that for $\{\breve{\Theta}_n, \breve{\Omega}_n, \breve{\varphi}_n, \breve{G}_n\} \in A_n$ with A_n defined in (B.8),

$$\liminf_{n\to\infty} cv_n\big(\breve{\Theta}_n, \breve{\Omega}_n, \breve{\varphi}_n, \breve{G}_n, 1-\alpha\big) > cv^S(\Gamma, \Omega, 1-\alpha) - \epsilon. \tag{B.19}$$

Recall that

$$\nu_n(\theta) = D_{F_n}(\theta)^{-\frac{1}{2}}\sqrt{n}\Big(\bar{m}_n(\theta) - E_{F_n}\big(m(W,\theta)\big)\Big)$$

and by Lemma D.2.1 in BCS15, we have $\nu_n \xrightarrow{d} \nu$ where ν is a Gaussian process with covariance kernel Ω. Define υ_n as a Gaussian process with covariance kernel $\breve{\Omega}_n$, thus we also have $\upsilon_n \xrightarrow{d} \nu$ on $\ell^\infty(\Theta)$. By Lemma 91, there is $\tilde{\upsilon}_n \sim \upsilon_n$ s.t. $\nu_n - \tilde{\upsilon}_n \xrightarrow{p} 0$ on $\ell^\infty(\Theta)$. Define

$$\tilde{\Delta}_n(\theta) \in \underset{\Delta\in\mathbf{\Delta}_n^{r_n}(\theta)}{\operatorname{argmin}} S\left(\tilde{\upsilon}_n(\theta) + \breve{\varphi}_n(\theta) + \breve{G}_n(\theta)\Delta, \breve{\Omega}_n(\theta)\right) + \frac{\lambda_n}{n}\Delta'\Delta. \tag{B.20}$$

Since $\tilde{\Delta}_n(\theta) \in \mathbf{\Delta}_n^{r_n}(\theta)$ and by construction $\breve{\Theta}_n \subset \Theta(r_n) \cap \Theta_I(r_n, F_n)^{\tau_{1n}}$, Assumption 35.2 implies that for all $\theta \in \breve{\Theta}_n$

$$\tilde{\Delta}_n(\theta) \in \sqrt{n}\left(\Theta(r_n) - T^{r_n}\theta\right)$$

with probability approaching 1. This is equivalent to

$$\Big\{T^{r_n}\theta + \frac{1}{\sqrt{n}}\tilde{\Delta}_n(\theta) : \theta \in \breve{\Theta}_n\Big\} \subset \Theta(r_n) \tag{B.21}$$

with probability approaching 1. Then we have

$$J_n(r_n) = \inf_{\theta\in\Theta(r_n)} S\left(\hat{D}_n(\theta)^{-\frac{1}{2}}\sqrt{n}\bar{m}_n(\theta), \hat{\Omega}_n(\theta)\right)$$

$$\leqslant \inf_{\theta\in\breve{\Theta}_n} S\Big(\hat{D}_n\big(T^{r_n}\theta + \frac{1}{\sqrt{n}}\tilde{\Delta}_n(\theta)\big)^{-\frac{1}{2}}\sqrt{n}\bar{m}_n\big(T^{r_n}\theta + \frac{1}{\sqrt{n}}\tilde{\Delta}_n(\theta)\big), \hat{\Omega}_n\big(T^{r_n}\theta + \frac{1}{\sqrt{n}}\tilde{\Delta}_n(\theta)\big)\Big)$$

$$+ o_p(1)$$

$$= \inf_{\theta \in \breve{\Theta}_n} S\Big(\nu_n(\theta) + h_{n,F_n}(T^{r_n}\theta) + G_{F_n}(\theta)\tilde{\Delta}_n(\theta), \Omega_{F_n}(\theta)\Big) + o_p(1) \tag{B.22}$$

the first inequality follows from (B.21). Then I show (B.22). First note that $0 \in \mathbf{\Delta}_n^{r_n}(\theta)$, thus (B.20) implies

$$S\left(\tilde{v}_n(\theta) + \breve{\varphi}_n(\theta) + \breve{G}_n(\theta)\tilde{\Delta}_n(\theta), \breve{\Omega}_n(\theta)\right) + \frac{\lambda_n}{n}\left\|\tilde{\Delta}_n(\theta)\right\|^2 \leqslant S\left(\tilde{v}_n(\theta) + \breve{\varphi}_n(\theta), \breve{\Omega}_n(\theta)\right) = O_p(1).$$

Therefore, on $\ell^\infty(\breve{\Theta}_n)$

$$\tilde{\Delta}_n(\theta) = O_p(\sqrt{\frac{n}{\lambda_n}}), \tag{B.23}$$

$$\left\lfloor \tilde{v}_n(\theta) + \breve{\varphi}_n(\theta) + \breve{G}_n(\theta)\tilde{\Delta}_n(\theta) \right\rfloor_a = O_p(1). \tag{B.24}$$

Second, on $\ell^\infty(\breve{\Theta}_n)$

$$\begin{aligned}
&\left\| \left\lfloor \nu_n(\theta) + h_{n,F_n}(T^{r_n}\theta) + G_{F_n}(\theta)\tilde{\Delta}_n(\theta) \right\rfloor_a \right\| \\
\leqslant &\left\| \left\lfloor \tilde{v}_n(\theta) + \breve{\varphi}_n(\theta) + \breve{G}_n(\theta)\tilde{\Delta}_n(\theta) \right\rfloor_a \right. \\
&\left. + \left\lfloor (\nu_n(\theta) - \tilde{v}_n(\theta)) + \left(G_{F_n}(\theta) - \breve{G}_n(\theta)\right)\tilde{\Delta}_n(\theta) \right\rfloor_a + \lfloor h_{n,F_n}(T^{r_n}\theta) - \breve{\varphi}_n(\theta) \rfloor_a \right\| \qquad \text{(B.25)} \\
\leqslant &\left\| \left\lfloor \tilde{v}_n(\theta) + \breve{\varphi}_n(\theta) + \breve{G}_n(\theta)\tilde{\Delta}_n(\theta) \right\rfloor_a \right\| \\
&+ \left\| o_p(1) + \tau_{2n} O_p\left(\sqrt{\frac{n}{\lambda_n}}\right) + o_p(\tau_{4n}) \right\| \qquad \text{(B.26)} \\
= &\left\| \left\lfloor \tilde{v}_n(\theta) + \breve{\varphi}_n(\theta) + \breve{G}_n(\theta)\tilde{\Delta}_n(\theta) \right\rfloor_a \right\| + o_p(1). \qquad \text{(B.27)}
\end{aligned}$$

