Inference in Models Defined by Infinitely Many Inequalities: A Survey

Xiaoxia Shi*

October 7, 2025

Abstract

This paper complements the surveys by Molinari (2020) and Canay and Shaikh (2017) by reviewing the extensive body of work on inference based on an infinite number of inequalities. In addition to synthesizing existing results, I aim to identify gaps in the literature and outline promising avenues for future research.

Keywords: Moment Inequalities, Hypothesis Testing, Test Inversion, Gaussian Approximation, Power Regret

*Department of Economics, University of Wisconsin-Madison (xshi@ssc.wisc.edu). I thank Isaiah Andrews, Tim Armstrong, Qingliang Fan, Anders Kock, Simon Lee, Zhipeng Liao, Francesca Molinari, Kevin Song, Zhenting Sun, Yoonjae Whang, and Haitian Xie for helpful discussions. The paper is written while I am visiting Guanghua Management School at Peking University.

The study of partially identified models originated from the recognition that economic models often fail to yield point identification. Foundational contributions by Manski (1989, 1990, 1993) developed the use of bounds and incomplete models to draw inference under minimal assumptions, particularly in treatment effect and selection settings. Building on these early insights, a rich literature has developed on partially identified models. The applications of such models span all major areas of empirical economics including labor economics (e.g. Blundell et al. (2007)), industrial organization (e.g. Ciliberto and Tamer (2009)), trade (e.g. Morales et al. (2019), Kalouptsidi et al. (2020)), market design (e.g. He (2017) and Fack et al. (2019)), macroeconomics (e.g. Giacomini and Kitagawa (2021)), network formation and interaction (e.g. Sheng (2020)), and political economics (e.g. Iaryczower et al. (2018)). A more thorough overview of the applications can be found in the survey papers Ho and Rosen (2017), Pakes et al. (2015), Molinari (2020), and Kline et al. (2021). Canay et al. (2023) provide a user's guide for doing inference in such models.

In many applications, partially identified models are characterized by infinitely many moment (in)equalities. The combination of partial identification and the sheer number of inequality constraints poses distinct challenges for statistical inference. Over the past two decades, a rich set of methods has been developed, including diverse testing procedures and new asymptotic tools. Nevertheless, important questions remain open. This review seeks to synthesize key contributions from the existing literature and to highlight unresolved issues and promising avenues for future research.

This paper complements the comprehensive review of other aspects of partially identified models, by Molinari (2020) and Canay and Shaikh (2017). The former provides a high-level overview of the key issues including consistent set estimation, uniform coverage, and the effect of misspecification, with more detailed treatment on random set theory and computation. The latter focuses on inference methods for models defined by a finite number of moment inequalities.

1 Setup and Examples

A generic form of models defined by inequalities is as follows:

$$g_P(\theta, t) \leq 0 \text{ for all } t \in \mathcal{T}, \tag{1}$$

for an index set $\mathcal{T}$, where $g_P(\cdot, \cdot)$ is an unknown function that is determined by the unknown data generating process (DGP) P, and θ is an unknown parameter living in the parameter space Θ. This setup allows equalities: when there is an equality restriction, we can simply write it as a pair of opposing inequalities. For much of the subsequent discussion, it is not necessary to write the equalities out, but it should be noted that some test statistics used in the literature may benefit from an explicit equality/inequality notation.

This model is defined by an infinite number of inequalities when $\mathcal{T}$ contains an infinite number of points. There are a few reasons that $\mathcal{T}$ may be infinite, which I illustrate with examples:

Example 1 (Conditional Moment Inequalities)**.** *Many models used in structural estimation are conditional ones, where the model specifies the conditional generating process of the endogenous variables given exogenous variables. When model incompleteness and/or data imperfection lead to moment inequalities, they are conditional moment inequalities given the exogeneous variables. Mathematically, they are*

$$\mathbb{E}_P[m(W, \theta)|Z] \leq 0 \textit{ almost surely}, \tag{2}$$

where $m(\cdot, \cdot)$ is a known R^d-valued moment function for an integer $d > 0$, W is the vector of observables which may include exogenous and endogenous variables, and Z is the vector of exogenous variables. When Z contains a continuous variables, (2) stands for a continuum of inequalities.

There are two ways to write (2) in the form of (1). The first is the non-parametric *conditional mean approach, where*

$$g_P(\theta, t) = \mathbb{E}_P[m(W, \theta)|Z = t], \text{ for all } t \in \mathcal{T}, \tag{3}$$

where $\mathcal{T} = Supp(Z)$*, where* $Supp(Z)$ *stands for the support of* Z*. The second is the instrumental function approach, where*

$$g_P(\theta, t) = \mathbb{E}_P[m(W, \theta)t(Z)], \text{ for all } t \in \mathcal{T}, \tag{4}$$

where $\mathcal{T}$ *is a sufficiently rich class of functions mapping the support of* Z *to* $[0, \infty)$*. The classes of functions that are sufficiently rich are discussed in Andrews and Shi (2013). These two representations are equivalent, but they motivate different inference procedures as I discuss later.*

Example 2 (Sharp Identified Set)**.** *A common feature of many incomplete structural models is that the correspondence between observables and unobservables are multi-valued. Let* Y *stand for the vector of observables and* ε *stand for the vector of unobservables. The model imposes the following restrictions:*

$$\varepsilon \in \mathcal{E}(\theta, Y), \tag{5}$$

where $\mathcal{E}(\theta, y)$ *is a known closed set for each value of* θ *and* y*, and* $\varepsilon \sim G_\varepsilon(\cdot|\theta)$*. Then the sharp identified set is the set of* $\theta \in \Theta$ *such that the following inequalities hold:*

$$P(\mathcal{E}(\theta, Y) \subseteq A) \leq G_\varepsilon(A|\theta). \tag{6}$$

for all measurable subsets A *of the support of* ε*. See e.g. Chesher and Rosen (2017). Clearly, (6) stands for an infinite number of inequalities. The inequalities in (6) can be written in* *the form of (1) as follows:*

$$g_P(\theta, t) = \mathbb{E}[1\{\mathcal{E}(\theta, Y) \subseteq t\}] - G_\varepsilon(t|\theta), \text{ for all } t \in \mathcal{T}, \tag{7}$$

where $\mathcal{T}$ is the set of all measurable subsets of the support of Y. For specific models, such as discrete choice models with instrumental variables, selectively observed data, and some auction models, Galichon and Henry (2011), Chesher et al. (2013) and Chesher and Rosen (2017) develop methods to reduce the class of all measurable subsets to a much smaller core-determining class, but even that class can contain many elements.

Example 3 (Support Function Characterization of Sharp Identified Set)**.** *When certain moment conditions instead of parametric distribution assumptions are imposed on ε, Beresteanu et al. (2011) show that (6) can be equivalently expressed in terms of support functions. For example, $\mathbb{E}[\varepsilon] = \mathbf{0}$ and (6) together can be written equivalently as:*

$$\mathbb{E}h_{\mathcal{E}(\theta,Y)}(u) \geq 0 \text{ for all } u \in \mathbb{S}^{d_u-1}, \tag{8}$$

where $h_A(u) = \sup_{e\in A} u'e$ is the support function of the set A, $\mathbb{S}^{d_u-1} = \{u \in \mathbb{R}^{d_u} : \|u\| = 1\}$ and d_u is the dimension of ε.

Example 4 (Infinite Number of Conditional Moment Inequalities)**.** *In Examples 2 and 3, I abstracted away from exogenous variables for simplicity. However, in practice, there often are exogenous variables, and the model is often regarding conditional distributions of Y given exogenous variables. To be precise, let X be the vector of exogenous variables. Instead of (5), we have $\varepsilon \in \mathcal{E}(\theta, Y, X)$, where $\varepsilon \sim G_\varepsilon(\cdot|X, \theta)$. Instead of (6), we have*

$$P(\mathcal{E}(\theta, Y) \subseteq A|X) \leq G_\varepsilon(A|X, \theta), \text{ a.s.} \tag{9}$$

for all measurable subsets A of the support of ε. Instead of $\mathbb{E}_P[\varepsilon] = 0$, we have $\mathbb{E}_P[\varepsilon|X] = 0$, *and instead of (8), we have*

$$\mathbb{E}[h_{\mathcal{E}(\theta,Y,X)}(u)|X] \geq 0 \text{ for all } u \in \mathcal{S}^{d_u-1}. \tag{10}$$

When X contains a continuous variable, both (9) and (10) involve infinite number of conditional moment inequalities. We can write them in the form of (1) by either the nonparametric conditional mean approach as in (3) or the instrumental function approach as in (4).

1.1 Inference by Test Inversion

The inequality model (1) often does not point identify θ. Instead, it defines an identified set for θ:

$$\Theta_0(P) = \{\theta \in \Theta : g_P(\theta, t) \leq 0 \text{ for all } t \in \mathcal{T}\}. \tag{11}$$

If $g_P(\cdot,\cdot)$ is known, one can calculate this set using numerical or analytical tools. This is the exercise that, for example, Chesher et al. (2013) do in the numerical part of their paper, where they design a DGP and calculate $\Theta_0(P)$ under this artificial DGP. Such exercises are useful for studying the identification power of various model assumptions under a designed P, but not applicable in an empirical environment.

In an empirical environment, the researcher has a dataset drawn from P. The goal is to infer about $\Theta_0(P)$ based on the dataset. In standard point-identified models, one often calculates a consistent parameter estimator and builds a confidence interval around it. However, in the literature of partially identified models, the estimation of $\Theta_0(P)$ is out-shadowed by confidence set construction.[1] Part of the reason may be that the sample analogue estimator $\widehat{\Theta}_n = \{\theta \in \Theta : \widehat{g}(\theta, t) \leq 0 \text{ for all } t \in \mathcal{T}\}$ is inward biased unless strong

[1]A confidence set generalizes the concept of a confidence interval. It is a subset of the parameter space that has certain coverage probability guarantee.

assumptions are imposed, and removing the inward bias requires tuning parameters that the resulting estimator is senstive to. The other part may be that confidence sets based on a consistent set estimator are difficult to develop except in special cases. I do not discuss the estimation problem further but refer interested readers to Section 4.2 of Molinari (2020) for a thorough review of existing methods.

I focus on confidence set construction, and in particular, on confidence sets that cover the true value of the parameter with a given probability (asymptotically):

$$\inf_{P\in\mathcal{P}} \inf_{\theta_0\in\Theta_0(P)} \Pr{}_P(\theta_0 \in CS_n(1-\alpha)) \geq 1-\alpha+o(1), \tag{12}$$

where $\mathcal{P}$ is a set of DGPs allowed by the model and $\alpha \in (0,1)$ is a nominal significance level. Since the researcher does not know the true DGP or which point in $\Theta_0(P)$ is the true value even given P, we would like the minimum coverage probability under all possible combinations of (P, θ_0) allowed by the model to be bounded from below. Confidence sets that satisfy (12) are said to have uniform asymptotic coverage for the true value of the parameter. [2]

Confidence sets satisfying (12) are constructed by test inversion. Specifically, one constructs a family of tests $\{\varphi_{n,\alpha}(\theta) : \theta \in \Theta\}$, where for each $\theta \in \Theta$, $\varphi_{n,\alpha}(\theta)$ is a test for the hypothesis

$$H_0 : g_P(\theta, t) \leq 0 \text{ for all } t \in \mathcal{T}. \tag{14}$$

[2]The literature has also defined a different notion of coverage:

$$\inf_{P\in\mathcal{P}} \Pr{}_P(\Theta_0(P) \subseteq CS_n(1-\alpha)) \geq 1-\alpha+o(1). \tag{13}$$

Confidence sets that satisfy this coverage guarantee are said to have uniform asymptotic coverage for the identified set. This notion of coverage is stronger than (12) and typically requires the confidence set to be wider. It is worth noting that some papers do not consider $\inf_{P\in\mathcal{P}}$ in (13). See Chernozhukov et al. (2007). That results in the pointwise-in P asymptotic coverage of the identified set, which is not stronger (or weaker) than (12).

Based on the tests, one defines the confidence set to be

$$CS_n(1-\alpha) = \{\theta \in \Theta : \varphi_{n,\alpha}(\theta) = 0\}, \tag{15}$$

that is, the set of θ values at which the test does not reject. Since $\mathbb{E}_P[\varphi_{n,\alpha}(\theta)] = 1 - \Pr_P(\varphi_{n,\alpha}(\theta) = 0)$, the coverage guarantee (12) is satisfied if the test has the following uniform asymptotic level control:

$$\sup_{P \in \mathcal{P}} \sup_{\theta \in \Theta_0(P)} \mathbb{E}_P[\varphi_{n,\alpha}(\theta)] \leq \alpha + o(1). \tag{16}$$

In practice, $CS_n(1-\alpha)$ is often computed by conducting the test $\varphi_{n,\alpha}(\theta)$ for a grid of θ values on Θ, and by inferring the boundary of $CS_n(1-\alpha)$ from the acceptance and the rejection regions on the grid. When the projection of $CS_n(1-\alpha)$ on a scalar parameter, say $\lambda(\theta)$, is desired, one can also compute the lower and upper end points of the projection via the following constrained optimization problems: $\min \setminus \max_{\theta \in \Theta : \varphi_{n,\alpha}(\theta)=0} \lambda(\theta)$.

As we can see, constructing the confidence set (15) is naturally related to the literature of jointly testing an infinite number of inequalities. The hypotheses tested in this literature include stochastic dominance (e.g. Barrett and Donald (2003),Donald and Hsu (2016), Chetverikov et al. (2021)), conditional stochastic dominance (e.g. Delgado and Escanciano (2013), Lok and Tabri (2021)), stochastic monotonicity (e.g. Lee et al. (2009), Seo (2018)), regression monotonicity (e.g. Ghosal et al. (2000),Hsu et al. (2019)), density ratio ordering (Carolan and Tebbs (2005), Beare and Moon (2015), Beare and Shi (2019)), conditional predictive superiority (e.g. Li et al. (2022)) and so on. These hypotheses can be viewed as a special case of (14) where a parameter θ is not there. They can be tested using the tests developed for (14), although in some of the afore-mentioned papers, specialized tests that are not applicable to (14) are developed. Here I focus on the generic tests and do not discuss the specialized procedures.

2 A Selective Review of Existing Tests

Now I discuss the existing tests for (14). Since the tests are constructed for each given θ, we omit θ and henceforce write $g_P(\theta, t)$ as $g_P(t)$ for notational simplicity.

There are two scenarios in which the literature has considered (14), depending on the large sample property of the estimator of $g_P(t) : t \in \mathcal{T}$. Let $\widehat{g}_n(t) : t \in \mathcal{T}$ denote the estimator. The two scenarios are as follows:

1. The sequence of stochastic processes $\{r_n(\widehat{g}_n(t) - g_P(t)) : t \in \mathcal{T}\}_{n=1}^{\infty}$ converges weakly in $\ell^{\infty}(\mathcal{T}, \mathbb{R}^d)$ to a tight Gaussian process $G(t) : t \in \mathcal{T}$, for a normalizing sequence $\{r_n\}_{n=1}^{\infty}$. [3]

2. $\{r_n(\widehat{g}_n(t) - g_P(t)) : t \in \mathcal{T}\}_{n=1}^{\infty}$ does not converge weakly to a tight Gaussian process for any normalizing sequence $\{r_n\}$.

