# Inference on Projections of Identified Sets *

Hiroaki Kaido[†] Francesca Molinari[‡] Jörg Stoye[§]

May 14, 2015

**Abstract**

This paper proposes a bootstrap-based procedure to build confidence intervals for single components of a partially identified parameter vector, and for linear combinations of such components. Our confidence interval is constructed from the projection of a confidence region for the entire parameter. We propose a novel way to calculate the *critical level*, the amount by which we relax the moment restrictions so that the component of interest, instead of the entire vector, is covered by the confidence interval with a prespecified probability. This new methodology allows us to show that in finite sample, our confidence interval is (weakly) shorter than the projection of confidence regions designed to cover the entire parameter. We provide simple conditions under which our confidence interval is asymptotically strictly shorter, and conditions under which our confidence interval has uniformly asymptotically exact coverage. We further show that our inference method controls asymptotic coverage uniformly over a large class of data distributions. This class of distributions is non-nested with the class of distributions over which the main alternative to our method, which is based on a profiled test statistic, is uniformly valid. Our bootstrap procedure iterates over linear programming problems, and as such is computationally attractive.

**Work in progress. Please do not cite or circulate.**

*We are grateful to Ivan Canay and seminar and conference participants at Brown, Chicago, Columbia, Cornell, Michigan, Michigan State, NYU, Penn State, UPenn, Syracuse, Toronto, Vanderbilt, Yale, Wisconsin, CEME, and Econometric Society's Winter Meeting 2015 for comments.

[†]Department of Economics, Boston University, hkaido@bu.edu. Financial support from NSF grant SES-1230071 is gratefully acknowledged.

[‡]Department of Economics, Cornell University, fm72@cornell.edu. Financial support from NSF grant SES-0922330 is gratefully acknowledged.

[§]Department of Economics, Cornell University, stoye@cornell.edu. Financial support from NSF grant SES-1260980 is gratefully acknowledged.

# 1 Introduction

A growing body of literature in econometric theory focuses on estimation and inference in partially identified models. For a given $d$-dimensional parameter vector $\theta$ characterizing the model, much work has been devoted to develop testing procedures and associated confidence sets in $\mathbb{R}^d$ that satisfy various desirable properties. These include coverage of each element of the $d$-dimensional identification region, denoted $\Theta_I$, or coverage of the entire set $\Theta_I$, with a prespecified –possibly uniform– asymptotic probability. From the perspective of researchers familiar with inference in point identified models, this effort is akin to building confidence ellipsoids for the entire parameter vector $\theta$. However, applied researchers are frequently interested in conducting inference for each component of a partially identified vector, or for linear combinations of components of the partially identified vector, similarly to what is typically done in multiple linear regression.

The goal of this paper is to provide researchers with a novel procedure to conduct such inference in partially identified models, while ensuring that asymptotic coverage is *uniformly correct* in a sense made precise below. The procedure is computationally attractive because critical values are computed by bootstrapping a *linear* programming problem.

Given the abundance of inference procedures for the entire parameter vector $\theta$, one might be tempted to just report the projection of one of them as confidence interval for the projections of $\Theta_I$ (e.g., for the bounds on each component of $\theta$). Such a confidence interval is asymptotically valid but typically conservative. The extent of the conservatism increases with the dimension of $\theta$ and is easily appreciated in the case of a point identified parameter. Consider, for example, a linear regression in $\mathbb{R}^{10}$, and suppose for simplicity that the limiting covariance matrix of the estimator is the identity matrix. Then a 95% confidence interval for each component of $\theta$ is obtained by adding and subtracting 1.96 to that component's estimate. In contrast, projection of a 95% Wald confidence ellipsoid on each component amounts to adding and subtracting 4.28 to that component's estimate. We refer to this problem as *projection bias*.

The key observation behind our approach is that projection bias can be anticipated. In the point identified case, this is straightforward. Returning to the example of multiple linear regression, if we are interested in a confidence interval with a certain asymptotic coverage for a component of the vector $\theta$, we can determine the level of a confidence ellipsoid whose projection yields just that confidence interval. When the limiting covariance matrix of the estimator is the identity matrix and $d = 2$, projection of a confidence ellipsoid with asymptotic coverage equal to 85.4% yields an interval equal to the component's estimate plus/minus 1.96, and therefore asymptotic coverage of 95% for that component; when $d = 5$, the required ellipsoid's coverage is 42.8%; when $d = 10$, the required ellipsoid's coverage is 4.6%.[1]

[1]The fast decrease in the required coverage level can be explained observing that the volume of a ball of radius $r$ in $\mathbb{R}^d$ decreases at least geometrically in $d$.

The main technical contribution of this paper is to show how this straightforward insight from the point identified case can be generalized in the partially identified case while preserving computational feasibility and desirable coverage properties.