where (B.25) follows from $\|\lfloor x_1 + x_2 \rfloor_a\| \leqslant \|\lfloor x_1 \rfloor_a + \lfloor x_2 \rfloor_a\|$, and (B.26) follows from (i)

$$\nu_n(\theta) - \tilde{v}_n(\theta) = o_p(1) \tag{B.28}$$

by the construction of $\tilde{v}_n$, and (ii)

$$G_{F_n}(\theta) - \breve{G}_n(\theta) = O(\tau_{2n}) \tag{B.29}$$

$$\lfloor h_{n,F_n}(T^{r_n}\theta) - \breve{\varphi}_n(\theta)\rfloor_a = O(\tau_{4n}) \tag{B.30}$$

by the construction of A_n. Therefore, by (B.24) and (B.27), we have that on $\ell^\infty(\breve{\Theta}_n)$

$$\Big\lfloor \nu_n(\theta) + h_{n,F_n}(T^{r_n}\theta) + G_{F_n}(\theta)\tilde{\Delta}_n(\theta)\Big\rfloor_a = O_p(1). \tag{B.31}$$

In sum, (B.23), (B.31) and the conclusion in step 1, (B.12), imply (B.22).

Moreover,

$$\begin{aligned}
&\inf_{\theta\in\breve{\Theta}_n} S\big(\nu_n(\theta) + h_{n,F_n}(T^{r_n}\theta) + G_{F_n}(\theta)\tilde{\Delta}_n(\theta), \Omega_{F_n}(\theta)\big) \\
&\qquad - \inf_{\theta\in\breve{\Theta}_n} S\big(\tilde{v}_n(\theta) + \breve{\varphi}_n(\theta) + \breve{G}_n(\theta)\tilde{\Delta}_n(\theta), \breve{\Omega}_n(\theta)\big) \\
\leqslant &\inf_{\theta\in\breve{\Theta}_n} S\big(\tilde{v}_n(\theta) + \breve{\varphi}_n(\theta) + \breve{G}_n(\theta)\tilde{\Delta}_n(\theta) + \lfloor h_{n,F_n}(T^{r_n}\theta) - \breve{\varphi}_n(\theta)\rfloor_a + o_p(1), \Omega_{F_n}(\theta)\big) \\
&\qquad - \inf_{\theta\in\breve{\Theta}_n} S\big(\tilde{v}_n(\theta) + \breve{\varphi}_n(\theta) + \breve{G}_n(\theta)\tilde{\Delta}_n(\theta), \breve{\Omega}_n(\theta)\big) \\
\leqslant &\inf_{\theta\in\breve{\Theta}_n} S\Big(\tilde{v}_n(\theta) + \breve{\varphi}_n(\theta) + \breve{G}_n(\theta)\tilde{\Delta}_n(\theta) + o_p(1), \breve{\Omega}_n(\theta) + o_p(1)\Big) \\
&\qquad - \inf_{\theta\in\breve{\Theta}_n} S\Big(\tilde{v}_n(\theta) + \breve{\varphi}_n(\theta) + \breve{G}_n(\theta)\tilde{\Delta}_n(\theta), \breve{\Omega}_n(\theta)\Big) \\
\leqslant &\sup_{\theta\in\breve{\Theta}_n}\Big| S\Big(\tilde{v}_n(\theta) + \breve{\varphi}_n(\theta) + G_{F_n}(\theta)\tilde{\Delta}_n(\theta) + o_p(1), \breve{\Omega}_n(\theta) + o_p(1)\Big) \\
&\qquad - S\Big(\tilde{v}_n(\theta) + \breve{\varphi}_n(\theta) + \breve{G}_n(\theta)\tilde{\Delta}_n(\theta), \breve{\Omega}_n(\theta)\Big)\Big| = o_p(1)
\end{aligned} \tag{B.32}$$

where the first inequality follows from (B.28), (B.29) and the second one from (B.30). The last equality holds by (B.24) and Lemma 90. Therefore by (B.22) and (B.32), we have

$$J_n(r_n) \leqslant \inf_{\theta\in\breve{\Theta}_n} S\big(\tilde{v}_n(\theta) + \breve{\varphi}_n(\theta) + \breve{G}_n(\theta)\tilde{\Delta}_n(\theta), \breve{\Omega}_n(\theta)\big) + o_p(1). \tag{B.33}$$

Then taking limsup on both sides of the following inequality

$$\begin{aligned}
&P_{F_n}\Big(\inf_{\theta\in\breve{\Theta}_n} S\big(\tilde{v}_n(\theta) + \breve{\varphi}_n(\theta) + \breve{G}_n(\theta)\tilde{\Delta}_n(\theta), \breve{\Omega}_n(\theta)\big) \leqslant cv^S(\Gamma, \Omega, 1-\alpha) - \epsilon\Big) \\
\leqslant &P_{F_n}\Big(\inf_{\theta\in\breve{\Theta}_n} S\big(\tilde{v}_n(\theta) + \breve{\varphi}_n(\theta) + \breve{G}_n(\theta)\tilde{\Delta}_n(\theta), \breve{\Omega}_n(\theta)\big) - J_n(r_n) \leqslant -\frac{\epsilon}{2}\Big)
\end{aligned}$$

$$+ P_{F_n}\Big(J_n(r_n) \leqslant cv^S(\Gamma,\Omega,1-\alpha) - \frac{\epsilon}{2}\Big)$$

and using (B.33) and $J_n(r_n) \xrightarrow{d} J(\Gamma,\Omega)$, we get

$$\limsup_{n\to\infty} P_{F_n}\Big(\inf_{\theta\in\breve{\Theta}_n} S\big(\tilde{\upsilon}_n(\theta) + \breve{\varphi}_n(\theta) + \breve{G}_n(\theta)\tilde{\Delta}_n(\theta), \breve{\Omega}_n(\theta)\big) \leqslant cv^S(\Gamma,\Omega,1-\alpha) - \epsilon\Big)$$

$$\leqslant P\big(J(\Gamma,\Omega) \leqslant cv^S(\Gamma,\Omega,1-\alpha) - \frac{\epsilon}{2}\big) < 1-\alpha. \tag{B.34}$$