2.1 Scenario 1

In the first scenario, standard empirical process techniques, in e.g. van der Vaart and Wellner (1996), can be applied. This allows one to aggregate the information contained in each dimension of $\widehat{g}_n(t)$ for each t in a variety of ways to form the test statistic. In particular, it allows us to study the distribution of tests statistics of the form

$$T_n^{\text{CvM}} = \int_{\mathcal{T}} S(r_n \widehat{g}_n(t), \overline{\Sigma}_n(t)) d\mu(t), \tag{17}$$

where $\overline{\Sigma}_n(t)$ is an estimator of the variance-covariance matrix of $r_n(\widehat{g}_n(t) - g_P(t))$ and $S(\cdot, \cdot)$ is a user-chosen function to aggregate different dimensions of $\widehat{g}_n(t)$. This is called the Cramér-von-Mises (CvM) type statistic in Andrews and Shi (2013). Commonly used S functions are $S^{\max}(g, \Sigma) = \max_j \frac{\max\{g_j, 0\}^2}{\sigma_j^2}$, $S^{\text{sum}}(g, \Sigma) = \sum_j \frac{\max\{g_j, 0\}^2}{\sigma_j^2}$., and $S^{\text{qlr}}(g, \Sigma) = \min_{x \leq 0} (g - x)' \Sigma^{-1} (g - x)$.

[3] Here $\ell^{\infty}(\mathcal{T}, \mathbb{R}^d)$ consists of all bounded functions $f : \mathcal{T} \to \mathbb{R}^d$.

An alternative to the CvM-type statistic is the Kolmogorov-Smirnov (KS)-type Statistic, for example,

$$T_n^{KS} = \sup_{t \in \mathcal{T}} S(r_n \widehat{g}_n(t)), \overline{\Sigma}_n(t)). \tag{18}$$

Both types of statistics are considered in the literature for Scenario 1. Andrews and Shi (2013) consider both for conditional moment inequality models, Andrews and Shi (2014) for a non-parametric conditional moment inequality model where conditional moment inequalities hold at a given value of a subset of the conditional variables. Andrews and Shi (2017) study a model defined by infinitely many conditional moment inequalities. Hsu et al. (2019) extend it to testing generalized regression monotonicity. In these papers, a large sample distributional approximation of the test statistic is derived instead of a limit distribution. However, it is possible to derive a limit distribution for the CvM-type T_n. I do so in Proposition 1, the proof of which is given in the appendix.

Proposition 1. *Let $\{P_n\}$ be a sequence of distributions such that H_0 in (14) holds for each n. Suppose that there is a subsequence of $\{n\}$ along which:*

(i) $r_n(\widehat{g}_n(t) - g_{P_n}(t)) \Rightarrow G(t)$ *in* $\ell^\infty(\mathcal{T}, \mathbb{R}^d)$*, where G is a tight Gaussian process;*

(ii) $\sup_{t \in \mathcal{T}} \|\overline{\Sigma}_n(t) - \overline{\Sigma}(t)\| \to_p 0$*, for some* $\overline{\Sigma} \in \ell^\infty(\mathcal{T}, \mathbb{S}_+^d)$ *with eigenvalues bounded below by $\varepsilon > 0$, where $\mathbb{S}_+^d$ is the set of positive definite matrices of size d;*

(iii) $-r_n g_{P_n}(t) \to h(t)$ *pointwise, for some* $h : \mathcal{T} \to [0, \infty]^d$*;*

(iv) $S : [-\infty, \infty)^d \times \mathbb{S}_+^d \to \mathbb{R}_+$ *is continuous, non-decreasing in the first argument, and satisfies* $S(m, \Sigma + \Sigma_1) \leq S(m, \Sigma)$ *for all m and* $\Sigma, \Sigma_1 \in \mathbb{S}_+^d$.

Then along this subsequence, $T_n^{\text{CvM}} := \int_{\mathcal{T}} S(r_n \widehat{g}_n(t), \overline{\Sigma}_n(t)) d\mu(t) \longrightarrow_d T_\infty^{\text{CvM}} := \int_{\mathcal{T}} S(G(t) - h(t), \overline{\Sigma}(t)) d\mu(t)$.

The proposition extends the distributional approximation result in Theorem 1 of Andrews and Shi (2013) to a limit distribution one, with a key condition added: condition (iii). This condition is not easy to verify when P_n changes with n unless $\mathcal{T}$ is countable. This is why the papers mentioned above this proposition opt to derive a distributional approximation instead, despite a drawback associated with the lack of a limit distribution, a drawback I discuss shortly below. Even with condition (iii) added, an analogous result for the KS statistic does not hold because the bounded convergence theorem is used in the proof and it does not apply for the supremum operator. [4]

Based on a result like that in Proposition 1, one can obtain a simulation-based critical value after estimating a lower bound for $h(\cdot)$. We can only bound $h(\cdot)$ instead of consistently estimating it because it is a limit of $r_n g_{P_n}(t)$ and $g_{P_n}(t)$ can at best be estimated r_n-consistently. On the other hand, a lower-bound is sufficient for constructing a valid test because replacing $h(\cdot)$ by a lower bound enlarges the integral $\int_{\mathcal{T}} S(G(t) - h(t), \overline{\Sigma}(t))d\mu(t)$ due to the monotonicity of $S(g, \Sigma)$ in its first argument. Therefore, the resulting simulated critical value is asymptotically valid (i.e. not leading to excesive over-rejection uner H_0).

Typically, there are two ways to bound $h(t)$. The first is sometimes called "least favorable" and sometimes called "plug-in asymptotics (PA)" in the literature. It is to bound $h(t)$ by 0 for all t, and is justified by the fact that $g_{P_n}(t) \leq 0$ for all P_n satisfying H_0 in (14).

The second is called "generalized moment selection (GMS)" in the literature. The idea is to approximate $h_j(t)$ by ∞ or something that diverges to ∞ if there is strong evidence that $h_j(t) = \infty$ and to replace it by zero or something that converges to zero otherwise. Andrews and Shi (2013) recommend the following GMS bound for $h(\cdot)$ (when $r_n = n^{1/2}$): $\underline{h}_{n,j}(t) = B_n 1\{\kappa_n^{-1} n^{1/2} \widehat{g}_{n,j}(t)/\overline{\sigma}_{n,j}(t) < -1\}$, where κ_n and B_n (e.g. $\kappa_n = (0.3\log(n))^{1/2}$, $B_n = (0.4\log(n)/\log(\log(n)))^{1/2}$) are user-chosen positive constants such that $\kappa_n \to \infty$, $B_n \to \infty$, and $B_n/\kappa_n \to 0$ as $n \to \infty$, and $\overline{\sigma}_{n,j}(t)$ is the jth diagonal element of $\overline{\Sigma}_n(t)$.

Once a feasible bound $\underline{h}_n(t)$ is chosen, one can define the critical value, $cv_n(\alpha)$, to be the

[4] Nevertheless, under a fixed P satisfying H_0 in (14) (as opposed to a drifting sequence $\{P_n\}$), Barrett and Donald (2003) establish the limit distribution for an identity weighted KS statistic for the hypothesis of stochastic dominance.[5] The arguments do not appear to be specific to stochastic dominance or to using identity weighting, but do seem to depend on the fixed P. It is an open question how it extends to drifting P. The drifting P result is necessary to prove an asymptotic size result that is uniform over DGPs, that is, a result like (16) with the $\sup_{P \in \mathcal{P}}$ (and $\sup_{\theta \in \Theta}$ if there is a θ).

simulated $1-\alpha$ quantile of:

$$T_n^{cv} = \int_{\mathcal{T}} S(G_n^*(t) + \underline{h}_n(t), \overline{\Sigma}_n(t)) d\mu(t), \tag{19}$$

where $G_n^*(t)$ is the random component (conditional on data) to be simulated. It can be the bootstrap empirical process $r_n(\widehat{g}_n^*(t) - \widehat{g}_n(t)) : t \in \mathcal{T}$ where $\widehat{g}_n^*(t)$ is $\widehat{g}_n(t)$ calculated from a bootstrap sample. It can also be a Gaussian process with variance covariance kernel $\widehat{\Sigma}_n(t_1, t_2) : t_1, t_2 \in \mathcal{T}$ which for each (t_1, t_2) is a consistent estimator of $\text{Cov}(r_n(\widehat{g}_n(t_1) - g_P(t_1)), r_n(\widehat{g}_n(t_2) - g_P(t_2)))$. Finally, the test is defined to be $\varphi_{n,\alpha} = 1\{T_n > cv_n(\alpha)\}$.

A few things are left out in the foregoing discussion. First, the eigenvalues of the matrix $\overline{\Sigma}(t)$ is required to be uniformly bounded away from zero. This is typically not satisfied if $\overline{\Sigma}_n(t)$ is a *consistent* estimator of the variance-covariance matrix of $r_n(\widehat{g}_n(t) - g_P(t))$. This is because often the latter matrix can be arbitrarily close to singularity when $\mathcal{T}$ has infinitely many elements. Thus, in order to take advantage of the empirical process result the way that Proposition 1 and results like Theorem 1 of Andrews and Shi (2013) do, the variance-covariance matrix needs to be regularized. Thus, $\overline{\Sigma}_n(t)$ is often taken as $\widehat{\Sigma}_n(t) + \varepsilon\widehat{\Sigma}_n$ where $\widehat{\Sigma}_n$ is $\widehat{\Sigma}_n(t)$ at a particular t where the variance-covariance matrix is not degenerate. The regularization parameter ε is not allowed to converge to zero as $n \to \infty$. It is interesting to ask what happens to the limit distribution of T_n if a sequence $\varepsilon_n \to 0$ is used instead of a fixed ε. For the CvM statistic, this is an open question.[6]

Second, the critical value that I define above is not exactly the same as that proposed in Andrews and Shi (2013) and the subsequent papers. Specifically, the critical values used in that literature are η plus the simulated $1-\alpha+\eta$ quantile, where η is the so-called infinitesimal constant. The constant is needed to prove the asymptotic level control of the test because the limit distribution of T_n is not derived in that literature, but instead, an approximate

[6]Kolmogorov-Smirnov (KS) type statistics without such a regularization is discussed extensively in Scenario 2 where penultimate distributional approximations are derived but not limit distributions. A limit distribution result for that such a KS statistic is derived in Armstrong (2015) for the conditional moment inequality hypothesis under the data generating processes that the conditional moments are binding (= 0) on a measure-zero set. The derivation does not apply when the binding set is not of measure-zero.

distribution that still depends on $r_n g_{P_n}(t)$ is derived. That alone is not sufficient to justify the use of the $1-\alpha$ quantile of the approximating distribution as the critical value.

Proposition 1 establishes the limit distribution of T_n^{CvM}, and Lemma B.3 in the supplemental appendix of Andrews and Shi (2013) proves the continuity and strict monotonicity of the limit distribution when the following S functions are used $S^{\max}(\cdot,\cdot)$ and $S^{\text{sum}}(\cdot,\cdot)$. In light of Lemma 5 of Andrews and Guggenberger (2010), these results together should obviate the need for the infinitesimal constant.

Third, the GMS procedure selects the binding inequalities through benchmarking the t statistic $r_n\widehat{g}_{n,j}(t)/\overline{\sigma}_{n,j}(t)$ against the fixed threshold $-\kappa_n$. Both the size and the power of the test can be somewhat senstive to the choice of κ_n. Alternatively, one can use, in place of $-\kappa_n$, a data dependent threshold that bounds $h(t)$ probablisitically. To be specific, defined z_n^β as the $100\beta\%$ quantile of $\inf_j \inf_{t\in\mathcal{T}} G_{nj}^*(t)/\overline{\sigma}_{n,j}(t)$ where $G_n^*(t)$ is explained under (19), and let $\underline{h}_{n,j}^\beta(t) = \max\{-r_n\widehat{g}_{n,j}(t) + z_n^\beta\overline{\sigma}_{n,j}(t), 0\}$. Asymptotically, $-r_n g_{P_n,j}(t) \geq \underline{h}_n^\beta(t) \;\; \forall t\in\mathcal{T}$ with probability approaching $1-\beta$. Define $cv_n^\beta(\alpha)$ to be $100(1-\alpha+\beta)\%$ conditional quantile of $\int_{\mathcal{T}} S(G_n^*(t) + \underline{h}_n^\beta(t), \overline{\Sigma}_n(t))d\mu(t)$. Then the test $\varphi_{n,\alpha}^\beta = 1\{T_n > cv_n^\beta(\alpha)\}$ can be proved to be uniformly asymptotically valid for any $\beta\in(0,\alpha)$. The choice of β only affects its power, unlike the choice of κ_n in the GMS test above, which affects both size and power. This technique is proposed in Romano et al. (2014) in the context of testing a finite set of moment inequalities. For testing many inequalities, it has not been used for CvM-type statistics, although it has been used for KS-type statistics developed in Scenario 2, which we move on to now.

2.2 Scenario 2

In the second scenario, weak convergence fails to hold. This arises, for instance, when $g_P(t)$ includes nonparametrically estimated conditional expectations or if $\mathcal{T}$ is a discrete set and $g_P(t)$ does not have a particular structure.

In the conditional moment inequality models described in Examples 1 and 4, if one writes down $g_P(t)$ using the instrumental function approach illustrated in (3), and the set of instrumental functions is appropriately chosen to be rich enough but not too large (ref. Andrews and Shi (2013)), then one can form a $\widehat{g}_n(t)$ to satisfy weak convergence and hence to use the techniques developed for Scenario 1. On the other hand, if one writes down $g_P(t)$ using the nonparametric conditional mean approach described in (4) and estimate $g_P(t)$ nonparametrically, weak convergence will not hold and one is in Scenario 2.

In Example 2 when the core determining class is a finite union of half spaces and in Example 3, with mild regularity conditions, one is in Scenario 1, as proved in Sections 7-9 in Andrews and Shi (2017). Otherwise, one is in Scenario 2.

In the literature that works under Scenario 2, the variety of test statistics considered has been much more limited. The predominant choice is a supremum (SUP)-t statistics, with the notable exception of Lee et al. (2013) and Lee et al. (2018). I discuss the approach in those two papers in a subsection below.

Suppose that $g_P(t)$ is scalar-valued.[7] The SUP-t statistic is

$$T_n^{\sup} = \sup_{t \in \mathcal{T}} \frac{r_n \widehat{g}_n(t)}{\widehat{\sigma}_n(t)}, \tag{20}$$

where $\widehat{\sigma}_n(t)^2$ is a consistent estimator of the asymptotic variance of $r_n(\widehat{g}_n(t) - g_P(t))$.[8]

Let $\widehat{Z}_n(t) = r_n(\widehat{g}_n(t) - g_{P_n}(t))/\widehat{\sigma}_n(t)$ for $t \in \mathcal{T}$. Note that this process may not weakly converge to a tight Gaussian process even when $r_n(\widehat{g}_n(t) - g_{P_n}(t)) : t \in \mathcal{T}$ does if $\widehat{\sigma}_n(t)$ is not bounded away from zero, let alone when $r_n(\widehat{g}_n(t) - g_{P_n}(t)) : t \in \mathcal{T}$ does not converge weakly. Thus, different tools are needed to analyze the distributional behavior of $T_n^{\sup}$.

One tool used in this literature is the strong approximation of $\widehat{Z}_n(t) : t \in \mathcal{T}$ by a

[7]It is without loss of generality to assume that $g_P(t)$ is scalar-valued when the SUP statistic is used. This is because each component of $g_P(t)$ enters the statistic separately. Thus, one can simply lump the component index with t.