We focus on the class of moment (in)equalities models, a special case of partial identification that has received much attention in the recent literature and in which $\Theta_I$ is equal to the set of values for $\theta$ that satisfy a finite number of moment equalities and inequalities. Combination of the moment (in)equalities framework with the support function approach (see Beresteanu and Molinari (2008)) was first proposed by Kaido (2012). For the case of moment functions that are convex in $\theta$, Kaido (2012) obtained testing procedures and confidence sets that are asymptotically valid pointwise. We generalize these results to the case that the moments are not convex in $\theta$, and significantly advance them to provide a testing procedure and associated confidence intervals that are uniformly valid over interesting classes of data generating processes.

The existing literature provides one main alternative to our method, designed to provide uniformly valid test procedures and confidence statements for projections of $\theta \in \Theta_I$.[2] This procedure is based on a profiled test statistic as introduced in Romano and Shaikh (2008) and significantly advanced in Bugni, Canay, and Shi (2014). As we explain below, the class of data generating processes over which our procedure is uniformly valid, is non-nested with the class of data generating processes over which the profiling method is uniformly valid. Another method proposed by Pakes, Porter, Ho, and Ishii (2011) for inference on projections, is based on bootstrapping directly the support function of a sample analog of the identified set. As we explain below, this method controls asymptotic coverage over a significantly smaller class of models than our method.

The importance of uniform coverage of confidence sets in partial identification was first emphasized by Imbens and Manski (2004), further clarified in Stoye (2009), and fully developed for moment (in)equalities models by Romano and Shaikh (2008), Andrews and Guggenberger (2009) and Romano and Shaikh (2010).[3] These authors show that in this context, several traps may emerge unless inference procedures are uniform. For example, if identified sets are intervals that are long relative to standard errors, then the testing problem is essentially one-sided, leading to shorter and more easily computed confidence intervals. Of course, in a pointwise perspective, every interval with positive length is asymptotically long relative to standard errors. So one might naively assume a "long" interval whenever the estimated

[2]We remark that this method provides uniformly valid confidence intervals also for non-linear functions of $\theta$, something that our method does not currently do.

[3]This might be puzzling because uniformity is less frequently emphasized in other contexts and because it is well known that uniformity over all data generating processes is elusive (Savage). Indeed, uniformity holds only over restricted classes of models that exclude Savage-type examples. Of course, other areas where uniformity is heavily studied exist and include contexts where one encounters similar traps to those described here. Examples include inference close to unit roots Mikusheva (2007), weak identification Andrews and Cheng (2012), and post-model selection inference (see Leeb and Pötscher (2005) for a negative take). See also the discussion, with more examples, in Andrews and Guggenberger (2009).

length of the interval is positive. Inference based on this approximation would, however, break down along Pitman drift parameter sequences where degenerate intervals, i.e. point identification, are reached in the limit. This example was explored by Imbens-Manski and Stoye.

In our problem, uniformity is desirable along a novel dimension: Holding one (reasonably well-behaved) model fixed, confidence regions should be equally valid for different directions of projection. It is surprisingly easy to fail this criterion. For example, if one does not properly account for flat faces which are orthogonal to the direction of projection, the resulting confidence interval will not be valid uniformly over directions of projection if the true identified set is a polyhedron. A polyhedron is not only a simple shape but also practically relevant: it arises for Best Linear Prediction (BLP) using interval data with discrete regressors. In this example, a method that does not apply at (or near) flat faces is not equally applicable to all linear hypotheses that one might want to test. This stands in stark contrast to point identified BLP estimation: Barring collinearity, an F-test will be applicable uniformly over simple linear hypotheses. Under this latter condition and some others, our method too applies uniformly over linear hypotheses.

**Overview of the Method.** Our proposal is to report as confidence interval for a chosen linear projection of each $\theta \in \Theta_I$, the support function in direction $\pm p$ of a $c(\theta)$-level set of a sample criterion function that aggregates sample violation of moment inequalities, where $c(\theta)$ is chosen to achieve the desired coverage. In the salient special case of moment inequalities, reporting the $c(\theta)$-level set of the sample criterion function boils down to relaxing all studentized moment inequalities by a properly determined amount $c(\theta)$, and reporting the support function in direction $\pm p$ of the set defined by the relaxed studentized inequalities.

The correct choice of $c(\theta)$ entails anticipating projection bias. In particular, for each candidate $\theta \in \Theta$, we calibrate $c(\theta)$ to insure that across bootstrap repetitions the projection of each $\theta \in \Theta_I$ is covered with a pre-specified probability of at least $1 - \alpha$. To assure that our methodology is computationally attractive, we work with a local linear approximation to the moment inequalities, which yields that $c(\theta)$ can be calibrated by iterating over a linear programming problem. We then establish uniform asymptotic validity of our procedure over the class of distributions that we allow for.