Moreover, $cv_n(\breve{\Theta}_n, \breve{\Omega}_n, \breve{\varphi}_n, \breve{G}_n, 1-\alpha)$ is the $1-\alpha$ quantile of $\inf_{\theta\in\breve{\Theta}_n} S\big(\tilde{\upsilon}_n + \breve{\varphi}_n + \breve{G}_n\tilde{\Delta}_n, \breve{\Omega}_n\big)$, thus (B.34) implies

$$\liminf_{n\to\infty} cv_n(\breve{\Theta}_n, \breve{\Omega}_n, \breve{\varphi}_n, \breve{G}_n, 1-\alpha) > cv^S(\Gamma,\Omega,1-\alpha) - \epsilon.$$

Step 3. I next show that

$$\liminf_{n\to\infty} P_{F_n}\Big(cv_n\big(\hat{\Theta}_I(r_n), \hat{\Omega}_n, \varphi_n, \hat{G}_n, 1-\alpha\big) > cv^S(\Gamma,\Omega,1-\alpha) - \epsilon\Big) = 1. \tag{B.35}$$

We can show that

$$\liminf_{n\to\infty} P_{F_n}(\mathcal{W}_n^*) = 1. \tag{B.36}$$

where

$$\mathcal{W}_n^* = \Big\{\big(\hat{\Theta}_I(r_n), \hat{\Omega}_n, \varphi_n, \hat{G}_n\big) \subset A_n\Big\}$$

and A_n is defined in (B.8). (i) Let

$$\Theta_n \equiv \left\{\theta\in\Theta : S\left(\sqrt{n}E_F[m(W,\theta)], \Sigma_F(\theta)\right) \leqslant a^\chi \ln\kappa_n^2\right\}. \tag{B.37}$$

Then by Lemma D.13 in BCS15, we have

$$\liminf_{n\to\infty} \inf_{(r,F)\in\mathcal{L}_0} P_F\Big(\hat{\theta}_{n,r} \subset \Theta_n \cap \Theta(r)\Big) = 1. \tag{B.38}$$

And by Assumption 33, Assumption 88.5, for all $\theta\in\Theta_n\cap\Theta(r)$, it holds that

$$d(\theta, \Theta_I(r,F)) \leqslant a\frac{(\ln\kappa_n^2)^{1/\chi}}{\sqrt{n}} = \tau_{1n} \tag{B.39}$$

when n is large enough. Thus we can rewrite (B.38) as

$$\liminf_{n\to\infty} \inf_{(r,F)\in\mathcal{L}_0} P_F\Big(d(\hat{\theta}_{n,r}, \Theta_I(r,F)) \leqslant \tau_{1n}\Big) = 1. \tag{B.40}$$

(ii) by Assumption 34.1 and BCS17 Lemma S.3.1.2,

$$\liminf_{n\to\infty} P_{F_n}\bigg(\sup_{\theta\in\hat{\Theta}_I(r_n)} \big\|\hat{G}_n(\theta) - G_{F_n}(\theta)\big\| \leqslant \tau_{2n}, \sup_{\theta\in\hat{\Theta}_I(r_n)} \big\|\hat{\Omega}_n(\theta) - \Omega_{F_n}(\theta)\big\| \leqslant \tau_{3n}\bigg) = 1$$

(iii) Note that for $\theta_n \in \hat{\Theta}_I(r_n)$

$$\begin{aligned}
\varphi_n(\theta_n) =& \Big\lfloor \hat{D}_n(\theta_n)^{-\frac{1}{2}} D_{F_n}(\theta_n)^{\frac{1}{2}} D_{F_n}(\theta_n)^{-\frac{1}{2}} \sqrt{n}\bar{m}_n(\theta_n)/\kappa_n \Big\rfloor_0 \\
=& \Big\lfloor \big(I_k + o_p(1)\big)\big(\nu_{F_n}(\theta_n) + h_{n,F_n}(T^{r_n}\theta_n) \\
&\qquad + G_{F_n}(\theta_n)\sqrt{n}(\theta_n - T^{r_n}\theta_n) + o_p(1)\big)/\kappa_n \Big\rfloor_0 \\
=& \Big\lfloor \big(I_k + o_p(1)\big) h_{n,F_n}(T^{r_n}\theta_n)/\kappa_n + O_p\left((\ln\kappa_n^2)^{1/\chi}/\kappa_n\right) \Big\rfloor_0 \\
\leqslant& h_{F_n}(T^{r_n}\theta_n) + O_p\left((\ln\kappa_n^2)^{1/\chi}/\kappa_n\right)
\end{aligned} \tag{B.41}$$

where the first line is the definition of φ_n, the second line follows from (B.13), the third line from $d(\theta_n, T^{r_n}\theta_n) = o(\tau_{1n})$ and Assumption 34.2, the last line follows from $\kappa_n \to +\infty$. Thus

$$\liminf_{n\to\infty} P_{F_n}\bigg(\inf_{\theta\in\hat{\Theta}_I(r_n)} \Big\|\big\lfloor h_{F_n}(T^{r_n}\theta) - \varphi_n(\theta)\big\rfloor_a\Big\| \leqslant \tau_{4n}\bigg) = 1.$$

By (i), (ii), (iii) above, we complete the proof of (B.36).