[8]Note that this statistic is the square root of T_n^{KS} above with S being $S^{\max}$ and with the diagonals of $\overline{\Sigma}_n(t)$ being $\widehat{\sigma}_n(t)^2$.

penultimate Gaussian process $Z_n^*(t) : t \in \mathcal{T}$:

$$\sup_{t \in \mathcal{T}} |\widehat{Z}_n(t) - Z_n^*(t)| = o_p(\delta_n), \tag{21}$$

where δ_n is a sequence such that $\delta_n \to 0$. See Chernozhukov et al. (2013). Strong approximation of this type can be established using coupling arguments such as the Yurinskii coupling for sums of independent random variables (Chapter 10 of Pollard (2002)) and for partial sums of mixingales (Li and Liao (2020)). I describe a distributional approximation result for $T_n^{\sup}$ based on this tool.

A related tool is the Gaussian approximation of the supreme of $\widehat{Z}_n(t)$, such as

$$\left| \sup_{t \in \mathcal{T}} \widehat{Z}_n(t) - \sup_{t \in \mathcal{T}} Z_n^*(t) \right| = o_p(\delta_n). \tag{22}$$

Not requiring full approximation of the entire stochastic process $\widehat{Z}_n(t) : t \in \mathcal{T}$ allows weaker assumptions to be made regarding $\widehat{g}_P(t) : t \in \mathcal{T}$. The verification and the usage of this approximation are considered in Chernozhukov et al. (2014b). Also in this genre, Li et al. (2022) use a Gaussian approximation for the supreme over many projection directions of $\widehat{Z}_n(t)$ to derive a test for superior predictivity. They prove the validity of the approximation for the nonparametric series estimator in dependent data.

A different type of tools is central limit theorems (CLT) in high dimensions (e.g. Chernozhukov et al. (2017) and Fang and Koike (2024)). Such results are stated with a finite but increasing $\mathcal{T}$. Specifically, denote the $\mathcal{T}$ used for a given sample size n as $\mathcal{T}_n$. The set $\mathcal{T}_n$ consists of p_n elements.[9] Let $\vec{Z}$ denote $(Z(t_1), \ldots, Z(t_{p_n}))'$. A CLT in high dimensions is of the form:

$$\sup_{A \in \mathcal{A}_n} |\Pr(\vec{\widehat{Z}}_n \in A) - \Pr(\vec{Z}_n^* \in A)| = o(\xi_n), \tag{23}$$

[9]The finite set $\mathcal{T}_n$ can be viewed as an increasingly fine discretization of an uncountable $\mathcal{T}$.

where $\mathcal{A}_n$ is a collection of measurable sets on $\mathbb{R}^{p_n}$, such as the collection of hyperrectangles or of Euclidean balls, and $\xi_n \to 0$. Such results are used to derive tests for increasingly many moment inequalities, for example, in Chetverikov (2018) and Bai et al. (2022).

One last tool used is the analytical bounds for the tail probability of $\sup_{t\in\mathcal{T}_n} \widehat{Z}_n(t)$, derived using the subadditivity of probability measures and tail bounds for standard normal random variables (ref. Chernozhukov et al. (2019)). These bounds are not as accurate as the CLT approximations, but they do not require bootstrap to implement and can allow p_n to be much larger than n.

Now I describe a result from Chernozhukov et al. (2013). Consider a sequence of data generating processes $\{P_n\}$ satisfying H_0. Let $\sigma_n(t)$ be the population counterpart of $\widehat{\sigma}_n(t)$. Let

$$T_n^* = \sup_{t\in\mathcal{T}_n^*} Z_n^*(t), \tag{24}$$

where $\mathcal{T}_n^* = \{t \in \mathcal{T} : r_n g_{P_n}(t) \geq -\sigma_n(t)k_n(\gamma_n)\}$, $k_n(\gamma)$ is the $100(1-\gamma)\%$ quantile of $\sup_{t\in\mathcal{T}} Z_n^*(t)$, and γ_n is a sequence of positive numbers such that $\gamma_n \to 0$.

Proposition 2. *Let $\mathcal{T}$ be a compact subset of $\mathbb{R}^d$ and suppose that $\mathcal{T}_n^*$ is compact for all n. Moreover, consider a sequence of DGPs $\{P_n\}$ satisfying H_0. Suppose that*

(i) Strong Approximation: *under the sequence $\{P_n\}$* (21) *holds,*

(ii) Anti-Concentration: $\sup_{x\in R} \Pr\left(|\sup_{t\in\mathcal{T}_n} Z_n^*(t) - x| \leq \delta_n\right) \to 0$ *for any compact subset $\mathcal{T}_n \subseteq \mathcal{T}$.*

(iii) Variance Convergence: $\sup_{t\in\mathcal{T}} \left|\frac{\sigma_n(t)}{\widehat{\sigma}_n(t)} - 1\right| = o_p(\delta_n k_n(\gamma_n)^{-1})$.

Then, uniformly over $x \in [0,\infty)$,

$$\Pr{}_{P_n}(T_n^{\sup} \leq x) \geq \Pr(T_n^* \leq x) - o(1). \tag{25}$$

Based on a distributional approximation result such as that in Proposition 2, we can define a simulated critical value after bounding the contact set $\mathcal{T}_n^*$, that is, finding an index set $\widehat{\mathcal{T}}_n$ such that

$$\Pr(\mathcal{T}_n^* \subseteq \widehat{\mathcal{T}}_n) \to 1. \tag{26}$$

Clearly, the least-favorable option, $\widehat{\mathcal{T}}_n = \mathcal{T}$, satisfies the requirement. Chernozhukov et al. (2013) consider a more sophisticated $\widehat{\mathcal{T}}_n$ that can be a much smaller set than $\mathcal{T}$ while still satisfying (26) when the inequalities $g_P(t) \leq 0$ are slack on most of $\mathcal{T}$. The smaller $\widehat{\mathcal{T}}_n$ serves a similar role as moment selection and can reduce under-rejection comparing to the least-favorable test.

Once $\widehat{\mathcal{T}}_n$ is constructed, one defines the critical value $cv_n^{\sup}(\alpha)$ to be the $100(1-\alpha)\%$ quantile of $\sup_{t\in\widehat{\mathcal{T}}_n} \tilde{Z}_n(t)$, where $\tilde{Z}_n(t)$ is the random component to be simulated. Its conditional distribution (given data) should approximate that of $\sup_{t\in\mathcal{T}_n^*} Z_n^*(t)$. A typical choise for $\widetilde{Z}(t) : t \in \mathcal{T}$ is a Gaussian multiplier bootstrap process because it is easier to establish strong approximation results for such processes than for exchangable bootstrap. Finally, the test is $\varphi_{n,\alpha}^{\sup} = 1\{T_n^{\sup} > cv_n^{\sup}(\alpha)\}$.

It is worth discussing the difference between the conditions for Propositions 1 and 2. Unlike Proposition 1, Proposition 2 does not require the weak convergence of the empirical process and it allows unregularized studentization of $\widehat{g}_n(t)$, that is, the $\widehat{\sigma}_n(t)$ does not need to be bounded away from zero. However, Proposition 2 requires an anti-concentration property for the distribution of $\sup_{t\in\mathcal{T}} Z_n^*(t)$ (Condition (ii)). It also needs $k_n(\gamma_n)$ to not grow too fast because otherwise Condition (iii) is violated. This is typically verified through a concentration property, or in other words, a tail probability bound, of the supremum of Gaussian processes. The anti-concentration and concentration properties have been established under general conditions in Chernozhukov et al. (2014a) and Chernozhukov et al. (2015) for $\sup_{t\in\mathcal{T}} Z_n^*(t)$.

Anti-concentration and concentration results are not available for more general functionals of Gaussian processes. This may be the reason that test statistics other than T_n^{sup} are overlooked in Scenario 2 except in Lee et al. (2013) and Lee et al. (2018) which we discuss below.

Nevertheless, if CLT results with a convergence rate, such as those in Chernozhukov et al. (2017), are used instead of strong approximation, one may not need to separately establish anti-concentration or concentration of the approximating Gaussian process, which may be a viable direction to develop tests based on a wider variety of test statistics in Scenario 2.

2.2.1 CvM-type Test in Scenario 2

In the special case of conditional moment inequality models, under representation (3), a CvM-type test statistic is considered in Lee et al. (2013) (LSW). In particular, they use the statistic

$$T_n^{LSW} = \sum_{j=1}^{d} \int_{\mathcal{Z}} \max\left\{0, \sum_{i=1}^{n} m_j(W_i) K_h(Z_i - z)\right\}^p w(z)dz, \tag{27}$$

where $p \geq 1$, $K_h(x) = \frac{1}{h}K(x/h)$, h is a bandwidth such that $h \to 0$ as $n \to \infty$, $K(\cdot)$ is a kernel function with bounded support, and $w(z)$ is a user-defined nonnegative and square-integrable weight function.

LSW discretize the integral into a sum of integrals over fine partions of $\mathcal{Z}$. Then they show via Poissonization that integrals on each part are close to being a one-dependent random field. Applying a central limit theorem on this random field allows them to conclude that $s_n^{-1}(T_n^{LSW} - a_n) \to \mathcal{N}(0, 1)$. Here s_n and a_n are normalizing sequences that depend on the conditional means and conditional variances of $m_j(W) : j = 1, \dots, d$ given Z. Then they show that s_n and an upper bound of a_n can be estimated (say by $\widehat{s}_n$ and $\widehat{a}_n$ respectively) sufficiently accurately so that $\widehat{s}_n^{-1}(T_n^{LSW} - \widehat{a}_n) \to_d \mathcal{N}(0, 1)$ when $g_P(t) = 0$ for all $t \in \mathcal{T}$ and is asymptotically stochastically bounded by $\mathcal{N}(0, 1)$ under H_0 when $g_P(t)$ is not zero for all

$t \in \mathcal{T}$. A test follows naturally from such a result. The technique is fine tuned and extended to general functional inequalities involving conditional expectations in Lee et al. (2018).

A key assumption in this approach is that the kernel function has bounded support. Otherwise, discretization and Poissonization cannot lead to a one-dependent random field. I illustrate the role that Poissonization plays in the proof of $s_n^{-1}(T_n^{LSW} - a_n) \to \mathcal{N}(0,1)$ in a simplistic example, where $d = 1$ and Z is discrete, in Appendix C.

3 Comparison and Hybridation of Tests

As reviewed in the previous section, for some inequality testing problems, such as conditional moment inequalities, there are quite a few uniformly asymptotically valid tests available in the literature. In some settings, there currently is only one test statistic proposed, that is T_n^{sup}, but new asymptotic theory developed in the literature may allow new tests to be introduced. Then, it is natural to ask what one should do with these options. How should one select a test to use? Should one select just one test or is there a way to combine different tests for better performance?

These questions are intriguing because the hypothesis in (14) is a composite null hypothesis and as such, there is no uniformly most powerful (UMP) test. Moreover, unlike in the context of equalities, the inequality hypothesis is not rotation invariant, and hence it is not obvious that one could restrict attention to a reasonable class of invariant tests and find a UMP one in the class.

In the absence of a UMP test, the relative power of competing tests depends on the specific alternative under which the data are generated. Since the data generating process is unobserved, selecting among tests becomes a classic decision-under-uncertainty problem. Two decision criteria has been suggested in the literature of inequality testing: a weighted average power (WAP) maximization criterion and a Maximin power criterion. I now review some existing work on this topic.

3.1 Existing Optimality Discussion

The WAP maximization criterion is considered in Chiburis (2008) for moment inequality hypotheses and in Elliott et al. (2015) to develop a general framework for a nearly optimal test when testing a null hypothesis in the presence of a nuisance parameter.[10] [11] These two papers do not compare tests in the literature. Instead, they try to design a test that maximizes WAP over all tests that control size.

Assume again that $g_P(t)$ is a scalar for each t. Chiburis's approach discretizes $\mathcal{T}$ into $\{t_1, \ldots, t_K\}$ and the space of $\mathbb{R}^K$ into S regions, and defines a test to be a vector of rejection probabilities on the S regions. He then numerically calculates the WAP of such a test for a given weight over the alternative space, as well as a given null rejection probabilities (NRP) for points on a grid for the null space. Finally, he calculates the optimal test by maximizing the WAP subject to the constraint that the NRPs are all less than or equal to a nominal significance level. The procedure can be computationally intensive or even infeasible when K is larger than a handful because many regions (large S) and many grid points need to be considered in order to get a reasonably good approximation.

Elliott et al. (2015) improve on Chiburis (2008) by making use of the Neyman-Pearson (NP) Lemma. When applied in our context, their approach also discretizes $\mathcal{T}$ and $\{\mu_P \in \mathbb{R}^K : \mu_P \leq 0\}$, but instead of solving a linear programming problem, it uses an iterative procedure to find the least favorable mixed null (a mixure of the distributions in $\{P : (g_P(t_1), \ldots, g_P(t_K)) \leq 0\}$). In each iteration, a NP test is constructed for the weighted mixture null with weights from the previous iteration against a given mixture alternative, and then the mixing weights for the null are updated to give more weights to P's under which the NP test over-rejects and less weights to P's under which the NP test under-rejects. The iteration procedure converges to a least favorable mixed null and the associated NP test is

[10] In inequality testing problems, the slackness parameter $\sqrt{n}g_P(t) : t \in \mathcal{T}$ are the nuisance parameters because their values are not uniquely determined by H_0.

[11] Andrews and Barwick (2012) also uses the concept of WAP maximization to select tuning parameters in tests for a finite number of inequalities.

nearly optimal.

As we can see, the WAP consideration leads to tests different from the tests reviewed in the previous section and thus is not helpful for guiding the selection or combination of those tests. In fact, if one is comfortable with choosing a weight over the alternative space and committed to maximizing the associated WAP, and if the problem is simple enough for the numerical algorithms in Chiburis (2008) or Elliott et al. (2015) to be feasible, one probably should simply use the nearly optimal test.

Often, though, the weight on the alternative space is difficult to interpret in practice.[12] This compounded with the computational burden of the nearly optimal tests limits their practical appeal.

The maximin power rule has been considered in Chetverikov (2018), Armstrong (2018), and Chernozhukov et al. (2019). Chetverikov (2018) studies the conditonal inequality moment hypothesis in (2). He shows that no test with asymptotic size control can have asymptotic power larger than the size uniformly against alternatives belonging to a Hölder ball with smoothness parameter τ and having a sup distance from the null hypothesis at least a_n. Here $a_n = o((\log(n)/n)^{\tau/(2\tau+d_Z)})$ and d_Z is the number of continuous conditioning variables. The intuition for such a result is that there are always least-rejectable Gaussian P_n's satisfying the sup distance requirement that is close to a P_0 in the null; the difference of the expectations of a binary statistic (that is, a test) under the P_n's and under P_0 is bounded by the differences of P_n's and P_0. These differences are small when a_n is small. The least-rejectable P_n's are typically ones such that the null hypothesis is violated on a fast shrinking neighborhood of a Z value as $n \to \infty$.

Chetverikov (2018) shows that an adaptive sup statistic-based test controls asymptotic size and is consistent against any local alternatives such that $a_n \to \infty$. In this sense, the adpative sup statistic-based test is called maximin power rate optimal. Armstrong (2018)

[12]It should be noted that alternatives local to the null hypothesis are the important ones for WAP consideration because most reasonable tests are consistent against fixed alternatives. Yet, data are inherently unable to provide precise information about local alternatives. Thus, we typically cannot rely on data to determine which alternatives are more relevant.

shows that the truly studentized KS test is also maximin power rate optimal and that a CvM test is not. Chernozhukov et al. (2019) consider testing a large number of unstructured inequalities. Using similar techniques, they also derive a maximin rate, and shows that the SUP-statistic based tests achieve this rate. As we can see, the maximin power consideration clearly recommends the truly studentized KS-type test statistic.