This class of data generating processes can be related to the existing literature as follows. We start from the same assumptions as Andrews and Soares (2010, AS10 henceforth), and in fact compare the length of our confidence interval to the length of the projection of their confidence set (which is constructed with the goal of covering each vector $\theta \in \Theta_I$ with a prespecified asymptotic probability uniformly). Similarly to the related literature, we ensure uniform validity in presence of drifting-to-binding inequalities by adopting Generalized Moment Selection as put forward by AS10, Bugni (2009), and Canay (2010). In addition, our procedure requires that the correlation matrix of the sample moment (in)equalities has eigenvalues uniformly bounded from below. This assumption was considered in AS10 (for a specific criterion function), but eliminated by Andrews and Barwick (2012).

We then further restrict the class of data generating processes that we work with, by assuming that for each individual constraint, a local linear approximation (i) is a good approximation to the constraint's graph and (ii) can be estimated. Around parameter values for which a moment inequality binds, the gradient of that ineqality must, therefore, be continuous and have strictly positive norm.[4] This assumption (and the lower bound on eigenvalues of the correlation matrix) is not required by the profiling method, and as such data generating processes for which it fails might be handled by Romano and Shaikh (2008) and Bugni, Canay, and Shi (2014), but cannot be handled by our procedure.

However, we do not further restrict the local geometry of $\Theta_I$. In particular, we allow for an extreme point of $\Theta_I$ in direction of projection to be (i) a point of differentiability of the boundary of $\Theta_I$, (ii) a point on a flat face that is orthogonal to the direction of projection, or (iii) a point on a flat face that is drifting-to-orthogonal to the direction of projection. Case (iii) is excluded by Romano and Shaikh (2008) and Bugni, Canay, and Shi (2014), and all three cases are excluded by Pakes, Porter, Ho, and Ishii (2011). We also allow for corners with extremely acute angles, meaning that the interior of $\Theta_I$ locally vanishes and that the joint linear approximation of constraints is not a good approximation to the local geometry of $\Theta_I$. This case is again excluded by Pakes, Porter, Ho, and Ishii (2011) and also by Chernozhukov, Hong, and Tamer (2007).[5] Compared to the related literature, our ability to handle this class of models comes at the price of an additional (non-drifting) tuning parameter. We explain in Section 3.4 why this additional parameter is needed, and why it is helpful. We also provide a (heuristic) method to choose it.

Going back to AS10, our method can be directly compared to projection of their confidence region if one uses comparable tuning parameters. By construction, the confidence intervals that we propose are (weakly) shorter in any finite sample. They asymptotically agree if and only if the asymptotic testing problem is equivalent to intersection bounds in $\mathbb{R}^1$, that is, if all binding constraints are locally orthogonal to the direction of projection.[6] This is furthermore the only case in which (in absence of drifting-to-binding inequalities) projection of AS10's confidence region is not, in fact, asymptotically conservative.

There are a few other papers in the literature, that aim at providing confidence intervals for projections of identified sets. These include Andrews, Berry, and Jia (2004), Chen, Tamer, and Torgovitsky (2011), Kitagawa (2012), Kline and Tamer (2015) and Wan (2013). However, each of these contributions provide confidence intervals that are valid pointwise (and some are Bayesian and not frequentist, as our approach), and therefore might not be valid uniformly

[4]As we need these conditions in a uniform sense, we actually impose Lipschitz continuity of, as well as a strictly positive lower bound on the norm of, gradients.

[5]Our ability to handle this case constitutes a major advance over a previous, widely presented version of this paper.

[6]The case is more generic than this description may sound because it obtains whenever the relevant support point of $\Theta_I$ is a point of differentiability of $\partial\Theta_I$.

over the class of models that we consider.

**Structure of the paper.** Section 2 sets up notation and describes the inferential problem that we focus on, providing the basic insight in our approach. Section 3 describes the bootstrap procedure based on linear programming that we propose for computing the level $\hat{c}_n(\theta)$. It then lays out our assumptions and presents our main results: asymptotic validity and improvement over methods based on projection of high dimensional confidence ellipsoids. The section is concluded with a discussion of the challenges posed by the local geometry of the identification region for uniform inference. Within this discussion, we further elucidate the relation between our method, and the existing literature.

**Sections reporting computational aspects of the method, Monte Carlo exercises, and concluding remarks, are TBA.**

# 2 Set up

We start by introducing basic notation for the moment (in)equalities framework. Let $X_i \in \mathcal{X} \subseteq \mathbb{R}^{d_X}$ be a random vector with distribution $P$ and let $\Theta \subseteq \mathbb{R}^d$ denote a parameter space. We then let $m_j : \mathcal{X} \times \Theta \to \mathbb{R}$ denote a measurable function characterizing the model, known up to parameter vector $\theta \in \Theta$, with $j = 1, \dots, J_1 + J_2$. The true parameter value $\theta$ is assumed to satisfy the moment inequality and equality restrictions:

$$\begin{aligned} &E_P[m_j(X_i, \theta)] \leq 0, \; j = 1, \cdots, J_1, \\ &E_P[m_j(X_i, \theta)] = 0, \; j = J_1 + 1, \cdots, J_1 + J_2. \end{aligned} \tag{2.1}$$

The *identification region* $\Theta_I(P)$ is the set of parameter values in $\Theta$ that satisfy these moment restrictions. In what follows, we simply write $\Theta_I$ whenever its dependence on $P$ is obvious. For a random sample $\{X_i, i = 1, \cdots, n\}$ of observations drawn from $P$, we let $\bar{m}_{j,n}(\theta) \equiv n^{-1} \sum_{i=1}^n m_j(X_i, \theta), j = 1, \cdots, J_1 + J_2$ denote the sample moments.