Next, note that

$$\begin{aligned}
&\liminf_{n\to\infty} P_{F_n}\Big(cv_n\big(\hat{\Theta}_I(r_n), \hat{\Omega}_n, \varphi_n, \hat{G}_n, 1-\alpha\big) > cv^S(\Gamma, \Omega, 1-\alpha) - \epsilon\Big) \\
\geqslant& \liminf_{n\to\infty} P_{F_n}\left(cv_n\big(\hat{\Theta}_I(r_n), \hat{\Omega}_n, \varphi_n, \hat{G}_n, 1-\alpha\big) > cv^S(\Gamma, \Omega, 1-\alpha) - \epsilon \middle| \mathcal{W}_n^*\right) P_{F_n}(\mathcal{W}_n^*) \\
\geqslant& \liminf_{n\to\infty} P_{F_n}\left(cv_n\big(\hat{\Theta}_I(r_n), \hat{\Omega}_n, \varphi_n, \hat{G}_n, 1-\alpha\big) > cv^S(\Gamma, \Omega, 1-\alpha) - \epsilon \middle| \mathcal{W}_n^*\right) \liminf_{n\to\infty} P_{F_n}(\mathcal{W}_n^*) \\
\geqslant& \liminf_{n\to\infty} P_{F_n}\bigg(\inf_{(\Theta,\Omega,\varphi,G)\in A_n} cv_n\big(\Theta, \Omega, \varphi, G, 1-\alpha\big) > cv^S(\Gamma, \Omega, 1-\alpha) - \epsilon\bigg)
\end{aligned}$$

$$\geqslant \liminf_{n\to\infty} P_{F_n}\left(cv_n\big(\Theta_n^*,\Omega_n^*,\varphi_n^*,G_n^*,1-\alpha\big) > cv^S(\Gamma,\Omega,1-\alpha)-\frac{\epsilon}{2}\right)$$

$$\geqslant P\left(\liminf_{n\to\infty}\left\{cv_n\big(\Theta_n^*,\Omega_n^*,\varphi_n^*,G_n^*,1-\alpha\big) > cv^S(\Gamma,\Omega,1-\alpha)-\frac{\epsilon}{2}\right\}\right)$$

$$\geqslant P\left(\liminf_{n\to\infty} cv_n\big(\Theta_n^*,\Omega_n^*,\varphi_n^*,G_n^*,1-\alpha\big) > cv^S(\Gamma,\Omega,1-\alpha)-\frac{\epsilon}{2}\right)=1$$

the first inequality follows from $P(\cdot)\geqslant P(\cdot\cap\mathcal{W}_n^*)=P(\cdot|\mathcal{W}_n^*)P(\mathcal{W}_n^*)$; the second one is straightforward; the third one follows from conditional on $\mathcal{W}_n^*$,

$$cv_n\big(\hat{\Theta}_I(r_n),\hat{\Omega}_n,\varphi_n,\hat{G}_n,1-\alpha\big)\geqslant \inf_{(\Theta,\Omega,\varphi,G)\in A_n} cv_n\big(\Theta,\Omega,\varphi,G,1-\alpha\big)$$

and (B.36); in the fourth one, $(\Theta_n^*,\Omega_n^*,\varphi_n^*,G_n^*)\in A_n$ satisfies

$$\inf_{(\Theta,\Omega,\varphi,G)\in A_n} cv_n\big(\Theta,\Omega,\varphi,G,1-\alpha\big)+\frac{\epsilon}{2} > cv_n\big(\Theta_n^*,\Omega_n^*,\varphi_n^*,G_n^*,1-\alpha\big)$$

the fifth one follows from Fatou's lemma, the sixth one is straightforward, and the last one follows from that (B.19) holds for all $\epsilon>0$.

B.1.3.3 Part 3

Consider $(r_n,F_n)\in\mathcal{F}_0$ with $\Omega_{F_n}\xrightarrow{u}\Omega$, $\Gamma_{n,F_n}(r_n)\xrightarrow{H}\Gamma$,where $(\Gamma,\Omega)\in\mathcal{S}(\Theta\times\mathbb{R}^k_{[\pm\infty]})\times\mathcal{C}(\Theta^2)$.

Step 1. I first show

$$\lim_{n\to\infty} P_{F_n}\left(J_n(r_n) > cv_n\big(\hat{\Theta}_I(r_n),\hat{\Omega}_n,\varphi_n,\hat{G}_n,1-\alpha\big)+\eta\right)\leqslant\alpha$$

for all $\eta>0$. Define $\epsilon>0$ so that $\eta-\epsilon>0$ and $cv^S(\Gamma,\Omega,1-\alpha)+\eta-\epsilon$ is a continuity point of the CDF of $J(\Gamma,\Omega)$. It then follows that

$$\begin{aligned}
&P_{F_n}\left(J_n(r_n) > cv_n\left(\hat{\Theta}_I(r_n),\hat{\Omega}_n,\varphi_n,\hat{G}_n,1-\alpha\right)+\eta\right)\\
\leqslant& P_{F_n}\left(cv_n\left(\hat{\Theta}_I(r_n),\hat{\Omega}_n,\varphi_n,\hat{G}_n,1-\alpha\right)\leqslant cv^S(\Gamma,\Omega,1-\alpha)-\epsilon\right)\\
&+P_{F_n}\left(J_n(r_n) > cv^S(\Gamma,\Omega,1-\alpha)+\eta-\epsilon\right).
\end{aligned}$$

Take the limit supremum on both sides and by (B.35) and $J_n(r_n) \xrightarrow{d} J(\Gamma, \Omega)$, we get

$$\begin{aligned}&\limsup_{n\to\infty} P_{F_n}\left(J_n(r_n) > cv_n\left(\hat{\Theta}_I(r_n), \hat{\Omega}_n, \varphi_n, \hat{G}_n, 1-\alpha\right) + \eta\right)\\ \leqslant &1 - P\left(J(\Gamma,\Omega) \leqslant cv^S(\Gamma,\Omega,1-\alpha) + \eta - \epsilon\right) \leqslant \alpha. \end{aligned} \tag{B.42}$$