However, rate optimality is a crude type of optimality that may be satisfied by many tests that have very different power performance in other apsects. Ignoring the other aspects can be costly. An obvious cost is the power against other alternatives, in particular, those under which the inequalities are violated on a non-shrinking or slowly shrinking set of $\mathcal{T}$ albeit by a small amount at each point. Andrews and Shi (2013) show that the tests that they propose have nontrivial power against some alternatives that converge to H_0 at $1/\sqrt{n}$ rate. The tests based on $T_n^{\rm sup}$ do not have nontrivial power against such alternatives.

Alternatives under which the inequalities are violated on a non-shrinking naturally occur in moment inequality models that contain explicit or implicit equalities. In such models, pairs or groups of inequalities interact and restrict each other to be binding on a nontrivial subset of $\mathcal{T}$. They often are violated at all points in that subset when we move the model parameters away from their identified set. [13] I illustrate this with a simple interval outcome regression example.

Example 5. *Consider a regression model*

$$Y = \beta_0 + \beta_1 X + \varepsilon, \tag{28}$$

where Y *is missing when a missing indicator* $M = 1$*. Suppose that* $Median(\varepsilon|X) = 0$*. Then*

[13]Although explicit or implicit equalities do not always imply point identification, point identified moment inequality models often contain such equalities. See Gandhi et al. (2023) for an example of point identified interval IV regression model, and Shi et al. (2018) for that of a point-identified moment inequality representation of a panel data multinomial choice model.

the model implies the following conditional moment inequality model:

$$m^U(X) := E[1\{Y \leq \beta_0 + \beta_1 X, M = 0\} + 1\{M = 1\}|X] - 0.5 \geq 0$$
$$m^L(X) := E[1\{Y \leq \beta_0 + \beta_1 X\}|X] - 0.5 \leq 0, \tag{29}$$

where the upper bound holds because $0 = E[1\{\varepsilon \leq 0\}|X] - 0.5 = E[1\{\varepsilon \leq 0, M = 0\} + 1\{\varepsilon \leq 0, M = 1\}|X] - 0.5 \leq E[1\{\varepsilon \leq 0, M = 0\} + 1\{M = 1\}|X] - 0.5$. *This is a simplified version of the model considered in Blundell et al. (2007) where* Y *is the wages of female individuals and* X *contain covariates that may affect wages. Missingness is caused by labor market nonparticipation.*

For illustration, let $M = 1\{X + Y < C\}$, *where* C *is the* 10*th percentile of* $X + Y$. *Let* $\beta_0 = \beta_1 = 1$ *and* X *and* ε *be independent* $\mathcal{N}(0, 1)$ *random variables. Note that under this DGP, there is very little missing at large values of* X. *Thus, at those values of* X, *the two bounds are close to forming an equality. This is illustrated in Figure 1. As illustrated in the figure, as* b_1 *moves toward* β_1, *the lower bound tilts toward the horizontal* 0 *line instead of parallel-shift down. Thus, the set of* X *values at which the bound is violated does not shrink in the process. Andrews and Shi (2013) and Andrews and Shi (2014) contain Monte Carlo results that show the power trade-off of different tests in a model similar to this.*

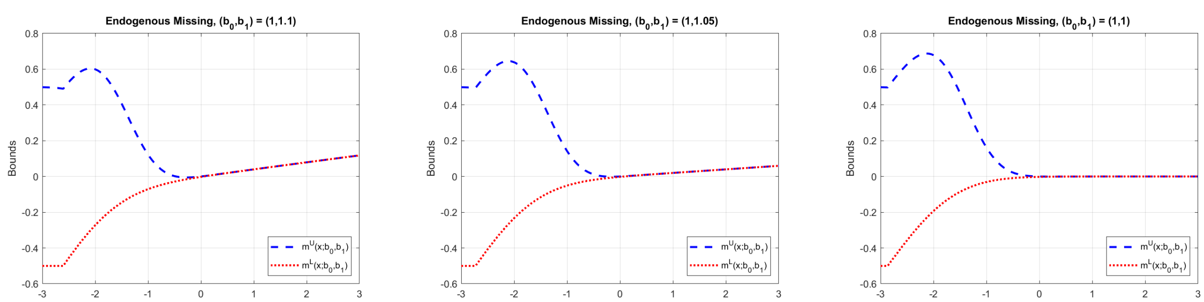

Figure 1: Bounds at Three Values of (b_0, b_1) For Example 5. Drawn by the author - original created for this book.

As the literature currently stands, there is no formal discussion of test comparison and combination that does not require one to commit to a prior distribution on the alternative space (as WAP maximization would) or to restrict attention to the most pessimistic rate (as Maximin power rate optimality would). In the next subsection, I suggest a direction toward such a formal discussion.

3.2 Minimax Power Regret: A Suggested Direction

In decision theory, an alternative decision criterion to maximin is minimax regret. Whereas the maximin criterion leads to the safest option by focusing on the worst-case outcome, the minimax regret criterion seeks to avoid ex post disappointment and may favor options with better upside potential if they limit worst-case regret.

In hypothesis testing, we can define the regret of a test relative to a class of tests with guaranteed (asymptotic) level α. Specifically, let $\Phi = \{\varphi_{s,\alpha} : s \in \mathcal{S}\}$ be the class of tests under consideration where $\mathcal{S}$ is a finite index set. Under a given alternative distribution P_1, we can define the regret of test $\varphi_{s,\alpha}$ relative to Φ as

$$R_\Phi(\varphi_{s,\alpha}, P_1) = \max_{s' \in \mathcal{S}} \mathbb{E}_{P_1}\varphi_{s',\alpha} - \mathbb{E}_{P_1}\varphi_{s,\alpha} \text{ for } s \in \mathcal{S}. \tag{30}$$

Thus, the maximum power regret of test $\varphi_{s,\alpha}$ on the alternative space ($\mathcal{P}_1$) is

$$\overline{R}_\Phi(\varphi_{s,\alpha}) := \max_{P_1 \in \mathcal{P}_1} R_\Phi(\varphi_{s,\alpha}, P_1). \tag{31}$$

The minimax power regret criterion recommends test s if $\overline{R}(\varphi_{s,\alpha}) \leq \overline{R}(\varphi_{s',\alpha})$ for all $s' \in \mathcal{S}$.

Let us consider an example, where K inequalities are tested:

$$g_P(t) \leq 0 \text{ for } t = 1, \ldots, K. \tag{32}$$

Let $\widehat{g}(t) : t = 1, \ldots, K$ be estimators of $g_P(t) : t = 1, \ldots, K$. Suppose that they are independent, and $(\widehat{g}(t) - g_P(t)) \sim \mathcal{N}(0, 1)$. This setting abstracts away from the many challenges in the general case, but suffices for illustrating the potential of the minimax power regret criterion.

Let us consider two test statistics:

$$T^{\text{sum}} = \sum_{t=1}^{K}([\widehat{g}(t))]_+)^2, \text{ and } T^{\text{sup}} = \max_{t=1,\dots,K} \widehat{g}(t), \tag{33}$$

where $[x]_+ = \max\{0, x\}$. Note that since we assume mutual independence of $\{\widehat{g}(t) : t = 1, \dots, K\}$, T^{sum} is the same as the quasi-likelihood ratio statistic: $T^{\text{qlr}} = \min_{h \geq \mathbf{0}}(\widehat{g} - h)'(\widehat{g} - h)$ where $\widehat{g} = (\widehat{g}(1), \dots, \widehat{g}(K))'$.

Let us consider the least favorable critical values:

$$cv_\alpha^{\text{sum}}(K) = F^{-1}_{\sum_{t=1}^{K}[Z_t]_+^2}(1-\alpha) \text{ and } cv_\alpha^{\text{sup}}(K) = F^{-1}_{\max_{t=1,\dots,K} Z_t}(1-\alpha), \tag{34}$$

where $(Z_1, \dots, Z_K)' \sim N(\mathbf{0}, I_K)$ and $F_X^{-1}(1-\alpha)$ stands for the $100(1-\alpha)\%$ quantile of X. Let $\varphi_{\alpha,K}^{\text{sum}} = 1\{T^{\text{sum}} > cv_\alpha^{\text{sum}}(K)\}$ and $\varphi_{\alpha,K}^{\text{sup}} = 1\{T^{\text{sup}} > cv_\alpha^{\text{sup}}(K)\}$.

Suppose that $\mathcal{P}_1$ is rich eough so that $\{g_P = (g_P(1), \dots, g_P(K))' : P \in \mathcal{P}_1\} = \mathbb{R}^K$. That is, we do not have prior information about $(g_P(1), \dots, g_P(K))'$ beyond that is provided by its estimator. The next proposition implies that

$$\liminf_{K\to\infty} \overline{R}_{\{\varphi_{\alpha,K}^{\text{sum}}, \varphi_{\alpha,K}^{\text{sup}}\}}(\varphi_{\alpha,K}^{\text{sum}}) \geq 1-\alpha \text{ and } \liminf_{K\to\infty} \overline{R}_{\{\varphi_{\alpha,K}^{\text{sum}}, \varphi_{\alpha,K}^{\text{sup}}\}}(\varphi_{\alpha,K}^{\text{sup}}) \geq 1-\alpha. \tag{35}$$

That is, both tests have large maximum power regret. Specifically, when only one inequality is violated, in order to have non-trivial power, the SUM test needs the violation to diverge much faster than the SUP test does because the critical value of the SUM test diverges faster than the SUP test as $K \to \infty$. On the other hand, when all the inequalities are violated by a similar amount, the SUP test statistic cannot aggregate all the violations as effectively as the SUM test. As a result, in order to have nontrivial power, the SUP test needs the violation (of each inequality) to be much larger than the SUM test does.

However, we can define a hybrid test based on the two as follows

$$\varphi^{\text{hyb}}_{\alpha,K} = \max\{\varphi^{\text{sum}}_{\alpha/2,K}, \varphi^{\text{sup}}_{\alpha/2,K}\}. \tag{36}$$

That is, the level α hybrid test rejects H_0 when either the level $\alpha/2$ SUP test rejects H_0 or the level $\alpha/2$ SUM test rejects H_0. The proposition also shows that under the sequences of P_1 that either the SUP test or the SUM test has a large power regret, the hybrid test has a zero power regret. In a way, the hybrid test adapts to the data generating process and automatically switches to the more powerful test.[14]

Proposition 3. (a) *Let $\{P_K\}$ be such that $g_{P_K} = a_K \mathbf{1}_K$, where $\mathbf{1}_K$ is a vector of ones, and a_K is a positive scalar sequence such that $\sqrt{\log K} a_K \to 0$ and $\sqrt{K} a_K \to \infty$ as $K \to \infty$. Then, $R_{\{\varphi^{\text{sum}}_{\alpha,K}, \varphi^{\text{sup}}_{\alpha,K}\}}(\varphi^{\text{sup}}_{\alpha,K}, P_K) \to 1 - \alpha$.*

(b) *Let $\{P_K\}$ be such that $g_{P_K}(1) = a_K$ and $g_{P_K}(t) = 0$ for all $t \neq 1$, where $a_K - \sqrt{2\log(K)} \to \infty$ and $a_K = o(K^{1/4})$. Then, $R_{\{\varphi^{\text{sum}}_{\alpha,K}, \varphi^{\text{sup}}_{\alpha,K}\}}(\varphi^{\text{sum}}_{\alpha,K}, P_K) \to 1 - \alpha$.*

(c) *Under each sequence in parts (a) and (b), $R_{\{\varphi^{\text{sum}}_{\alpha,K}, \varphi^{\text{sup}}_{\alpha,K}, \varphi^{\text{hyb}}_{\alpha,K}\}}(\varphi^{\text{hyb}}_{\alpha,K}, P_K) \to 0$.*

I now investigate the implication of the proposition in a simulation exercise. I first simulate these rejection probabilities at h's that have an Euclidean distance of 4 to the null space $\{h \in \mathbb{R}^2 : h \leq 0\}$.[15] In particular, I consider three sets of points on $\mathbb{R}^2$ that are 4-away from the null space:

(a) $g_P(t) = 4/\sqrt{K-1}$ for $t = 2, \ldots, K$, and $g_P(1)$ takes values on a grid on the interval $[-3, 0]$;

[14] It should be noted that hybriding is not without cost. Adjusting the level from α to $\alpha/2$ guarantees level control of the hybrid test, but it reduces power across the board. The power reduction may not be made up by hybriding under some data generating processes not described in Proposition 3. Such cost is expected because of the lack of a UMP test and is part of the reason that criteria like maximin power and minimax regret are useful in this context.

[15] The number 4 is somewhat arbitrarily chosen. Other distances give similar results, with smaller distance showing less dramatic contrasts between the tests.

(b) $g_P(1)$ takes values on a grid on the interval $(0, 4]$, and $g_P(t) = \frac{\sqrt{4 - g_P(1)^2}}{\sqrt{K-1}}$ for $t = 2, \ldots, K$;

(c) $g_P(1) = 4$ and $g_P(2)$ takes values on a grid on the interval $(-3/\sqrt{K-1}, 0)$.

The points can be plotted as a curve on the space of $(g_P(1), \|g_P(II)\|)'$ where $g_P(II) = (g_P(2), \ldots, g_P(K))'$. This curve is shown in the upper left graph in Figure 2. The graph makes it clear that each point on the curve corresponds to the angle that the ray from $(0, 0)$ to $(-\infty, 0)$ needs to rotate (clockwise) to reach the point. Denote that angle γ. The segment on which $\gamma > \pi$ corresponds to the DGPs where only one inequality is violated and the rest are increasingly slack as we increase γ. The segment on which $\gamma \in (\pi/2, \pi)$ corresponds to the DGPs where all inequalities are violated, and the segment on which $\gamma < \pi/2$ corresponds to the DGPs where $J - 1$ inequalites are vioated while one inequality is increasingly slack as γ decreases.

We can plot the rejection probabilities of each test for each point on the curve against γ. These plots are shown in the three other graphs in Figure 2. Consistent with the conclusions of Proposition 3, the SUP test has large regret at small γ's (when a large number of inequalities are violated), while the SUM test has large regret at large γ's (when only one inequality is violated). The hybrid test seems to achieve a balance between the two and has smaller regret than both on most γ's when $K = 5$ and on all γ's when $K = 10$ and 50. The advantage of the hybrid test is more and more clear as K increases. It is also worth noting that the hybrid test not only has small regret, but appears to have the best worst-case power as well.

The power lines are slices on the power surface and thus might not reflect the regrets at all P_1. I also construct the heat map of the regret of each test with respect to the 3-test collection. The heat maps are shown in Figure 3. As we can see from the heat bar on the right of each graph, the maximum regret of the hybrid test is much lower than the maximum regrets of the SUM and the SUP tests.

Proposition 3 and the simulation exercise suggest that the minimax power regret criterion may be a useful tool for test comparison. It has the potential to provide justfication for various test hybridation procedures.