A key tool for our inference procedure is the support function of a parameter set. We denote the unit sphere in $\mathbb{R}^d$ by $\mathbb{S}^{d-1} \equiv \{p \in \mathbb{R}^d : \|p\| = 1\}$, an inner product between two vectors $x, y \in \mathbb{R}^d$ by $x'y$, and use the following standard definition of support function and support set:

DEFINITION 2.1: *Given a closed set $A \subset \mathbb{R}^d$, its* support function *is*

$$s(p, A) = \sup\{p'a, \, a \in A\}, \; p \in \mathbb{S}^{d-1},$$

*and its* support set *is*

$$H(p, A) = \{a \in \mathbb{R}^d : p'a = s(p, A)\} \cap A, \; p \in \mathbb{S}^{d-1}.$$

It is useful to think of $p'a$ as a projection of $a \in \mathbb{R}^d$ to a one-dimensional subspace spanned by the direction $p$. For example, when $p$ is a vector whose $j$-th coordinate is 1 and other coordinates are 0s, $p'a = a_j$ is the projection of $a$ to the $j$-th coordinate. The support function of a set $A$ gives the supremum of the projections of points belonging to this set. Since $\inf\{p'a, a \in A\} = -s(-p, A)$, the projection of any $a \in A$ lies in the interval $[-s(-p, A), s(p, A)]$. Throughout, we call this interval the *projection* of $A$.[7]

Our goal is to cover each element in the projection of $\Theta_I$ in direction $p \in \mathbb{S}^{d-1}$. Toward this end, consider a confidence interval $CI_n$ obtained by projecting a confidence region $\mathcal{C}_n$ for the entire parameter vector $\theta$. This is akin to projecting the Wald ellipsoid in the point identified setting. In the moment (in)equalities setting, confidence regions designed to cover $\theta$ satisfying (2.1) have the following common form:

$$\mathcal{C}_n(c_n) \equiv \{\theta \in \Theta : T_n(\theta) \leq c_n(\theta)\}, \tag{2.2}$$

where $T_n(\theta)$ is a test statistic, and $c_n(\theta)$ is a critical value (see e.g. Chernozhukov, Hong, and Tamer, 2007; Andrews and Soares, 2010). Critical values $c_n(\theta)$ available in the literature are calibrated so that the entire parameter $\theta$ is covered by the confidence region with at least a prespecified probability $1-\alpha$ asymptotically. This means that $c_n(\theta)$ is large enough to ensure that any linear projection of $\theta$ is covered with probability $1-\alpha$ by the projection of $\mathcal{C}_n(c_n)$, i.e. $p'\theta \in [-s(-p, \mathcal{C}_n(c_n)), s(p, \mathcal{C}_n(c_n))]$ for *all* $p \in \mathbb{S}^{d-1}$. Clearly, this is more than needed as we aim at covering the projection of $\theta$ for *one* direction. Projecting $\mathcal{C}_n(c_n)$, therefore, tends to produce a conservative confidence interval.

As discussed in the introduction, this projection bias, however, can be removed by explicitly calculating a critical value that is just enough to ensure the coverage of the projection of interest. As opposed to the simple adjustment of the confidence level for the Wald confidence ellipsoid, the calculation of such a critical value in the moment (in)equalities setting is nontrivial, and it requires a careful analysis of the local behavior of the moment restrictions at each point in the identification region. This is because the projection of the confidence region depends on (i) the behavior of the sample moments entering the inequality restrictions, which can change discontinuously depending on whether they bind at $\theta$ or not and (ii) the local geometry of the identification region at $\theta$. Here, by local geometry, we mean the shape of the constraint set formed by the sample moment restrictions and its relation to the level set of the objective function $p'\theta$. These features can be quite different at different points in the identification region, which in turn makes uniform inference for the projection

[7]The projection of any $a \in A$ lies in the projection of $A$, but the converse is not always true. If $A$ is not connected, there may exist $x \in [-s(-p, A), s(p, A)]$ for which no $a$ can be found such that $x = p'a$. Most applications of partial identification to date are for models that yield connected identification regions. Some exceptions include Molinari (2008) and Chesher, Rosen, and Smolinski (2012). However, in some of the models considered in these papers, the researcher knows ex ante that $\Theta_I$ is disconnected, and also what subset of $\Theta$ contains each connected subset of $\Theta_I$. Hence, our procedure can be repeated for each connected component separately to obtain confidence intervals with smaller probability of false coverage.

challenging. In particular, the second issue does not arise if one only considers inference for the entire parameter vector, and hence this new challenge requires a new methodology. The core innovation of this paper is to provide a novel and computationally attractive procedure to construct a critical value that overcomes these challenges.