Step 2. For $\eta = 0$, there are two cases. First, suppose $cv^S(\Gamma, \Omega, 1-\alpha) = 0$, then by Assumption 87,

$$\limsup_{n\to\infty} P_{F_n}\left(J_n(r_n) > cv_n\big(\hat{\Theta}_I(r_n), \hat{\Omega}_n, \varphi_n, \hat{G}_n, 1-\alpha\big)\right) \leqslant \limsup_{n\to\infty} P_{F_n}\left(J_n(r_n) > 0\right) \leqslant \alpha.$$

Second, suppose $cv^S(\Gamma, \Omega, 1-\alpha) > 0$. Consider a sequence $\{\epsilon_m\}_{m\geqslant 1}$ s.t. $\epsilon_m \downarrow 0$ and $cv(\Gamma, \Omega, 1-\alpha) - \epsilon_m$ is a continuity point of the CDF of $J(\Gamma, \Omega)$ for all $m \in \mathbb{N}$. For any $m \in \mathbb{N}$, it follows from (B.42) that

$$\begin{aligned}&\limsup_{n\to\infty} P_{F_n}\left(J_n(r_n) > cv_n\big(\hat{\Theta}_I(r_n), \hat{\Omega}_n, \varphi_n, \hat{G}_n, 1-\alpha\big)\right)\\ \leqslant &1 - P\left(J(\Gamma,\Omega) \leqslant cv^S(\Gamma,\Omega,1-\alpha) - \epsilon_m\right).\end{aligned}$$

Taking $\epsilon_m \downarrow 0$ and using Assumption 87.1 gives the RHS equal to α. Together with the discussion in Section B.1.3.1 and (B.4), this completes the proof.

B.2 Lemmas

Lemma 90. *If $\nu_{1n} - \nu_{2n} = o_p(1)$, $\lfloor \nu_{2n} \rfloor_a = O_p(1)$, and $\Omega_{1n}, \Omega_{2n} \in \Psi_1$, $\Omega_{1n} - \Omega_{2n} = o_p(1)$, it holds that*

$$S(\nu_{1n}, \Omega_{1n}) = S(\nu_{2n}, \Omega_{2n}) + o_p(1).$$

Proof. We want to show that for all $\epsilon > 0$, there is N_1 such that for all $n > N_1$,

$$P\left(|S(\nu_{1n}, \Omega_{1n}) - S(\nu_{2n}, \Omega_{2n})| > \epsilon\right) \leqslant \epsilon.$$

Since $\lfloor \nu_{2n} \rfloor_a = O_p(1)$, for all $\epsilon > 0$, there is M such that for all $n \in \mathbb{N}$,

$$P\left(\nu_{2n} \in [-M, +\infty)^p \times [-M, M]^{k-p}\right) \geqslant 1 - \frac{\epsilon}{3}.$$

And by Lemma 92, there is $\delta > 0$ such that $\|\nu_{1n} - \nu_{2n}\| \leqslant \delta$, $\|\Omega_{1n} - \Omega_{2n}\| \leqslant \delta$ and $\nu_{2n} \in [-M, +\infty)^p \times [-M, M]^{k-p}$ imply that

$$|S(\nu_{1n}, \Omega_{1n}) - S(\nu_{2n}, \Omega_{2n})| \leqslant \epsilon.$$

Since $\nu_{1n} - \nu_{2n} = o_p(1)$ and $\Omega_{1n} - \Omega_{2n} = o_p(1)$, there is N_2 such that $\forall n \geqslant N_2$

$$P\left(\|\nu_{1n} - \nu_{2n}\| > \delta\right) \leqslant \frac{\epsilon}{3},\ P\left(\|\Omega_{1n} - \Omega_{2n}\| > \delta\right) \leqslant \frac{\epsilon}{3}.$$

Therefore, for $n \geqslant \max\{N_1, N_2\}$,

$$\begin{aligned}
&P\left(|S(\nu_{1n}, \Omega_{1n}) - S(\nu_{2n}, \Omega_{2n})| > \epsilon\right) \\
&\leqslant 1 - P\left(\nu_{2n} \in [-M, +\infty)^p \times [-M, M]^{k-p}, \|\nu_{1n} - \nu_{2n}\| \leqslant \delta, \|\Omega_{1n} - \Omega_{2n}\| \leqslant \delta\right) \\
&\leqslant P\left(\nu_{2n} \notin [-M, +\infty)^p \times [-M, M]^{k-p}\right) + P\left(\|\nu_{1n} - \nu_{2n}\| > \delta\right) + P\left(\|\Omega_{1n} - \Omega_{2n}\| > \delta\right) \\
&\leqslant \epsilon.
\end{aligned}$$

□

Lemma 91. *Assume that $\nu_n, \upsilon_n, \nu : \mathcal{W} \to \ell(\Theta)$, $\nu_n \xrightarrow{d} \nu$ and $\upsilon_n \xrightarrow{d} \nu$ where ν is a Gaussian process with covariance kernel $\Omega \in \mathcal{C}(\Theta^2)$, then there is $\upsilon_n^* : \mathcal{W} \to \ell(\Theta)$ (suitably enriched if needed) such that $\upsilon_n^* \sim \upsilon_n$ and $\nu_n - \upsilon_n^* \xrightarrow{p} 0$.*