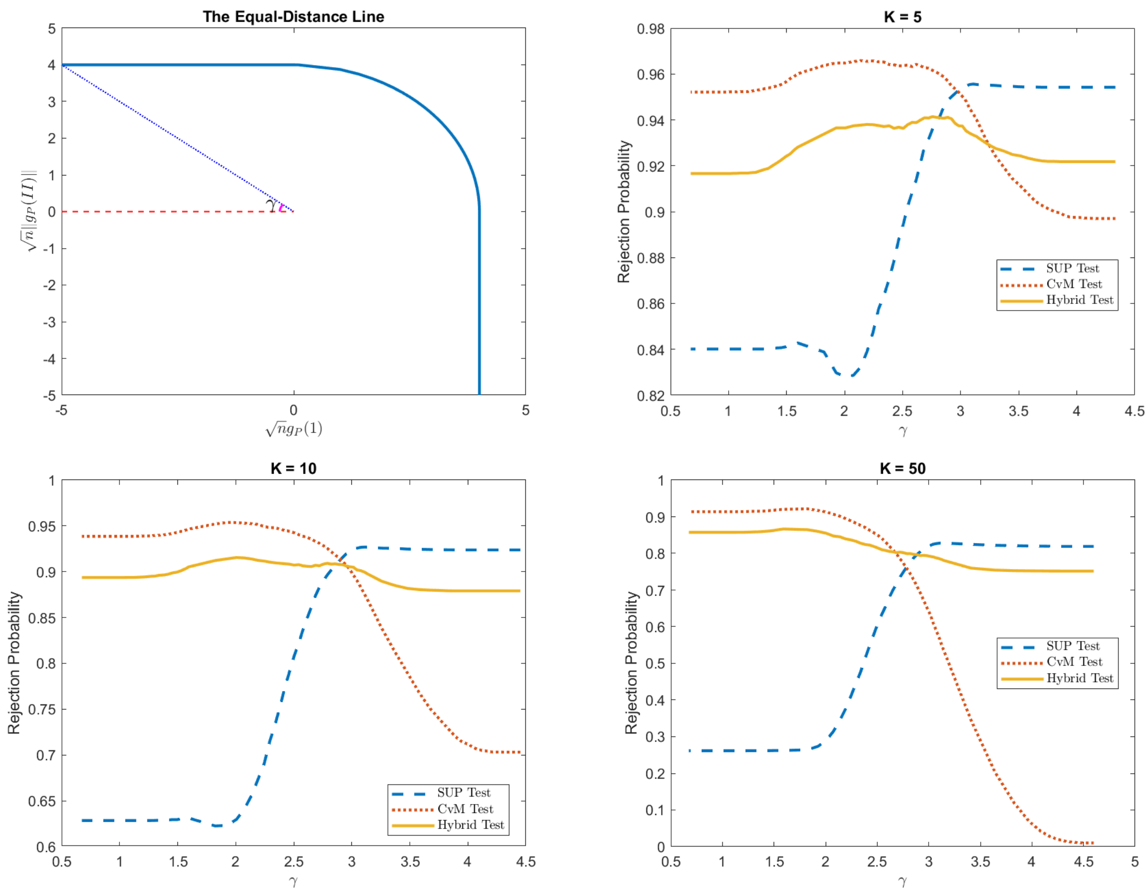

Figure 2: Power Lines for the SUP, SUM, and Hybrid Tests. Drawn by the author - original created for this book.

4 Conclusion

This paper surveys the literature on inference based on infinitely many inequalities. The emphasis is on various testing procedures for infinite dimensional inequality hypotheses. I summarized the two main branches of the literature on testing such hypotheses and explained the differences between their theoretical foundations. I reviewed the two optimality concepts used in the literature, discussed their limitations. I pointed out a few directions for future

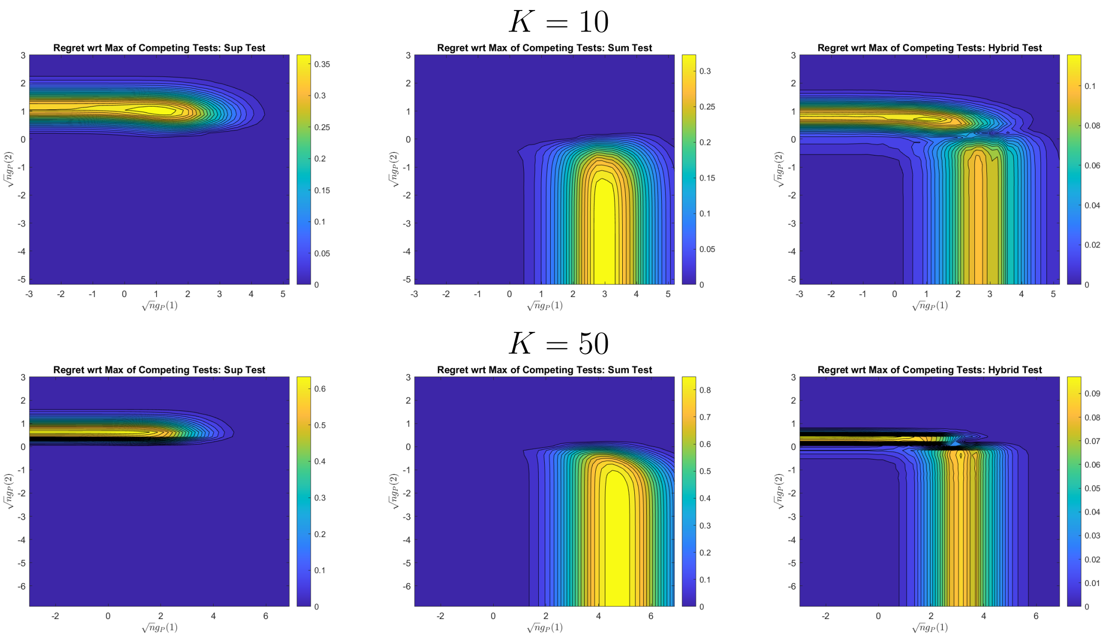

Figure 3: Heat Maps for the regret of the SUP, SUM, and Hybrid Tests against $(g_P(1), g_P(2))$ $(g_P(t) = g_P(2)$ for $t > 2)$. Drawn by the author - original created for this book.

research including new CvM-type tests in Scenario 2 and a new direction for comparing and hybriding tests.

Appendix

A Useful Lemmas

In this section, I present the lemma used to prove the Propositions. We use the notation $[x]_+ = \max\{x, 0\}$.

Lemma 1. *Let $\{Z(t)\}_{t=1}^K$ be i.i.d. $\mathcal{N}(0,1)$ random variables, and let $a_K \to 0$ as $K \to \infty$. Define the function $f_a(z) = [z + a]_+^2$.*

$$S_K := \frac{1}{\sqrt{K}} \sum_{t=1}^{K} (f_{a_K}(Z(t)) - \mathbb{E}[f_{a_K}(Z(t))]) \xrightarrow{d} N(0, 5/4).$$

Proof. Since $a_K \to 0$, for any fixed $z \in \mathbb{R}$, we have

$$f_{a_K}(z) = \max(z + a_K, 0)^2 \to \max(z, 0)^2 := f(z), \tag{37}$$

so $f_{a_K}(z) \to f(z)$ pointwise. Note that for all $a \in \mathbb{R}$,

$$f_a(z)^2 = \max(z + a, 0)^4 \leq (z + a)^4 \leq 8z^4 + 8a^2. \tag{38}$$

Thus,

$$\mathbb{E}[f_{a_K}(Z(t))^2] \leq 8E[Z(t)^4] + 8a_K^4 \to 8E[Z(t)^4] < \infty, \tag{39}$$

since $Z(t) \sim \mathcal{N}(0, 1)$. Therefore, the random variables $f_{a_K}(Z(t))$ have uniformly bounded second moments. This implies the Lindeberg condition holds and ensures uniform integrability.

For each fixed K, the terms $X_t^{(K)} := f_{a_K}(Z(t)) - \mathbb{E}[f_{a_K}(Z(t))]$ are i.i.d. mean-zero random variables with variance $\sigma_K^2 := \mathrm{Var}(f_{a_K}(Z))$ and $a_K \to 0$. Below we show that $\sigma_K^2 \to 5/4$. Hence, by the Lindeberg–Lévy Central Limit Theorem, we have:

$$S_K = \frac{1}{\sqrt{K}} \sum_{t=1}^{K} X_t^{(K)} \xrightarrow{d} N(0, 5/4).$$

Now we show that $\sigma_K^2 \to 5/4$. Since $f_{a_K}(Z) \to f(Z) := [Z]_+^2$ pointwise, and $\{f_{a_K}(Z)\}$ is uniformly integrable in L^2, it follows that $\sigma_K^2 = \mathrm{Var}(f_{a_K}(Z)) \to \mathrm{Var}(f(Z)) = \mathrm{Var}([Z]_+^2) = \frac{5}{4}$. This concludes the proof. □

Lemma 2. *Let $Z \sim \mathcal{N}(0, 1)$, and define $f(a) := \mathbb{E}[\max(Z + a, 0)^2]$. Then the function is twice-continuously differentiable and $f'(0) = \frac{2}{\sqrt{2\pi}}$.*

Proof. We first write

$$f(a) = \mathbb{E}[\max(Z + a, 0)^2] = \int_{-a}^{\infty} (z + a)^2 \phi(z)\, dz, \tag{40}$$

where $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ is the standard normal density. Differentiate under the integral sign using Leibniz's rule:

$$\begin{aligned} f'(a) &= \frac{d}{da} \int_{-a}^{\infty} (z + a)^2 \phi(z)\, dz \\ &= -(z + a)^2 \phi(z)\big|_{z=-a} + \int_{-a}^{\infty} \frac{\partial}{\partial a} (z + a)^2 \phi(z)\, dz \\ &= \int_{-a}^{\infty} 2(z + a)\phi(z)\, dz \end{aligned} \tag{41}$$

Similarly, we can obtain the second derivative: $f''(a) = \int_{-a}^{\infty} 2\phi(z)dz = 2\Phi(a)$, where $\Phi(\cdot)$ stands for the cumulative distribution functin of $\mathcal{N}(0, 1)$. Therefore, the function is twice-continuously differentiable.

To evaluate $f'(0)$, change variables to $u = z + a$, so that $f'(a) = 2 \int_0^{\infty} u \cdot \phi(u - a)du$.

Setting $a = 0$, we obtain:

$$f'(0) = 2 \int_0^{\infty} u\phi(u)du = 2 \cdot \mathbb{E}[Z \cdot \mathbf{1}_{Z>0}] = 2 \int_0^{\infty} z \cdot \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz = \frac{2}{\sqrt{2\pi}}. \tag{42}$$

□

Lemma 3. *Let Z be a standard normal random variable and $\{W_i\}_{i=1}^n$ be a set of i.i.d. random variables such that $\mathbb{E}[W_i] = 0$ and $\mathbb{E}[W_i^2] = 1$. Suppose that $\mathbb{E}[|W_i|^r] < \infty$ for $r \geq 3$. Let $\overline{W}_n = n^{-1} \sum_{i=1}^n W_i$. Then for p such that $1 \leq p < r$, for any $g \leq 0$, and for n such that $n^{-1/2}\mathbb{E}[|W_i|^3] \leq 1$, we have $\left| \mathbb{E}(\left[\sqrt{n}\overline{W}_n + g\right]_+^p) - \mathbb{E}([Z + g]_+^p) \right| \leq$*

$$C_1 n^{-1/2} \mathbb{E}[|W_i|^3] + C_2 n^{-(r-2)/2} \mathbb{E}[|W_i|^r] + C_3 \left(n^{-1/2}\mathbb{E}[|W_i|^3]\right)^p, \tag{43}$$

where C_1 , C_2, and C_3 are positive constants that only depend on r and p.

Proof of Lemma 3. We prove the lemma by verifying the conditions of Lemma A2 in Lee et al. (2013) for the function $\varphi(x) = [x+g]_+^p$. First, we verify that

$$\sup_{x \in R} \frac{|\varphi(x) - \varphi(0)|}{1 + |x|^r \min\{|x|, 1\}} \le 1 < \infty. \tag{44}$$

This is shown by the derivation:

$$\begin{aligned} \sup_{x<1} \frac{|\varphi(x) - \varphi(0)|}{1 + |x|^r \min\{|x|, 1\}} &= \sup_{x<1} \frac{[x+g]_+^p}{1 + |x|^r \min\{|x|, 1\}} \le \sup_{x<1} [x+g]_+^p \le 1, \\ \sup_{x\ge 1} \frac{|\varphi(x) - \varphi(0)|}{1 + |x|^r \min\{|x|, 1\}} &\le \sup_{x\ge 1} \frac{[x+g]_+^p}{1 + x^r} \le \sup_{x \ge 1} x^{p-r} \le 1, \end{aligned} \tag{45}$$

where we used $g \le 0$ and $p < r$.

Under the condition in (44) and the conditions given in Lemma 3, Lemma A2 of Lee et al. (2013) implies that $\left| \mathbb{E}(\left[\sqrt{n}\overline{W}_n + g\right]_+^p) - \mathbb{E}([Z+g]_+^p) \right| \le$

$$c_1\{n^{-1/2}\mathbb{E}[|W_i|^3] + n^{-(r-2)/2} E[|W_i|^r]\} + c_2 \mathbb{E}\left[\omega_\varphi\left(Z; n^{-1/2} c_3 \mathbb{E}[|W_i|^3]\right)\right], \tag{46}$$

where $\omega_\varphi(x; \varepsilon) = \sup_{y:|x-y|\le\varepsilon} |\varphi(x) - \varphi(y)|$, and c_1, c_2, and c_3 only depend on r. Consider the derivation: for any $y \in [x-\varepsilon, x+\varepsilon]$,

$$\begin{aligned} |\varphi(x) - \varphi(y)| = |[x+g]_+^p - [y+g]_+^p| &\le p[x+\varepsilon+g]_+^{p-1}|x-y| \le p[x+\varepsilon]_+^{p-1}\varepsilon \\ &\le p2^{p-1}(|x|^{p-1}\varepsilon + \varepsilon^p). \end{aligned} \tag{47}$$

Thus,

$$\mathbb{E}\omega_\varphi\left(Z; n^{-1/2}c_3\mathbb{E}[|W_i|^3]\right) \le p2^{p-1}\left(\mathbb{E}[|Z|^{p-1}]n^{-1/2}c_3\mathbb{E}[|W_i|^3] + n^{-p/2}c_3^p\mathbb{E}[|W_i|^3]^p\right). \tag{48}$$

Therefore, $\left|\mathbb{E}(\left[\sqrt{n}\overline{W}_n + g\right]_+) - \mathbb{E}([Z+g]_+)\right| \leq$

$$(c_1 + p2^{p-1}c_2c_3\mathbb{E}|Z|^{p-1})n^{-1/2}\mathbb{E}[|W_i|^3] + c_1 n^{-(r-2)/2}E[|W_i|^r] + c_2c_3^p p2^{p-1}n^{-p/2}\mathbb{E}[|W_i|^3]^p. \quad (49)$$

This proves the lemma with $C_1 = (c_1 + p2^{p-1}c_2c_3\mathbb{E}|Z|_+^{p-1})$, $C_2 = c_1$, and $C_3 = c_2c_3^p p2^{p-1}$. □

B Proof of the Propositions

Proof of Proposition 1. Conditions (i) and (ii) imply, via the almost sure representation theorem, that there exists a version of $(r_n(\widehat{g}_n(t) - g_{P_n}(t)), \overline{\Sigma}_n(t)) : t \in \mathcal{T}$ with the same distribution that converges almost surely to a version of $(G(t), \overline{\Sigma}(t)) : t \in \mathcal{T}$. Without loss of generality, assume we are working with this version. Then, almost surely,

$$(r_n(\widehat{g}_n(t) - g_{P_n}(t)), \overline{\Sigma}_n(t)) \to (G(t), \overline{\Sigma}(t)) \quad \text{uniformly over } t \in \mathcal{T}. \quad (50)$$