With a particular choice of a test statistic, the calculation of a critical value can be recast as a problem of finding how much one needs to relax sample counterparts of the moment restrictions so that the projection of $\mathcal{C}_n(c_n)$ covers $p'\theta$ with a prespecified asymptotic probability, uniformly in $P$. Let $\mathcal{C}_n(c_n)$ be a confidence region as in (2.2) with the test statistic

$$T_n(\theta) = \max\{\max_{j=1,\cdots,J_1} \sqrt{n}[\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)]_+, \max_{j=J_1+1,\cdots,J_2} \sqrt{n}|\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)|\}, \tag{2.3}$$

where $\hat{\sigma}_{n,j}(\theta)$ is a suitable estimator of the asymptotic standard deviation, $\sigma_{P,j}(\theta)$, of $\sqrt{n}\bar{m}_{n,j}(\theta)$. The support function of $\mathcal{C}_n(c_n)$ is then the optimal value of the following nonlinear program (NLP):

$$\begin{aligned} s(p, \mathcal{C}_n(c_n)) = &\sup_{\theta\in\Theta} \; p'\theta \\ &s.t. \quad \sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta) \leq c_n(\theta), \; j = 1, \cdots, J_1 + 2J_2, \end{aligned} \tag{2.4}$$

where we define the last $J_2$ moments as $\bar{m}_{n,J_1+J_2+k}(\theta) = n^{-1}\sum_{i=1}^n -m_{J_1+k}(X_i, \theta)$ for $k = 1, \cdots, J_2$. In other words, we split moment equality constraints into two opposing inequality constraints relaxed by $c_n(\theta)$ and impose them in addition to the first $J_1$ inequalities relaxed by the same amount. In total, we therefore have $J \equiv J_1 + 2J_2$ inequality constraints.

Remark 2.1: While our analysis is carried out working with the criterion function in equation 2.3, it is easy to show that our method (including the bootstrap procedure described in Section 3.1) applies similarly to a criterion function of the form

$$\tilde{T}_n(\theta) = \sum_{j=1,\cdots,J_1} \sqrt{n}[\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)]_+ + \sum_{j=J_1+1,\cdots,J_2} \sqrt{n}|\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)|, \tag{2.5}$$

Criterion function $T_n$ corresponds to criterion function $S_3$ in AS10; criterion function $\tilde{T}_n$ is akin to criterion function $S_1$ in AS10. In addition, AS10 propose a QLR based test statistic previously considered in Rosen (2008). This test statistic does not land itself easily to linearization, and as such we do not consider it in this paper.

Define the asymptotic size of the confidence interval by

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Theta_I(P)} P(p'\theta \in CI_n), \tag{2.6}$$

where $\mathcal{P}$ is a class of distributions that we specify below. Let the two-sided confidence interval be defined by

$$CI_n \equiv [-s(-p, \mathcal{C}_n(c_n)), s(p, \mathcal{C}_n(c_n))]. \tag{2.7}$$

Consider a sequence of parameter and distribution pairs $(\theta_n, P_n) \in \{(\theta, P) : \theta \in \Theta_I(P), P \in \mathcal{P}\}$. Then, the projection of $\theta_n$ is covered when

$$\begin{aligned}
&-s(-p, \mathcal{C}_n(c_n)) \leq p'\theta_n \leq s(p, \mathcal{C}_n(c_n)) \\
&\Leftrightarrow \begin{Bmatrix} -\sup -p'\vartheta \\ s.t.\, \vartheta \in \Theta, \quad \frac{\sqrt{n}\bar{m}_{n,j}(\vartheta)}{\hat{\sigma}_{n,j}(\vartheta)} \leq c_n(\vartheta), \forall j \end{Bmatrix} \leq p'\theta_n \leq \begin{Bmatrix} \sup p'\vartheta \\ s.t.\, \vartheta \in \Theta, \quad \frac{\sqrt{n}\bar{m}_{n,j}(\vartheta)}{\hat{\sigma}_{n,j}(\vartheta)} \leq c_n(\vartheta), \forall j \end{Bmatrix} \\
&\Leftrightarrow \begin{Bmatrix} -\sup_\lambda -p'\lambda \\ s.t. \lambda \in \sqrt{n}(\Theta - \theta_n), \quad \frac{\sqrt{n}\bar{m}_{n,j}(\theta_n + \lambda/\sqrt{n})}{\hat{\sigma}_{n,j}(\theta_n + \lambda/\sqrt{n})} \leq c_n(\theta_n + \lambda/\sqrt{n}), \forall j \end{Bmatrix} \leq 0 \\
&\qquad \leq \begin{Bmatrix} \sup_\lambda p'\lambda \\ s.t. \lambda \in \sqrt{n}(\Theta - \theta_n), \quad \frac{\sqrt{n}\bar{m}_{n,j}(\theta_n + \lambda/\sqrt{n})}{\hat{\sigma}_{n,j}(\theta_n + \lambda/\sqrt{n})} \leq c_n(\theta_n + \lambda/\sqrt{n}), \forall j \end{Bmatrix}
\end{aligned} \tag{2.8}$$

where the second equivalence follows from rewriting the problem which maximizes $p'\vartheta$ with respect to $\vartheta$ localized as $\vartheta = \theta_n + \lambda/\sqrt{n}$ by another problem which maximizes the same objective function with respect to the localization parameter $\lambda$. The argument above suggests that one can control the confidence size if one finds the right amount of $c_n$ such that 0 lies within the optimal values of the NLPs in (2.8) with probability $1 - \alpha$.