Proof. I first show that there are identically distributed copies of ν_n and υ_n, denoted as $\tilde{\nu}_n$ and $\tilde{\upsilon}_n$, defined on $(\mathcal{W}, \mathcal{A}, F_n)$ (suitably enriched if needed) such that $\tilde{\nu}_n - \tilde{\upsilon}_n \xrightarrow{p} 0$. By almost sure representation, there is $\bar{\nu}_n \sim \nu_n$, $\bar{\nu} \sim \nu$ s.t. $\bar{\nu}_n - \bar{\nu} \xrightarrow{\text{a.s.}} 0$ and $\bar{\bar{\upsilon}}_n \sim \upsilon_n$, $\bar{\bar{\nu}} \sim \nu$ s.t. $\bar{\bar{\upsilon}}_n - \bar{\bar{\nu}} \xrightarrow{\text{a.s.}} 0$. Let α be the law of $(\bar{\nu}_n, \bar{\nu})$ and β be the law of $(\bar{\bar{\nu}}, \bar{\bar{\upsilon}}_n)$. By Theorem 1.31 in Dudley (2014), consider $X, Y, Z = R^{d_\nu}$, $\bar{\nu}_n$ defined on X, $\bar{\nu}, \bar{\bar{\nu}}$ defined on Y, $\bar{\bar{\upsilon}}_n$ defined on Z. Since the marginal distribution of α and β on Y is the same, we can find a law γ on $X \times Y \times Z$ (denote as the law of $(\tilde{\nu}_n, \tilde{\nu}, \tilde{\upsilon}_n)$) such that $(\tilde{\nu}_n, \tilde{\nu})$ has the same law as α and $(\tilde{\nu}, \tilde{\upsilon}_n)$ has the same law as β. Thus

$$\tilde{\nu}_n - \tilde{\nu} \xrightarrow{\text{a.s.}} 0, \tilde{\upsilon}_n - \tilde{\nu} \xrightarrow{\text{a.s.}} 0 \quad \Rightarrow \tilde{\nu}_n - \tilde{\upsilon}_n \xrightarrow{\text{a.s.}} 0 \Rightarrow \tilde{\nu}_n - \tilde{\upsilon}_n \xrightarrow{p} 0.$$

Then by the same argument in Chernozhukov et al. (2013) Lemma 9 (change δ_n in Lemma 9 to 1), there is an identical copy v_n^* of v_n such that $\nu_n - v_n^* \xrightarrow{p} 0$. □

Lemma 92. *For all $\epsilon > 0$, $M > 0$, there is $\delta > 0$ such that $\nu_1 \in [-M, \infty)^p \times [-M, M]^{k-p}$, $\|\nu_1 - \nu_2\| \leqslant \delta$ and $\Omega_1, \Omega_2 \in \Psi_1$, $\|\Omega_1 - \Omega_2\| \leqslant \delta$ imply that $|S(\nu_1, \Omega_1) - S(\nu_2, \Omega_2)| \leqslant \epsilon$.*

Proof. First note that $\forall \nu_{11} \in [-M, \infty]^p, \nu_{12} \in [-M, M]^{k-p}$,

$$\begin{aligned} S\left((\nu_{11}, \nu_{12}), \Omega_1\right) &\leqslant S\left((-M\mathbf{1}_p, \nu_{12}), \Omega_1\right) \\ &\leqslant \max_{\tilde{\nu}_{12} \in [-M,M]^{k-p}, \Omega_1 \in \Psi_1} S\left((-M\mathbf{1}_p, \tilde{\nu}_{12}), \Omega_1\right) \\ &< \infty \end{aligned}$$

where $\mathbf{1}_p = \{ \underbrace{1, ..., 1}_{p \text{ elements}} \}$, the first inequality follows from Assumption 88.1, the second one is straightforward, and the last one from Assumption 88.4 and the compactness of $[-M, M]^{k-p}$ and Ψ_1. Then consider function $\bar{S} : [\Phi(-M), 1]^p \times [\Phi(-M), \Phi(M)]^{k-p} \to \mathbb{R}$ defined by

$$\bar{S}(\bar{\nu}, \Omega) = S\left(\Phi^{-1}(\bar{\nu}), \Omega\right),$$

where for $\bar{\nu} = (\bar{\nu}_1, ..., \bar{\nu}_k) \in \mathbb{R}^k$, $\Phi^{-1}(\bar{\nu}) = \left(\Phi^{-1}(\bar{\nu}_1), ..., \Phi^{-1}(\bar{\nu}_k)\right)$. Φ is the CDF of standard Normal distribution. By Assumption 88.4, we have that $\bar{S}$ is continuous for all $\bar{\nu} \in [\Phi(-M), 1]^p \times [\Phi(-M), \Phi(M)]^{k-p}$, which implies that $\bar{S}$ is uniformly continuous on $[\Phi(-M), 1]^p \times [\Phi(-M), \Phi(M)]^{k-p}$. That is, there is $\delta > 0$ s.t. if $\|\bar{\nu}_1 - \bar{\nu}_2\| \leqslant \delta$ and $\|\Omega_1 - \Omega_2\| \leqslant \delta$, it holds that

$$\left|\bar{S}(\bar{\nu}_1, \Omega_1) - \bar{S}(\bar{\nu}_2, \Omega_2)\right| \leqslant \epsilon.$$

Note that $\|\nu_1 - \nu_2\| \leqslant \delta$ implies $\|\Phi(\nu_1) - \Phi(\nu_2)\| \leqslant \Phi'(0) \|\nu_1 - \nu_2\| < \delta$, thus if $\|\nu_1 - \nu_2\| \leqslant \delta$ and $\|\Omega_1 - \Omega_2\| \leqslant \delta$, it holds that

$$|S(\nu_1, \Omega_1) - S(\nu_2, \Omega_2)| = \left|\bar{S}\left(\Phi(\nu_1), \Omega_1\right) - \bar{S}\left(\Phi(\nu_2), \Omega_2\right)\right| \leqslant \epsilon.$$

□

Bibliography

Amengual, D., Bei, X., Carrasco, M., & Sentana, E. (2022). Score-type tests for normal mixture models. *CEMFI Working Paper*, (2213).

Amengual, D., Bei, X., & Sentana, E. (2020). Hypothesis tests with a repeatedly singular information matrix. *CEMFI Working Paper*, (2002).