Fix a sample path along which the convergence in (50) holds and $\sup_{t\in\mathcal{T}} G(t) < \infty$. Consider convergence along this path. Then, by the continuity of S (Condition (iv)) and Condition (iii),

$$S(r_n\widehat{g}_n(t), \overline{\Sigma}_n(t)) \to S(G(t) - h(t), \overline{\Sigma}(t)) \quad \text{for all } t \in \mathcal{T}. \quad (51)$$

Moreover, the uniform convergence of $\overline{\Sigma}_n(\cdot)$ to $\overline{\Sigma}(\cdot)$ and condition (ii) imply that the minimum eigenvalue of $\overline{\Sigma}_n(t)$ is eventually uniformly bounded below by $\varepsilon/2$. This, combined with $g_{P_n}(t) \leq 0$, the non-decreasing property of S in its first argument and the non-increasing property in its second argument (condition (iv)), implies that $S(r_n\widehat{g}_n(t), \overline{\Sigma}_n(t)) \leq S(r_n(\widehat{g}_n(t) - g_{P_n}(t)), \frac{\varepsilon}{2}I)$. By the continuity of $S(\cdot, \frac{\varepsilon}{2}I)$ (via condition (iv)) and $\sup_{t\in\mathcal{T}} G(t) < \infty$, the Heine–Cantor theorem implies that $S(r_n(\widehat{g}_n(t) - g_{P_n}(t), \frac{\varepsilon}{2}I)$ converges uniformly to

$S(G(t), \frac{\varepsilon}{2}I)$. Thus,

$$\sup_{t\in\mathcal{T}} S(r_n\widehat{g}_n(t), \overline{\Sigma}_n(t)) < \sup_{t\in\mathcal{T}} S(G(t), \tfrac{\varepsilon}{2}I) + \varepsilon, \quad \text{eventually.} \tag{52}$$

Hence, the bounded convergence theorem applies and yields

$$\int_{\mathcal{T}} S(r_n\widehat{g}_n(t), \overline{\Sigma}_n(t))\, d\mu(t) \to \int_{\mathcal{T}} S(G(t) - h(t), \overline{\Sigma}(t))\, d\mu(t). \tag{53}$$

This holds for all sample paths where (50) holds, and $\sup_{t\in\mathcal{T}} G(t) < \infty$. Thus, it holds almost surely and therefore also in distribution. This concludes the proof. □

Proof of Proposition 2. Consider the derivation

$$\begin{aligned}
T_n^{\sup} &= \sup_{t\in\mathcal{T}}[\widehat{Z}_n(t) + r_n g_{P_n}(t)/\widehat{\sigma}_n(t)] \\
&\leq \sup_{t\in\mathcal{T}}[Z_n^*(t) + r_n g_{P_n}(t)/\widehat{\sigma}_n(t)] + \sup_{t\in\mathcal{T}} |Z_n^*(t) - \widehat{Z}_n(t)| \\
&\leq \max\left\{\sup_{t\in\mathcal{T}_n}\left[Z_n^*(t) + \frac{r_n g_{P_n}(t)}{\widehat{\sigma}_n(t)}\right], \sup_{t\in\mathcal{T}/\mathcal{T}_n^*}\left[Z_n^*(t) - k_n(\gamma_n)\frac{\sigma_n(t)}{\widehat{\sigma}_n(t)}\right]\right\} + o_p(\delta_n) \\
&\leq \max\left\{\sup_{t\in\mathcal{T}_n} Z_n^*(t), \sup_{t\in\mathcal{T}/\mathcal{T}_n^*}\left[Z_n^*(t) - k_n(\gamma_n)\frac{\sigma_n(t)}{\widehat{\sigma}_n(t)}\right]\right\} + o_p(\delta_n) \\
&\leq \max\left\{T_n^*, \ \sup_{t\in\mathcal{T}} Z_n^*(t) - k_n(\gamma_n) + k_n(\gamma_n)\sup_{t\in\mathcal{T}}\left|\frac{\sigma_n(t)}{\widehat{\sigma}_n(t)} - 1\right|\right\} + o_p(\delta_n),
\end{aligned} \tag{54}$$

where the first inequality holds because $\widehat{Z}_n(t) \leq Z_n^*(t) + \sup_{t\in\mathcal{T}_n} |Z_n^*(t) - \widehat{Z}_n(t)|$, the second inequality holds by $\sup_{t\in\mathcal{T}}[\cdot] = \max\{\sup_{t\in\mathcal{T}_n^*}[\cdot], \sup_{t\in\mathcal{T}/\mathcal{T}_n^*}[\cdot]\}$, the definition of $k_n(\gamma_n)$, and Condition (i), the third inequality holds because $g_{P_n}(t) \leq 0$ for all $t \in \mathcal{T}$ since H_0 holds under P_n, and the fourth inequality uses the triangular inequality and the fact that $\mathcal{T}/\mathcal{T}_n^* \subseteq \mathcal{T}$. Therefore, for any $x \geq 0$,

$$\begin{aligned}
&\Pr(T_n^{\sup} \leq x) \\
&\geq \Pr(T_n^* + o_p(\delta_n) \leq x) - \Pr\left(\sup_{t\in\mathcal{T}} Z_n^*(t) > k_n(\gamma_n) - o_p(\delta_n) - k_n(\gamma_n)\sup_{t\in\mathcal{T}}\left|\frac{\sigma_n(t)}{\widehat{\sigma}_n(t)} - 1\right|\right).
\end{aligned} \tag{55}$$

Note that $\Pr(T_n^* + o_p(\delta_n) \leq x) \geq \Pr(T_n^* \leq x) - \Pr(|T_n^* - x| \leq o_p(\delta_n))$ and that $\Pr(|\sup_{t\in\mathcal{T}_n^*} Z_n^*(t) - x| \leq o_p(\delta_n)) \to 0$ uniformly over $x \in R$ by Condition (ii). Similarly the subtracted term in the above display is bounded above by

$$\Pr\left(\sup_{t\in\mathcal{T}} Z_n^* > k_n(\gamma_n)\right) + \Pr\left(\left|\sup_{t\in\mathcal{T}} Z_n^* - k_n(\gamma_n)\right| < o_p(\delta_n) + k_n(\gamma_n)\sup_{t\in\mathcal{T}}\left|\frac{\sigma_n(t)}{\widehat{\sigma}_n(t)} - 1\right|\right)$$
$$= \Pr\left(\sup_{t\in\mathcal{T}} Z_n^* > k_n(\gamma_n)\right) + o(1), \tag{56}$$

where the equality holds by Condition (ii) and $k_n(\gamma_n)\sup_{t\in\mathcal{T}}\left|\frac{\sigma_n(t)}{\widehat{\sigma}_n(t)} - 1\right| = o_p(\delta_n)$ (Condition (iii)). By the definition of $k_n(\gamma_n)$, $\Pr(\sup_{t\in\mathcal{T}} Z_n^* > k_n(\gamma_n)) \leq \gamma_n$. Thus it is also $o(1)$. Therefore, $\Pr(T_n^{\sup} \leq x) \geq \Pr(T_n^* \leq x) + o(1)$ uniformly over $x \in [0,\infty)$ $\square$

Proof of Proposition 3. (a) It suffices to show that $\lim_{K\to\infty}\max\{\mathbb{E}_{P_K}\varphi_{\alpha,K}^{\sup}, \mathbb{E}_{P_K}\varphi_{\alpha,K}^{\text{sum}}\} = 1$ and that $\lim_{K\to\infty}\mathbb{E}_{P_K}\varphi_{\alpha,K}^{\sup} = \alpha$. Consider the derivation

$$\begin{aligned}
\mathbb{E}_{P_K}\varphi_{\alpha,K}^{\text{sum}} &= \Pr{}_{P_K}\left(\sum_{t=1}^{K}([\widehat{g}(t)]_+)^2 > cv_\alpha^{\text{sum}}(K)\right)\\
&= \Pr{}_{P_K}\left(\sum_{t=1}^{K}([Z(t) + a_K]_+)^2 > cv_\alpha^{\text{sum}}(K)\right)\\
&= \Pr{}_{P_K}\left(W_K(a_K) + e_K > K^{-1/2}(cv_\alpha^{\text{sum}}(K) - K/2)\right),
\end{aligned} \tag{57}$$

where $W_K(a) = K^{-1/2}\sum_{t=1}^{K}([Z(t)+a]_+^2 - \mathbb{E}[Z(t)+a]_+^2)$ and $e_K = K^{1/2}(\mathbb{E}[Z(t)+a_K]_+^2 - 1/2)$. By Lemma 1, we have $W_K(a_K) \to_d N(0, 5/4)$. By Lemma 2,

$$e_K = K^{1/2}\left(\frac{2}{\sqrt{2\pi}}a_K + o(a_K)\right) \tag{58}$$

Also, by the definition of $cv_\alpha^{\text{sum}}(K)$, $K^{-1/2}(cv_\alpha^{\text{sum}} - K/2)$ is the $100(1-\alpha)\%$ quantile of $W_K(0)$. By Lemma 1, $W_K(0) \to_d N(0, 5/4)$. Thus,

$$K^{-1/2}(cv_\alpha^{\text{sum}} - K/2) \to \sqrt{5}z_\alpha/2, \tag{59}$$

where z_α is the $100(1-\alpha)\%$ quantile of $\mathcal{N}(0,1)$.

Therefore, the last line of (57) equals

$$\begin{aligned}
&\Pr{}_{P_K}\Big((a_K K^{1/2})^{-1} W_K(a_K) + 2/\sqrt{2\pi} + o(1) > 5(a_K K^{1/2})^{-1} z_\alpha/2\Big) \\
&= \Pr{}_{P_K}\Big(o_p(1) + 2/\sqrt{2\pi} + o(1) > o(1)\Big) \to 1, \qquad (60)
\end{aligned}$$

where the equality holds because $\sqrt{K} a_K \to \infty$.

For $\varphi^{\sup}_{\alpha,K}$, consider the derivation

$$\begin{aligned}
\mathbb{E}_{P_K}\varphi^{\sup}_{\alpha,K} &= \Pr{}_{P_K}\left(\max_{t=1,\dots,T}\widehat{g}(t) > cv^{\sup}_\alpha(K)\right) \\
&= \Pr{}_{P_K}\left(\max_{t=1,\dots,K}(Z(t) + g_{P_K}(t)) > cv^{\sup}_\alpha(K)\right) \\
&= \Pr{}_{P_K}\left(a_K + \max_{t=1,\dots,K} Z(t) > cv^{\sup}_\alpha(K)\right). \qquad (61)
\end{aligned}$$

By the Gaussian extreme value theorem, $b_K(\max_{t=1,\dots,K} Z(t) - c_K) \to_d G$, where G has the cumulative distribution function $\exp(-\exp(-x))$, $b_K = \sqrt{2\log K}$ and $c_K = \sqrt{2\log(K)} - \frac{(\log\log(K)+\log(4\pi))}{2\sqrt{2\log(K)}}$. Thus, the last line of (61) is equal to,

$$\Pr{}_{P_K}\left(b_K a_K + b_K(\max_{t=1,\dots,K} Z(t) - c_K) > b_K(cv^{\sup}_\alpha(K) - c_K)\right). \qquad (62)$$

By design, $b_K a_K \to 0$. Thus,

$$b_K a_K + b_K(\max_{t=1,\dots,T} Z(t) - c_K) \to_d G. \qquad (63)$$

Also by the definition of $cv^{\sup}_\alpha(K)$, we have that $b_K(cv^{\sup}_\alpha(K) - c_K)$ is the $100(1-\alpha)\%$ quantile of $b_K(\max_{t=1,\dots,T} Z(t) - c_K)$. The distribution of G is continuous and strictly increasing. Thus, $b_K(cv^{\sup}_\alpha(K) - c_K)$ converges in probability to the $100(1-\alpha)\%$ quantile of G. This and (63) together implies that $\mathbb{E}_{P_K}\varphi^{\sup}_{\alpha,K} \to \alpha$.

(b) It is sufficient to show that $\lim_{K\to\infty} \max\{\mathbb{E}_{P_K}\varphi^{\text{sup}}_{\alpha,K}, \mathbb{E}_{P_K}\varphi^{\text{sum}}_{\alpha,K}\} = 1$ and that $\lim_{K\to\infty} \mathbb{E}_{P_K}\varphi^{\text{sum}}_{\alpha,K} = \alpha$. Consider the derivation: $\mathbb{E}_{P_K}\varphi^{\text{sup}}_{\alpha,K} =$

$$
\begin{aligned}
&\Pr{}_{P_K}\left(\max\{Z(1) + a_K, \max_{t=2,\dots,K} Z(t)\} > cv^{\text{sup}}_\alpha(K)\right) \\
&= \Pr{}_{P_K}\left(\max\{(Z(1) + a_K - c_K), (\max_{t=2,\dots,K} Z(t) - c_K)\} > (cv^{\text{sup}}_\alpha(K) - c_K)\right) \\
&\geq \Pr{}_{P_K}\left(Z(1) + a_K - c_K > (cv^{\text{sup}}_\alpha(K) - c_K)\right). \qquad (64)
\end{aligned}
$$

In part (a), we argue that $b_K(cv^{\text{sup}}_\alpha(K) - c_K)$ converges in probability to the $100(1-\alpha)$ quantile of G, where $b_K = \sqrt{2\log(K)}$. Thus,

$$
cv^{\text{sup}}_\alpha(K) - c_K \to_p 0. \qquad (65)
$$

This and $a_K - \sqrt{2\log K} \to \infty$ together imply that the last line of (64) converges to 1. Thus, $\mathbb{E}_{P_K}\varphi^{\text{sup}}_{\alpha,K} \to 1$.

For $\varphi^{\text{sum}}_{\alpha,K}$, consider the derivation: $\mathbb{E}_{P_K}\varphi^{\text{sum}}_{\alpha,K} =$

$$
\begin{aligned}
&\Pr{}_{P_K}\left([Z(1) + a_K]_+^2 + \sum_{t=2}^{K}[Z(t)]_+^2 > cv^{\text{sum}}_\alpha(K)\right) \qquad (66) \\
&= \Pr{}_{P_K}\left(\frac{[Z(1) + a_K]_+^2}{\sqrt{K}} - \frac{1}{2\sqrt{K}} + \frac{\sqrt{K-1}}{\sqrt{K}} W_{K-1}(0) > \frac{1}{\sqrt{K}}\left(cv^{\text{sum}}_\alpha(K) - \frac{K}{2}\right)\right),
\end{aligned}
$$

Since $a_K = o(K^{1/4})$ and $Z(1) \sim \mathcal{N}(0,1)$, we have that $\frac{[Z(1)+a_K]_+^2}{\sqrt{K}} = o_p(1)$. In the proof of part (a), we have shown that $W_{K-1}(0) \to_d N(0, 5/4)$ and $\frac{1}{\sqrt{K}}\left(cv^{\text{sum}}_\alpha(K) - \frac{K}{2}\right) \to \sqrt{5}z_\alpha/2$. Combining these facts, we deduce that the last line of (66) converges to

$$
\Pr(N(0, 5/4) > \sqrt{5}z_\alpha/2) = \alpha. \qquad (67)
$$

Therefore, $\mathbb{E}_{P_K}\varphi^{\text{sum}}_{\alpha,K} \to \alpha$, concluding the proof of part (b).