To reduce the computational cost associated with calibrating the value of $c_n$, we approximate the probability of the event in equation (2.8), by taking a linear expansion in $\lambda$ of the constraint set. In particular, for the $j$-th constraint, adding and subtracting $E_P[m_j(X_i, \theta + \lambda/\sqrt{n})]$ yields

$$\begin{aligned}
&\frac{\sqrt{n}\bar{m}_j(X_i, \theta_n + \lambda/\sqrt{n})}{\hat{\sigma}_j(\theta_n + \lambda/\sqrt{n})} \\
&= \sqrt{n}\frac{(\bar{m}_j(X_i, \theta_n + \lambda/\sqrt{n}) - E_P[m_j(X_i, \theta_n + \lambda/\sqrt{n})])}{\hat{\sigma}_j(\theta_n + \lambda/\sqrt{n})} + \sqrt{n}\frac{E_P[m_j(X_i, \theta_n + \lambda/\sqrt{n})]}{\hat{\sigma}_j(\theta_n + \lambda/\sqrt{n})} \\
&= \{\mathbb{G}_{P,j,n}(\theta_n + \lambda/\sqrt{n}) + D_{P,j}(\bar{\theta}_n)'\lambda + h_{P,j,n}(\theta_n)\}(1 + \eta_{j,n}(\theta_n)),
\end{aligned} \tag{2.9}$$

where $\mathbb{G}_{P,j,n}(\cdot) = \sqrt{n}(\bar{m}_{n,j}(\cdot) - E_P[m_j(X_i, \cdot)])/\sigma_{P,j}(\cdot)$ is a normalized empirical process indexed by $\theta \in \Theta$, $D_{P,j}(\cdot) \equiv \nabla_\theta\{E_P[m_j(X_i, \cdot)]/\sigma_{P,j}(\cdot)\}$ is the gradient of the normalized moment, and $h_{P,j,n}(\cdot) \equiv \sqrt{n}E_P[m_j(X_i, \cdot)]/\sigma_{P,j}(\cdot)$ is the population moment scaled by $\sqrt{n}$. The second equality follows from the mean value theorem, where $\bar{\theta}_n$ represents a mean value between $\theta_n$ and $\theta_n + \lambda/\sqrt{n}$, which can differ across components of the gradient, and $\eta_{j,n}(\cdot) \equiv \sigma_{P,j}(\cdot)/\hat{\sigma}_{j,n}(\cdot) - 1$ can be shown to converge in probability to 0 uniformly.

Under suitable regularity conditions set forth in Section 3.2 (which include differentiability of $E_P[m_j(X_i, \theta)]$ in $\theta$ for each $j$), we show that the probability that the nonlinear program in equation (2.8) takes a value greater or equal to zero, is suitably approximated by the probability that a program linear in $\lambda$ takes a value greater or equal to zero. The constraint set of this linear program is given by the sum of (i) an empirical process $\mathbb{G}_{P,j}(\theta)$ evaluated at $\theta$ (that we can approximate by a bootstrap) (ii) a rescaled gradient times $\lambda$, $D_{P,j}(\theta)'\lambda$ (that we can uniformly consistently estimate on compact sets), and (iii) the parameter $h_{P,j,n}(\theta)$ that measures the extent to which each moment inequality is binding. This suggests a computationally attractive bootstrap procedure based on linear programs.

As commonly done in nonlinear econometric models, we use linearizations to obtain a first-order approximation to the statistic of interest. In our setting, the object of interest is the support function of the confidence region. Calculating the support function subject to the moment (in)equality constraints is similar to calculating a nonlinear estimator (e.g. GMM estimator) in the sense that both seek for a particular parameter value, which "solves" a system of sample moment restrictions. For our problem, we seek for a parameter value satisfying suitably relaxed moment inequalities and equalities whose projection is maximal, while GMM, for example, seeks for a parameter value that minimizes the norm of sample moments, or necessarily a value that solves its first-order conditions. Hence, the solution concepts are different. However, the methodology for obtaining approximations is common. Recall that one may obtain an influence function of the GMM estimator by linearizing the moment restrictions in the first-order conditions around the true parameter value and by solving for the estimator. In a complete analogy to this example, calculating the optimal value of the linear program discussed above can be interpreted as applying a particular solution concept (the maximum value of the linear projections) to a system of moment (in)equality constraints linearized around the true value.