Amengual, D., Bei, X., & Sentana, E. (2022). Normal but skewed? *Journal of Applied Econometrics*, *37*, 1295–1313.

Amengual, D., Bei, X., & Sentana, E. (2023). Highly irregular serial correlation tests. *CEMFI Working Paper*, (2302).

Amengual, D., & Sentana, E. (2015). Is a normal copula the right copula? *CEMFI Working Paper*, (1504).

Amengual, D., Sentana, E., & Tian, Z. (2022). Gaussian rank correlation and regression. *Advances in Econometrics*, *43B*, 269–306.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, 821–856 Publisher: JSTOR.

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). Misspecified Moment Inequality Models: Inference and Diagnostics. *The Review of Economic Studies*
rdad033.

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*.

Becker, J., Grun, G., Seemann, R., Mantz, H., Jacobs, K., Mecke, K., & Blossey, R. (2003). Complex dewetting scenarios captured by thin-film models. *Nat. Mater.*, *2*(1), 59–63.

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

Belloni, A., Bugni, F., & Chernozhukov, V. (2018). Subvector inference in partially identified models with many moment inequalities. *arXiv preprint arXiv:1806.11466*.

Belloni, A., Chernozhukov, V., Chetverikov, D., & Fernández-Val, I. (2019). Conditional quantile processes based on series or many regressors. *Journal of Econometrics*, *213*(1), 4–29.

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.

Bera, A., Ra, S., & Sarkar, N. (1998). Hypothesis testing for some nonregular cases in econometrics (C. Chakravarty & Mukherjee, Eds.), 319–351.

Bera, A., & Bilias, Y. (2001). Rao's score, neyman's c(α) and silvey's lm tests: An essay on historical developments and some new results. *Journal of Statistical Planning and Inference*, *97*, 9–44.

Berger, R. L., & Hsu, J. C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets [Publisher: JSTOR]. *Statistical Science*, 283–302.

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

Björkegren, D. (2019). The adoption of network goods: Evidence from the spread of mobile phones in Rwanda. *The Review of Economic Studies*, *86*(3), 1033–1060
Publisher: Oxford University Press.

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
Publisher: The University of Chicago Press Chicago, IL.

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

Bonn, D., Eggers, J., Indekeu, J., Meunier, J., & Rolley, E. (2007). *Wetting and spreading* [Submitted to *Rev. Mod. Phys.*]. http://itf.fys.kuleuven.be/~joi/papers/Wetting%20and%20spreading.pdf

Bottai, M. (2003). Confidence regions when the fisher information is zero. *Biometrika*, *90*, 73–84.

Breusch, T., & Pagan, A. (1980). The lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies*, *47*, 239–253.

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.

Canay, I. A., & Shaikh, A. M. (2017). Practical and theoretical advances in inference for partially identified models. *Advances in Economics and Econometrics*, *2*, 271–306.

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

Chang, H.-C., & Demekhin, E. (2002). Complex wave dynamics on thin films. Elsevier.

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
Publisher: Wiley Online Library.

Chesher, A., & Rosen, A. M. (2020a). Generalized instrumental variable models, methods, and applications. In *Handbook of Econometrics* (pp. 1–110, Vol. 7). Elsevier.

Chesher, A., & Rosen, A. M. (2020b). *Structural modeling of simultaneous discrete choice*. Cemmap, Centre for Microdata Methods; Practice, The Institute for Fiscal Studies, Department of Economics, UCL.

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
Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/07350015.2023.2203768.

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.

Christensen, T., & Connault, B. (2023). Counterfactual sensitivity and robustness. *Econometrica*, *91*(1), 263–298
Publisher: Wiley Online Library.

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

Ciliberto, F., & Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, *77*(6), 1791–1828
Publisher: Wiley Online Library.

Colbét, S. (1999). Name with proper names Hello Kitty. *Imag. Res.*, *1*, 123–127.

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

Constantine, G., & Savits, T. (1996). A multivariate faa di bruno formula with applications. *Transactions of the American Mathematical Society*, *348*, 503–520.

Cox, D., & Hinkley, D. (1974). *Theoretical statistics*. Chapman; Hall.

Cox, G., & Shi, X. (2019). Simple Adaptive Size-Exact Testing for Full-Vector and Subvector Inference in Moment Inequality Models. *arXiv preprint arXiv:1907.06317*.

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
Publisher: American Economic Association.

Davidson, J. (1994). *Stochastic limit theory: An introduction for econometricians*. Oxford University Press.

de Gennes, P.-G., Brochard-Wyart, F., & Quere, D. (2003). *Capillarity and wetting phenomena: Drops, bubbles, pearls, waves*. Springer Verlag.

di Bruno, F. F. (1859). *Théorie générale de l'élimination*.

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

Dovonon, P., & Renault, E. (2013). Testing for common conditionally heteroskedastic factors. *Econometrica*, *81*, 2561–2586.

Dudley, R. M. (2014). *Uniform central limit theorems* (Vol. 142). Cambridge university press.

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
Publisher: Oxford University Press.

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

Engle, R. (1983). Wald, likelihood ratio, and lagrange multiplier tests in econometrics. In M. D. Intriligator & Z. Griliches (Eds.), *Handbook of econometrics* (pp. 796–801). Elsevier.

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
Publisher: Oxford University Press.

Fang, Z., Santos, A., Shaikh, A., & Torgovitsky, A. (2020). Inference for Large-Scale Linear Systems with Known Coefficients. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2020-134).

Gallant, A., & Nychka, D. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica*, *55*, 363–390.

Godfrey, L. (1988). *Misspecification tests in econometrics*. Cambridge University Press.

Gualdani, C., & Sinha, S. (2019). Partial identification in matching models for the marriage market. *arXiv preprint arXiv:1902.05610*.

Hall, A. (1990). Lagrange multiplier tests for normality against seminonparametric alternatives. *Journal of Business and Economic Statistics*, *8*, 417–426.

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.