(c) It suffices to show that $\lim_{K\to\infty} \mathbb{E}_{P_K}\varphi^{\text{hyb}}_{\alpha,K} = 1$ under each of the sequences in part (a) and in part (b).

Consider a sequence satisfying the conditions in part (a). Then, by the arguments in the proof of part (a),

$$\lim_{K\to\infty} \mathbb{E}_{P_K}\varphi^{\text{sum}}_{\alpha/2,K} = 1. \tag{68}$$

No arguments in the proof there needs to be changed except that α is replaced by $\alpha/2$. By definition, $\mathbb{E}_{P_K}\varphi^{\text{hyb}}_{\alpha,K} \geq \mathbb{E}_{P_K}\varphi^{\text{sum}}_{\alpha/2,K}$. Therefore, $\mathbb{E}_{P_K}\varphi^{\text{hyb}}_{\alpha,K} \to 1$.

Consider a sequence satisfying the conditions in part (b). Then the arguments in the proof of that part show that

$$\lim_{K\to\infty} \mathbb{E}_{P_K}\varphi^{\text{sup}}_{\alpha/2,K} = 1. \tag{69}$$

No modificaiton to the arguments is needed except that α is replaced by $\alpha/2$. Moreover, by definition, $\mathbb{E}_{P_K}\varphi^{\text{hyb}}_{\alpha,K} \geq \mathbb{E}_{P_K}\varphi^{\text{sup}}_{\alpha/2,K}$. Therefore, $\mathbb{E}_{P_K}\varphi^{\text{hyb}}_{\alpha,K} \to 1$. □

C Illustration of the Poissonization Approach

Although LSW address the case where the conditioning variable is continuous, the idea of Poissonization can be most cleanly illustrated in the case of a discrete X. I follow LSW to consider the inequality testing problem and hence use the one-sided L_1 norm below. It is worth noting that the technique originally was designed for equality testing using the two-sided L_1 norm by Giné et al. (2003).

Consider the model $g_P(t) \leq 0$ for all $t \in \mathcal{T}_n$, where

$$g_P(t) = \mathbb{E}_P[Y|X = t], \tag{70}$$

Y is a scalar random variable, and $\mathcal{T}_n = \{t_1, t_2, \ldots, t_{k_n}\}$, where $k_n \to \infty$ as $n \to \infty$. Let

$[y]_+ = \max\{y, 0\}$.

Let $\{Y_i, X_i\}_{i=1}^n$ be an i.i.d. sample from P. Consider the test statistic

$$T_n^{L_1} = \sum_{j=1}^{k_n} [\widehat{\mathbb{E}}_n[Y|X = t_j]]_+ \widehat{p}_j, \tag{71}$$

where $\widehat{p}_j = n^{-1} \sum_{i=1}^n 1\{X_i = t_j\}$, and $\widehat{\mathbb{E}}_n[Y|X = t_j] = \widehat{p}_j^{-1} n^{-1} \sum_{i=1}^n Y_i 1\{X_i = t_j\}$.

It is easy to see that $T_n^{L_1}$ can be written as

$$T_n^{L_1} = \sum_{j=1}^{k_n} \left[n^{-1} \sum_{i=1}^n Y_i 1\{X_i = t_j\} \right]_+ . \tag{72}$$

Note that the k_n values of X_i split the sample into k_n subsamples, and each j summand in the above expression is a function of one of the subsamples. With a structure like this, we can use a Poissonization technique to make the summands independent across j. That then allows us to use central limit theorems to analyze its asymptotic distribution. This is the approach Lee et al. (2013) and Lee et al. (2018) take. We illustrate this approach below.

We prove the following proposition that establishes asymptotic normality of $T_n^{L_1}$ under a sequence of data generating processes $\{P_n\}_{n=1}^\infty$ such that $g_{P_n}(t) \leq 0$ for all $t \in \mathcal{T}_n$. To state the result, define $\mu_{nt} = \mathbb{E}_{P_n}[Y|X = t]$, $p_{nt} = \Pr_{P_n}(X = t)$, and $v_{nt}^2 = \mathrm{Var}_{P_n}(Y 1\{X = t\})$. Let Z denote a $\mathcal{N}(0, 1)$ random variable and N denote a Poisson random variable with mean n. Let Z and N be independent. Let

$$\begin{aligned} a_n &= n^{-1} \sum_{j=1}^{k_n} v_{nt_j} \mathbb{E}(N^{1/2}[Z + \sqrt{N} \mu_{nt_j} p_{nt_j} / v_{nt_j}]_+), \\ s_n^2 &= n^{-2} \sum_{j=1}^{k_n} v_{nt_j}^2 \mathrm{Var}(N^{1/2}[Z + \sqrt{N} \mu_{nt_j} p_{nt_j} / v_{nt_j}]_+), \end{aligned} \tag{73}$$

Let $\kappa_{3nt} = \mathbb{E}_{P_n}[|Y|^3|X = t]$.

Proposition 4. *Suppose that $\kappa_{3nt} < \infty$ and $v_{nt} > 0$ for all n, t and that $\sup_n \mathbb{E}_{P_n} Y^2 < \infty$. Also suppose that there exists a $u \in (0, 1)$ such that the following conditions hold*

(i) $n^{-1}\sum_{j=1}^{k_n} v_{nt_j}^{-2}\kappa_{3nt_j}p_{nt_j} = o(s_n)$,

(ii) $n^{-2}\sum_{j=1}^{k_n}(n^{1/2}v_{nt_j}^{-1}\kappa_{3nt_j}p_{nt_j} + v_{nt_j}^{-4}\kappa_{3nt_j}^2 p_{nt_j}^2) = o(s_n^2)$,

(iii) $n^{-2-u}\sum_{j=1}^{k_n} v_{nt_j}^{2+u}\mathbb{E}(N^{1+u/2}[Z + N^{1/2}\mu_{nt_j}p_{nt_j}/v_{nt_j}]_+^{2+u}) = o(s_n^{2+u})$

(iv) $n^{-(3+u)/2}\sum_{j=1}^{k_n} v_{nt_j}^{u-1}\kappa_{3nt_j}p_{nt_j} + n^{-2-u}\sum_{j=1}^{k_n} v_{nt_j}^{-2(2+u)}\kappa_{3nt_j}^{2+u}p_{nt_j}^{2+u} = o(s_n^{2+u})$,

(v) $n^{-3/2} + n^{-1}|\mathbb{E}_{P_n}Y|^2 = o(s_n^2)$.

Then under $\{P_n\}_{n=1}^{\infty}$ such that $g_{P_n}(t) \leq 0$ for all $t \in \mathcal{T}_n$, as $n \to \infty$,

$$(T_n^{L_1} - a_n)/s_n \to_d \mathcal{N}(0,1). \tag{74}$$

Proof. For notational simplicity, let $\xi_{nt} = \sum_{i=1}^n Y_i 1\{X_i = t\}$. Then

$$T_n^{L_1} = n^{-1}\sum_{j=1}^{k_n}[\xi_{nt}]_+. \tag{75}$$

Across j, ξ_{nt_j} is not independent due to the dependence of the events $X_i = t_j$ across j. However, let N be a Poisson random variable with mean n that is indepdent of $\{Y_i, X_i\}_{i=1}^n$, and let the Poissonized version of ξ_{nt} be

$$\xi_{Nt} = \sum_{i=1}^{N}(Y_i 1\{X_i = t\}). \tag{76}$$

Then $\xi_{Nt_j} : j = 1,\ldots,k_n$ are independent across j. To see this, let $\{N_1,\ldots,N_{k_n}\}$ be independent random variables where N_j follows the Poisson distribution with mean np_{nt_j} for each $j = 1,\ldots,k_n$. Let $Y_i^{(t_j)} : i = 1,\ldots,\infty, j = 1,\ldots,k_n$ be independent across (i,j) and $Y_i^{(t_j)}$ follows the conditional distribution of Y_i given $X_i = t_j$. Then the random vector $(\xi_{Nt_1},\ldots,\xi_{Nt_{k_n}})'$ has the same distribution as that of $\left(\sum_{i=1}^{N_1} Y_i^{(t_1)},\ldots,\sum_{i=1}^{N_{t_{k_n}}} Y_i^{(t_{k_n})}\right)'$. Since $\sum_{i=1}^{N_j} Y_i^{(t_j)} : j = 1,\ldots,k_n$ are independent across j, so are $\xi_{Nt_j} : j = 1,\ldots,k_n$.

Using ξ_{Nt_j}, define the Poissonized version of $T_n^{L_1}$ as:

$$\widetilde{T}_n^{L_1} = n^{-1} \sum_{j=1}^{k_n} [\xi_{Nt_j}]_+. \tag{77}$$

We show that $\widetilde{T}_n^{L_1}$ is asymptotically normal by checking the conditions of the Lindeberg central limit theorem for the sum across j.

First, note that

$$\mathbb{E}_{P_n}[\xi_{Nt}]_+ = v_{nt}\mathbb{E}\left\{ N^{1/2}\mathbb{E}_{P_n}\left(\left[\sqrt{N}\overline{W}_N + g_{Nt}\right]_+ \middle| N \right) \right\}. \tag{78}$$

where $g_{Nt} = \sqrt{N}\mu_{nt}p_{nt}/v_{nt}$, $\overline{W}_N = N^{-1}\sum_{i=1}^N W_i$ and $W_i = (Y_i 1\{X_i = t\} - \mu_{nt}p_{nt})/v_{nt}$. By definition, $\mathbb{E}_{P_n}[W_i] = 0$ and $\mathbb{E}_{P_n}[W_i^2] = 1$. Also,

$$\mathbb{E}_{P_n}[|W_i|^3] \leq 4v_{nt}^{-3}(\kappa_{3nt}p_{nt} + |\mu_{nt}|^3 p_{nt}^3) \leq 8v_{nt}^{-3}\kappa_{3nt}p_{nt} < \infty. \tag{79}$$

Thus, apply Lemma 3 with $r = 3$ and $p = 1$, we have

$$\begin{aligned} \left| \mathbb{E}_{P_n}\left(\left[\sqrt{N}\overline{W}_N + g_{Nt}\right]_+ \middle| N \right) - \mathbb{E}([Z + g_{Nt}]_+ | N) \right| &\leq N^{-1/2}C\mathbb{E}_{P_n}[|W_i|^3] \\ &\leq 8CN^{-1/2}v_{nt}^{-3}\kappa_{3nt}p_{nt}, \end{aligned} \tag{80}$$

where C is a constant that does not depend on n or t. Therefore,

$$|\mathbb{E}_{P_n}[\xi_{Nt}]_+ - v_{nt}[\mathbb{E}(N^{1/2}[Z + g_{Nt}]_+)]| \leq 8Cv_{nt}^{-2}\kappa_{3nt}p_{nt}. \tag{81}$$

Thus, by condition (i)

$$\mathbb{E}_{P_n}\widetilde{T}_n^{L_1} = a_n + O\left(n^{-1}\sum_{j=1}^{k_n} v_{nt_j}^{-2}\kappa_{3nt_j}p_{nt_j} \right) = a_n + o(s_n). \tag{82}$$

Similarly to (80), apply Lemma 3 with $r = 3$ and $p = 2$, and we can derive:

$$\begin{aligned}
&\left|\mathbb{E}_{P_n}\left(\left[\sqrt{N}\overline{W}_N + g_{Nt}\right]_+^2 |N\right) - \mathbb{E}([Z + g_{Nt}]_+^2|N)\right| \\
&\leq (C_1 + C_2)N^{-1/2}\mathbb{E}_{P_n}[|W_i|^3] + C_3(N^{-1/2}\mathbb{E}_{P_n}[|W_i|^3])^2. \qquad (83)
\end{aligned}$$

where C_1, C_2, C_3 do not depend on n or t. Thus

$$\begin{aligned}
&\left|\mathbb{E}_{P_n}([\xi_{Nt}]_+^2) - v_{nt}^2\mathbb{E}\left\{N\mathbb{E}\left([Z + g_{Nt}]_+^2|N\right)\right\}\right| \\
&\leq (C_1 + C_2)\mathbb{E}[N^{1/2}]v_{nt}^2\mathbb{E}_{P_n}|W_i|^3 + C_3v_{nt}^2(\mathbb{E}_{P_n}|W_i|^3)^2. \qquad (84)
\end{aligned}$$

Combining this and (81), we get

$$\begin{aligned}
&\left|\mathrm{Var}_{P_n}([\xi_{Nt}]_+) - v_{nt}^2\mathrm{Var}(N^{1/2}[Z + g_{Nt}]_+)\right| \\
&\leq (C_1 + C_2)\mathbb{E}[N^{1/2}]v_{nt}^2\mathbb{E}_{P_n}|W_i|^3 + C_3v_{nt}^2(\mathbb{E}_{P_n}|W_i|^3)^2 + \\
&\quad 8Cv_{nt}^{-2}\kappa_{3nt}p_{nt}(8Cv_{nt}^{-2}\kappa_{3nt}p_{nt} + v_{nt}[\mathbb{E}(N^{1/2}[Z + g_{Nt}]_+)]) \\
&\leq 8(C_1 + C_2)n^{1/2}v_{nt}^{-1}\kappa_{3nt}p_{nt} + 64C_3v_{nt}^{-4}\kappa_{3nt}^2p_{nt}^2 + \\
&\quad 64C^2v_{nt}^{-4}\kappa_{3nt}^2p_{nt}^2 + 8Cv_{nt}^{-1}\kappa_{3nt}p_{nt}n^{1/2}\mathbb{E}[Z]_+ \\
&= O(n^{1/2}v_{nt}^{-1}\kappa_{3nt}p_{nt} + v_{nt}^{-4}\kappa_{3nt}^2p_{nt}^2), \qquad (85)
\end{aligned}$$

where the second inequality used $[Z + g_{Nt}]_+ \leq [Z]_+$ (since $\mu_{nt} \leq 0$ under H_0) and $\mathbb{E}[N^{1/2}] \leq \sqrt{\mathbb{E}[N]} = n^{1/2}$. Therefore,

$$\begin{aligned}
\mathrm{Var}_{P_n}(\widetilde{T}_n^{L_1}) &= n^{-2}\sum_{j=1}^{k_n}\mathrm{Var}_{P_n}([\xi_{Nt_j}]_+) \\
&= n^{-2}\sum_{j=1}^{k_n}v_{nt_j}^2\mathrm{Var}(N^{1/2}[Z + g_{Nt_j}]_+) + O\left(n^{-2}\sum_{j=1}^{k_n}(n^{1/2}v_{nt_j}^{-1}\kappa_{3nt_j}p_{nt_j} + v_{nt_j}^{-4}\kappa_{3nt_j}^2p_{nt_j}^2)\right) \\
&= s_n^2 + O\left(n^{-2}\sum_{j=1}^{k_n}(n^{1/2}v_{nt_j}^{-1}\kappa_{3nt_j}p_{nt_j} + v_{nt_j}^{-4}\kappa_{3nt_j}^2p_{nt_j}^2)\right) = s_n^2 + o(s_n^2), \qquad (86)
\end{aligned}$$

where the last equality holds by condition (ii).