# 3 Asymptotic Validity of Inference

## 3.1 An LP-based bootstrap critical value

In light of the discussion in the previous section, for a given $\theta \in \Theta$ we calibrate the level $c_n(\theta)$ using a bootstrap procedure that iterates over linear programs (LP). Our bootstrap critical value is defined as

$$\hat{c}_n(\theta) \equiv \inf\{c \in \mathbb{R}_+ : P(-\mathcal{Z}_n^b(-p, c, \theta) \leq 0 \leq \mathcal{Z}_n^b(p, c, \theta)) \geq 1 - \alpha\}, \tag{3.1}$$

for a process $\mathcal{Z}_n^b$, which can be calculated as follows. Below, let $m_j^*(\cdot, \theta) \equiv m_j(\cdot, \theta) - \bar{m}_{j,n}(\theta)$ be the re-centered moment function and let $\hat{D}_{n,j}(\theta)$ be an estimator of $D_{P,j}(\theta)$.

**Step 0** For given $c$, repeat the following bootstrap simulation for $b = 1, ..., B$ times:

**Step 1** Draw a new sample $(X_i^b)_{i=1}^n$ by independently resampling the data with replacement.

**Step 2** Compute $\mathcal{Z}_n^b(q,c,\theta)$ for $q=p$ and $-p$ by solving the following program:

$$\mathcal{Z}_n^b(q,c,\theta) \equiv \sup_{\lambda} \; q'\lambda$$
$$s.t.\ \lambda \in \rho B_d, \quad \frac{1}{\sqrt{n}}\sum_{i=1}^n m_j^*(X_i^b,\theta)/\hat{\sigma}_{n,j}(\theta) + \hat{D}_{n,j}(\theta)'\lambda + \hat{\zeta}_{n,j}(\theta) \le c,\ j=1,\cdots,J, \tag{3.2}$$

where $\rho > 0$ is a constant chosen by the researcher, $B_d = \{x \in \mathbb{R}^d : |x_j| \le 1, \forall j\}$ is a unit box in $\mathbb{R}^d$, and $\hat{\zeta}_{n,j}(\theta)$ is one of the generalized moment selection (GMS) functions proposed by Andrews and Soares (2010) and defined by

$$\hat{\zeta}_{n,j}(\theta) = \begin{cases} 0 & \text{if } \kappa_n^{-1}\sqrt{n}\bar{m}_{j,n}(\theta)/\hat{\sigma}_{j,n}(\theta) \ge -1 \\ -\infty & \text{if } \kappa_n^{-1}\sqrt{n}\bar{m}_{j,n}(\theta)/\hat{\sigma}_{j,n}(\theta) < -1. \end{cases} \tag{3.3}$$

**Step 3** Compute

$$\hat{c}_n(\theta) \equiv \inf\{c : P^*(-\mathcal{Z}_n^b(-p,c,\theta) \le 0 \le \mathcal{Z}_n^b(p,c,\theta)) \ge 1-\alpha\}. \tag{3.4}$$

The linear program in (3.2) can be solved efficiently using commonly used softwares.[8]

REMARK 3.1: For concreteness in Step 2 above we propose the use of a specific GMS function among the ones in AS10. As shown in the Appendix, our results apply to all GMS functions in AS10, except one.[9]

The idea behind this bootstrap procedure is as follows. First, the bootstrapped empirical process and the estimator of the gradient give approximations to the first two terms in the constraint in (2.8). Second, the GMS function approximates the local slackness parameter $h_{P,j,n}(\theta)$ conservatively. Suppose, for example, the $j$-th constraint is locally binding at $\theta$, i.e. $h_{P,j,n}(\theta)$ is negative but is close to 0. If the GMS procedure selects this constraint, it sets the third term in (3.2) to 0. This makes our critical value slightly more conservative because replacing $h_{P,j,n}(\theta)$ with 0 requires us to relax the constraints by a larger amount to make the maximized value of the linear program greater or equal to 0 with probability $1-\alpha$. Since the local slackness parameter cannot be consistently preestimated uniformly, we employ the GMS procedure proposed by Andrews and Soares (2010) to obtain the conservative distortion.

In addition to the linear approximation of the constraints in (2.8), our bootstrap procedure restricts $\lambda$ to the "$\rho$-box" $\rho B_d$ for some $\rho > 0$. This restriction is introduced to allow our

[8] Examples of high-speed solvers for linear programs include CVXGEN (available from http://cvxgen.com) and Gurobi (available from http://www.gurobi.com).

[9] These are $\varphi^1 - \varphi^4$ in AS10, all of which depend on $\kappa_n^{-1}\sqrt{n}\bar{m}_{j,n}(\theta)/\hat{\sigma}_{j,n}(\theta)$. We do not consider GMS function $\varphi^5$ in AS10, which depends also on the covariance matrix of the moment functions.

methodology to remain uniformly valid even in situations where the set of moment restrictions tend to a configuration that makes inference for the projection challenging. We discuss this point in detail in Section 3.4.