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica*, *42*, 679–694.

Heckman, J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, *5*, 475–492.

Helpman, E., Itskhoki, O., Muendler, M.-A., & Redding, S. J. (2016). Trade and inequality: From theory to estimation. *The Review of Economic Studies*, *84*(1), 357–405
Publisher: Review of Economic Studies Ltd.

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

Horowitz, J., & Savin, N. (2000). Empirically relevant critical values for hypothesis tests: A bootstrap approach. *Journal of Econometrics*, *95*, 375–389.

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
Publisher: Wiley Online Library.

Jarque, C., & Bera, A. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, *6*, 255–259.

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

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

Jondeau, E., & Rockinger, M. (2001). Gram-charlier densities. *Journal of Economic Dynamics and Control*, *25*, 1457–1483.

Juhl, T., & Xiao, Z. (2009). Tests for changing mean with monotonic power. *Journal of Econometrics*, *148*(1), 14–24
Publisher: Elsevier.

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

Kaido, H., Molinari, F., & Stoye, J. (2022). Constraint qualifications in partial identification. *Econometric Theory*, *38*(3), 596–619.

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

Kawai, K., & Watanabe, Y. (2013). Inferring strategic voting. *American Economic Review*, *103*(2), 624–662
Publisher: American Economic Association.

Kédagni, D., Li, L., & Mourifié, I. (2020). Discordant relaxations of misspecified models. *arXiv preprint arXiv:2012.11679*.

Kejriwal, M., & Perron, P. (2010). Testing for multiple structural changes in cointegrated regression models. *Journal of Business & Economic Statistics*, *28*(4), 503–522
Publisher: Taylor & Francis.

King, M. A., Sentana, E., & Wadhwani, S. (1990). Volatiltiy and links between national stock markets.

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 [Publisher: Wiley Online Library]. *Econometrica*, *73*(4), 1103–1123.

Kline, B., & Tamer, E. (2016). Bayesian inference in a class of partially identified models. *Quantitative Economics*, *7*(2), 329–366.

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

Kudo, A. (1963). A multivariate analogue of the one-sided test. *Biometrika*, *50*(3/4), 403–418.

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
Publisher: JSTOR.

Lee, L. F., & Chesher, A. (1986). Specification testing when score test statistics are identically zero. *Journal of Econometrics*, *31*, 121–149.

León, Á., Mencía, J., & Sentana, E. (2009). Parametric properties of semi-nonparametric distributions, with applications to option valuation. *Journal of Business & Economic Statistics*, *27*(2), 176–192.

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. (2018). Identification of treatment effects under conditional partial independence. *Econometrica*, *86*(1), 317–351
Publisher: Wiley Online Library.

Masten, M. A., & Poirier, A. (2021). Salvaging falsified instrumental variable models. *Econometrica*, *89*(3), 1449–1469
Publisher: Wiley Online Library.

Mattingley, J., & Boyd, S. (2012). CVXGEN: A code generator for embedded convex optimization. *Optimization and Engineering*, *13*(1), 1–27
Publisher: Springer.

Meitz, M., & Saikkonen, P. (2021). Testing for observation-dependent regime switching in mixture autoregressive models. *Journal of Econometrics*, *222*, 601–624.

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 Publisher: Wiley Online Library.

Molchanov, I. S., & Molchanov, I. S. (2005). *Theory of random sets* (Vol. 19). Springer.

Molinari, F. (2020). Microeconometrics with partial identification. *Handbook of econometrics*, *7*, 355–486.

Moreira, H. A., & Moreira, M. J. (2013). Contributions to the theory of optimal tests.

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

Nevo, A., & Rosen, A. M. (2012). Identification with imperfect instruments. *Review of Economics and Statistics*, *94*(3), 659–671
Publisher: The MIT Press.

Newey, W., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In R. Engle & D. McFadden (Eds.), *Handbook of econometrics* (pp. 2111–2245). Elsevier.

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, *4*, 2111–2245.

Norets, A., & Tang, X. (2014). Semiparametric inference in dynamic binary choice models. *Review of Economic Studies*, *81*(3), 1229–1262
Publisher: Oxford University Press.

O'Hagan, A., & Leonard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, *63*, 201–203.

Pakes, A., Porter, J., Ho, K., & Ishii, J. (2015). Moment inequalities and their application. *Econometrica*, *83*(1), 315–334
Publisher: Wiley Online Library.

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

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

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

Rao, C. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, *44*, 50–57.

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.

Rotnitzky, A., Cox, D., Bottai, M., & Robins, J. (2000). Likelihood-based inference with singular information matrix. *Bernoulli*, *6*, 243–284.

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

Sargan, J. (1983). Identification and lack of identification. *Econometrica*, *51*, 1605–1633.

Sentana, E. (1995). Quadratic ARCH models. *The Review of Economic Studies*, *62*(4), 639–661.

Silvey, S. D. (1959). The lagrangian multiplier test. *Annals of Mathematical Statistics*, *30*, 389–407.

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

Stoye, J. (2010). Partial identification of spread parameters. *Quantitative Economics*, *1*(2), 323–357
Publisher: Wiley Online Library.

Stoye, J. (2020a). *Simple misspecification adaptive inference for interval identified parameters* (tech. rep.). cemmap working paper.

Stoye, J. (2020b). A simple, short, but never-empty confidence interval for partially identified parameters. *arXiv preprint arXiv:2010.10484*.

Ura, T. (2018). Heterogeneous treatment effects with mismeasured endogenous treatment. *Quantitative Economics*, *9*(3), 1335–1370
Publisher: Wiley Online Library.

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

van der Vaart, A. (1998). *Asymptotic statistics*. Cambridge.

Wollmann, T. G. (2018). Trucks without bailouts: Equilibrium product characteristics for commercial vehicles. *American Economic Review*, *108*(6), 1364–1406
Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Wu, J. (2016). A test for changing trends with monotonic power. *Economics Letters*, *141*, 15–19
Publisher: Elsevier.

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.