Also apply Lemma 3 with $r = 3$ and $p = 2 + u$ for $u \in (0, 1)$, and we have

$$\begin{aligned} &|\mathbb{E}_{P_n}([\sqrt{N}\overline{W}_N + g_{Nt}]_+^{2+u}|N) - \mathbb{E}([Z + g_{Nt}]_+^{2+u}|N)]| \\ &\le (C_1 + C_2)N^{-1/2}\mathbb{E}_{P_n}[|W_i|^3] + C_3(N^{-1/2}\mathbb{E}_{P_n}[|W_i|^3])^{2+u}. \end{aligned} \tag{87}$$

Then, using $\xi_{Nt} = v_{nt}N^{1/2}[\sqrt{N}\overline{W}_N + g_{Nt}]_+$, we have

$$\begin{aligned} &\left|\mathbb{E}_{P_n}([\xi_{Nt}]_+^{2+u}) - v_{nt}^{2+u}\mathbb{E}\left\{N^{1+u/2}\mathbb{E}\left([Z + g_{Nt}]_+^{2+u}|N\right)\right\}\right| \\ &\le (C_1 + C_2)\mathbb{E}[N^{(1+u)/2}]v_{nt}^{2+u}\mathbb{E}_{P_n}|W_i|^3 + C_3 v_{nt}^{2+u}(\mathbb{E}_{P_n}|W_i|^3)^{2+u} \\ &= O(n^{(1+u)/2}v_{nt}^{u-1}\kappa_{3nt}p_{nt} + v_{nt}^{-2(2+u)}\kappa_{3nt}^{2+u}p_{nt}^{2+u}). \end{aligned} \tag{88}$$

Then, for any $\varepsilon > 0$, and $u \in (0, 1)$,

$$\begin{aligned} &n^{-2}\sum_{j=1}^{k_n}\mathbb{E}_{P_n}\{[\xi_{Nt_j}]_+^2 1\{n^{-1}[\xi_{Nt_j}]_+ > \varepsilon s_n\} \\ &\le n^{-2-u}s_n^{-u}\varepsilon^{-u}\sum_{j=1}^{k_n}\mathbb{E}_{P_n}[\xi_{Nt_j}]_+^{2+u} \\ &= n^{-2-u}s_n^{-u}\varepsilon^{-u}\sum_{j=1}^{k_n}v_{nt_j}^{2+u}\mathbb{E}(N^{1+u/2}[Z + g_{Nt_j}]_+^{2+u}) \\ &\qquad + O\left(n^{-2-u}s_n^{-u}\sum_{j=1}^{k_n}(n^{(1+u)/2}v_{nt_j}^{u-1}\kappa_{3nt_j}p_{nt_j} + v_{nt_j}^{-2(2+u)}\kappa_{3nt_j}^{2+u}p_{nt_j}^{2+u})\right) \\ &= \varepsilon^{-u}s_n^{-u}o(s_n^{2+u}) + O\left(s_n^{-u}\sum_{j=1}^{k_n}(n^{-(3+u)/2}v_{nt_j}^{u-1}\kappa_{3nt_j}p_{nt_j} + n^{-2-u}v_{nt_j}^{-2(2+u)}\kappa_{3nt_j}^{2+u}p_{nt_j}^{2+u})\right) \\ &= \varepsilon^{-u}s_n^{-u}o(s_n^{2+u}) + o(s_n^{-u}s_n^{2+u}) = o(s_n^2), \end{aligned} \tag{89}$$

where the first inequality holds by Markov's inequality, the first equality holds by (88), the second equality holds by condition (iii) and the third equality holds by condition (iv). This verifies the Lindeberg condition. Thus,

$$s_n^{-1}(\widetilde{T}_n^{L_1} - a_n) = \frac{(\mathbb{E}_{P_n}[\widetilde{T}_n^{L_1}] - a_n)}{s_n} + \frac{\sqrt{\mathrm{Var}_{P_n}(\widetilde{T}_n^{L_1})}}{s_n} \frac{\widetilde{T}_n - \mathbb{E}_{P_n}[\widetilde{T}_n^{L_1}]}{\sqrt{\mathrm{Var}_{P_n}(\widetilde{T}_n^{L_1})}}$$

$$= o(1) + (1 + o(1)) \frac{\widetilde{T}_n - \mathbb{E}_{P_n}[\widetilde{T}_n^{L_1}]}{\sqrt{\mathrm{Var}_{P_n}(\widetilde{T}_n^{L_1})}} \to_d \mathcal{N}(0,1). \tag{90}$$

Now we dePoissonize $\widetilde{T}_n^{L_1}$, that is, we show that $T_n^{L_1}$ and $\widetilde{T}_n^{L_1}$ are close:

$$|T_n^{L_1} - \widetilde{T}_n^{L_1}| \le n^{-1} \sum_{j=1}^{k_n} |[\xi_{Nt_j}]_+ - [\xi_{nt_j}]_+|$$

$$\le n^{-1} \sum_{j=1}^{k_n} \left| \sum_{i=n\wedge N+1}^{n\vee N} [Y_i 1\{X_i = t_j\} - \mu_{nt_j} p_{nt_j}] \right| - n^{-1} \sum_{j=1}^{k_n} |N - n| \mu_{nt_j} p_{nt_j}$$

$$\le n^{-1} \sqrt{\sum_{j=1}^{k_n} \left| \sum_{i=n\wedge N+1}^{n\vee N} [Y_i 1\{X_i = t_j\} - \mu_{nt_j} p_{nt_j}] \right|^2} - n^{-1}|N - n| \mathbb{E}_{P_n}[Y], \tag{91}$$

where the second inequality holds by the triangle inequality where the minus sign in front of the second summand is used because $\mu_{nt_j} \le 0$ under H_0. Now note that $\mathbb{E}|N - n| \le \sqrt{\mathrm{Var}(N)} = n^{1/2}$. Also note that

$$\mathbb{E}_{P_n} \sum_{j=1}^{k_n} \left| \sum_{i=n\wedge N+1}^{n\vee N} [Y_i 1\{X_i = t_j\} - \mu_{nt_j} p_{nt_j}] \right|^2 = \sum_{j=1}^{k_n} \mathbb{E} \left[\sum_{i=n\wedge N+1}^{n\vee N} \mathrm{Var}_{P_n}(Y_i 1\{X_i = t_j\}) \right]$$

$$\le \sum_{j=1}^{k_n} \mathbb{E} \left[\sum_{i=n\wedge N+1}^{n\vee N} \mathbb{E}_{P_n}[Y_i^2 1\{X_i = t_j\}] \right]$$

$$= \mathbb{E}|N - n| \mathbb{E}_{P_n}[Y^2] = O(n^{1/2}). \tag{92}$$

Therefore, $|T_n^{L_1} - \widetilde{T}_n^{L_1}| = O(n^{-3/4}) + O(n^{-1/2}|\mathbb{E}_{P_n} Y|)$. Thus $s_n^{-1}(T_n^{L_1} - \widetilde{T}_n^{L_1}) = O((n^{-3/4} + n^{-1/2}|\mathbb{E}_{P_n} Y|) s_n^{-1} = o(1)$. This concludes the proof. $\square$

References

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

Andrews, D. W. K. and Barwick, P. J. (2012). Inference for parameters defined by moment inequalities: A recommended moment selection procedure. *Econometrica*, 80(6):2805–2826.

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

Andrews, D. W. K. and Shi, X. (2014). Nonparametric inference based on conditional moment inequalities. *Journal of Econometrics*, 179(1):31–45.

Andrews, D. W. K. and Shi, X. (2017). Inference based on many conditional moment inequalities. *Journal of Econometrics*, 196(2):275–287.

Armstrong, T. B. (2015). Asymptotically exact inference in conditional moment inequality models. *Journal of Econometrics*, 186(1):51–65.

Armstrong, T. B. (2018). On the choice of test statistic for conditional moment inequalities. *Journal of Econometrics*, 203(2):241–255.

Bai, Y., Santos, A., and Shaikh, A. M. (2022). A two-step method for testing many moment inequalities. *Journal of Business & Economic Statistics*, 40(3):1070–1080.

Barrett, G. F. and Donald, S. G. (2003). Consistent tests for stochastic dominance. *Econometrica*, 71(1):71–104.

Beare, B. K. and Moon, J.-M. (2015). Nonparametric tests of density ratio ordering. *Econometric Theory*, 31(3):471–492.

Beare, B. K. and Shi, X. (2019). An improved bootstrap test of density ratio ordering. *Econometrics and Statistics*, 10:9–26.

Beresteanu, A., Molchanov, I., and Molinari, F. (2011). Sharp identification regions in models with convex moment predictions. *Econometrica*, 79(6):1785–1821.

Blundell, R., Gosling, A., Ichimura, H., and Meghir, C. (2007). Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica*, 75(2):323–363.

Canay, I., Illanes, G., and Velez, A. (2023). A user's guide for inference in models defined by moment inequalities. *Journal of Econometrics*, forthcoming.

Canay, I. A. and Shaikh, A. M. (2017). Practical and theoretical advances in inference for partially identified models. In Honoré, B., Pakes, A., Piazzesi, M., and Samuelson, L., editors, *Advances in Economics and Econometrics: Eleventh World Congress*, pages 271–306. Econometric Society Monographs, Cambridge University Press.

Carolan, C. A. and Tebbs, J. M. (2005). Nonparametric tests for and against likelihood ratio ordering in the two-sample problem. *Biometrika*, 92(1):159–171.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014a). Anti-concentration and honest, adaptive confidence bands. *Annals of Statistics*, 42(5):1787–1818.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014b). Gaussian approximation of suprema of empirical processes. *Annals of Statistics*, 42(4):1564–1597.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields*, 162(1-2):47–70.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2017). Central Limit Theorems and Bootstrap in High Dimensions. *The Annals of Probability*, 45(4):2309–2352.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2019). Inference on causal and structural parameters using many moment inequalities. *The Review of Economic Studies*, 86(5):1867–1900.

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

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

Chesher, A. and Rosen, A. M. (2017). Generalized instrumental variable models. *Econometrica*, 85(3):959–989.

Chesher, A., Rosen, A. M., and Smolinski, K. (2013). An instrumental variable model of multiple discrete choice. *Quantitative Economics*, 4(2):157–196.

Chetverikov, D. (2018). Adaptive tests of conditional moment inequalities. *Econometric Theory*, 34(1):186–227.

Chetverikov, D., Wilhelm, D., and Kim, D. (2021). An adaptive test of stochastic monotonicity. *Econometric Theory*, 37(3):495–536.

Chiburis, R. C. (2008). Approximately most powerful tests for moment inequalities. Working paper, Princeton University, Department of Economics, Princeton, New Jersey. Unpublished manuscript.

Ciliberto, F. and Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, 77(6):1791–1828.

Delgado, M. A. and Escanciano, J. C. (2013). Conditional stochastic dominance testing. *Journal of Business & Economic Statistics*, 31(1):16–28.

Donald, S. G. and Hsu, Y. (2016). Improving the power of tests of stochastic dominance. *Econometric Reviews*, 35(4):553–585.

Elliott, G., Müller, U. K., and Watson, M. W. (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica*, 83(2):771–811.

Fack, G., Grenet, J., and He, Y. (2019). Beyond truth-telling: Preference estimation with centralized school choice and college admissions. *American Economic Review*, 109(4):1486–1529.

Fang, X. and Koike, Y. (2024). Large-dimensional central limit theorem with fourth moment error bounds on convex sets and balls. *The Annals of Applied Probability*, 34:2065–2106.

Galichon, A. and Henry, M. (2011). Set identification in models with multiple equilibria. *Review of Economic Studies*, 78(4):1264–1298.

Gandhi, A., Lu, Z., and Shi, X. (2023). Estimating demand for differentiated products with zeroes in market share data. *Quantitative Economics*, 14(2):381–418.

Ghosal, S., Sen, A., and van der Vaart, A. W. (2000). Testing monotonicity of regression. *Annals of Statistics*, 28(4):1054–1082.

Giacomini, R. and Kitagawa, T. (2021). Robust bayesian inference for set-identified models. *Econometrica*, 89(4):1519–1556.

Giné, E., Mason, D. M., and Zaitsev, A. Y. (2003). The l_1-norm density estimator process. *Annals of Probability*, 31(2):719–768.

He, Y. (2017). Gaming the boston school choice mechanism in beijing. Technical Report 15-551, Toulouse School of Economics.

Ho, K. and Rosen, A. M. (2017). Partial identification in applied research: Benefits and challenges. In Honoré, B., Pakes, A., Piazzesi, M., and Samuelson, L., editors, *Advances in Economics and Econometrics: Eleventh World Congress*, pages 307–359. Econometric Society Monographs, Cambridge University Press.

Hsu, Y., Liu, C., and Shi, X. (2019). Testing generalized regression monotonicity. *Econometric Theory*, 35(6):1146–1200.

Iaryczower, M., Shi, X., and Shum, M. (2018). Can words get in the way? the effect of deliberation in collective decision making. *Journal of Political Economy*, 126(2):688–734.

Kalouptsidi, M., Kitamura, Y., Lima, L., and Souza-Rodrigues, E. A. (2020). Partial identification and inference for dynamic models and counterfactuals. NBER Working Paper 26761, National Bureau of Economic Research.

Kline, B., Pakes, A., and Tamer, E. (2021). Moment inequalities and partial identification in industrial organization. In Ho, K., Hortacsu, A., and Lizzeri, A., editors, *Handbook of Industrial Organization*, volume 4, pages 345–431. Elsevier.

Lee, S., Linton, O. B., and Whang, Y. (2009). Testing for stochastic monotonicity. *Econometrica*, 77(2):585–602.

Lee, S., Song, K., and Whang, Y. (2013). Testing functional inequalities. *Journal of Econometrics*, 172(1):14–32.

Lee, S., Song, K., and Whang, Y.-J. (2018). Testing for a general class of functional inequalities. *Econometric Theory*, 34(5):1018–1064.

Li, J. and Liao, Z. (2020). "uniform nonparametric inference for time series. *Journal of Econometrics*, 129:28–51.

Li, J., Liao, Z., and Quaedvlieg, R. (2022). Conditional superior predictive ability. *Review of Economic Studies*, 89(2):843–875.

Lok, T. M. and Tabri, R. V. (2021). An improved bootstrap test for restricted stochastic dominance. *Journal of Econometrics*, 224(2):371–393.

Manski, C. F. (1989). Anatomy of the selection problem. *Journal of Human Resources*, 24(3):343–360.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *American Economic Review, Papers and Proceedings*, 80(2):319–323.

Manski, C. F. (1993). Identification problems in the social sciences. *Sociological Methodology*, 23:1–56.

Molinari, F. (2020). Microeconometrics with partial identification. In Durlauf, S. N., Hansen, L. P., Heckman, J. J., and Matzkin, R. L., editors, *Handbook of Econometrics*, volume 7A, pages 355–486. Elsevier.

Morales, E., Sheu, G., and Zahler, A. (2019). Extended gravity. *Review of Economic Studies*, 86(6):2668–2712.

Pakes, A., Porter, J. R., Ho, K., and Ishii, J. (2015). Moment inequalities and their application. *Econometrica*, 83(1):315–334.

Pollard, D. (2002). *A User's Guide to Measure Theoretic Probability*, volume 8 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press.

Romano, J. P., Shaikh, A. M., and Wolf, M. (2014). A practical two-step method for testing moment inequalities. *Econometrica*, 82(5):1979–2002.

Seo, J. (2018). Tests of stochastic monotonicity with improved size and power. *Journal of Econometrics*, 207(1):53–70.

Sheng, S. (2020). A structural econometric analysis of network formation games through subnetworks. *Econometrica*, 88(5):1829–1858.

Shi, X., Shum, M., and Song, W. (2018). Estimating semi-parametric panel multinomial choice models using cyclic monotonicity. *Econometrica*, 86(2):737–761.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, New York.