### 3.2 Assumptions

Our first assumption is on the parameter space and the criterion function. Below, $\epsilon$ and $M$ are used to denote generic constants which may be different in different appearances.

Assumption 3.1: *$\Theta \subseteq \mathbb{R}^d$ is compact and convex with a nonempty interior.*

Compactness is a standard assumption on $\Theta$ for extremum estimation. In addition we require convexity as we use mean value expansions of $E_P[m_j(X_i, \theta)]$ in $\theta$ as shown in equation (2.9). We then define our model as follows.

Assumption 3.2: *The model $\mathcal{P}$ for $P$ satisfies the following conditions:*

*(i)* $E_P[m_j(X_i, \theta)] \leq 0,\ j = 1, \dots, J_1$ *and* $E_P[m_j(X_i, \theta)] = 0,\ j = J_1 + 1, \cdots, J_1 + J_2$ *for some* $\theta \in \Theta$*;*

*(ii)* $\{X_i, i \geq 1\}$ *are i.i.d. under* $P$*;*

*(iii)* $\sigma^2_{P,j}(\theta) \in (0, \infty)$ *for* $j = 1, \cdots, J$ *for all* $\theta \in \Theta$*;*

*(iv) For constants* $\delta > 0$ *and* $0 < M < \infty$ *and all* $j = 1, \cdots, J$*,*

$$E_P\Big[\sup_{\theta \in \Theta} \Big|\frac{m_j(X_i, \theta)}{\sigma_{P,j}(\theta)}\Big|^{2+\delta}\Big] \leq M; \tag{3.5}$$

*(v) Let* $\tilde{m}(X_i, \theta) \equiv (m_1(X_i, \theta), \cdots, m_{J_1+J_2}(X_i, \theta))'$*. Let* $\tilde{\Omega}_P(\theta) = Corr_P(\tilde{m}(X_i, \theta))$*. The smallest eigenvalue of* $\tilde{\Omega}_P(\theta)$ *is greater than* $\omega$ *for some* $\omega > 0$*.*

*(vi) There is a positive constant* $\epsilon$ *such that* $\Theta_I(P) \subset \Theta^{-\epsilon}$*, where* $\Theta^{-\epsilon} = \{\theta \in \Theta :\ d_H(\theta, \mathbb{R}^d \setminus \Theta) \geq \epsilon\}$*.*

Assumption 3.2 (i)-(iv) based on Andrews and Soares (2010) are standard in the literature. Assumption 3.2 (v) requires that the correlation matrix of the sample moments has eigenvalues uniformly bounded from below. It is used to show that the probability that the nonlinear constraint set is empty while the linearized constraint set is non-empty, is uniformly arbitrarily small. This is used to establish asymptotic validity of our linear approximation. Assumption 3.2 (vi) requires that the identified set is in an $\epsilon$-contraction of the parameter space, where $\epsilon$ is a uniform constant. This implies that the behavior of the support function of $\mathcal{C}_n(c_n)$ is determined only by the moment restrictions asymptotically under any $P \in \mathcal{P}$. This assumption could be dropped if the parameter space can be defined via moment inequalities, e.g. $\Theta = [0, 1]^d$ or similar.

For any sequence of random variables $\{X_n\}$ and a positive sequence $a_n$, we write $X_n = o_{\mathcal{P}}(a_n)$ if for any $\epsilon, \eta > 0$, there is $N \in \mathbb{N}$ such that $\sup_{P\in\mathcal{P}} P(|X_n/a_n| > \epsilon) < \eta, \forall n \geq N$. We also write $X_n = O_{\mathcal{P}}(a_n)$ if for any $\eta > 0$, there is a $M \in \mathbb{R}_+$ and $N \in \mathbb{N}$ such that $\sup_{P\in\mathcal{P}} P(|X_n/a_n| > M) < \eta, \forall n \geq N$. The following assumption collects regularity conditions on the gradient and variance of the moments.

Assumption 3.3: *The model $\mathcal{P}$ for $P$ satisfies the following additional conditions:*

*(i) For each $j$, there exist $D_{P,j}(\theta) \equiv \nabla_\theta\{E_P[m_j(X, \theta)]/\sigma_{P,j}(\theta)\}$ and its estimator $\hat{D}_{j,n}(\theta)$ such that $\sup_{\theta\in\Theta} \|\hat{D}_{j,n}(\theta) - D_{P,j}(\theta)\| = o_{\mathcal{P}}(1)$. Further, there exist $\underline{M}, \overline{M} > 0$ such that $\underline{M} < \|D_{P,j}(\theta)\| \leq \overline{M}$ for all $\theta \in \partial\Theta_I(P)$ and $j = 1, \cdots, J$;*

*(ii) There exists $M > 0$ such that $\max_{j=1,\cdots,J} \|D_{P,j}(\theta) - D_{P,j}(\theta')\| \leq M\|\theta - \theta'\|$ for all $\theta, \theta' \in \Theta$;*

*(iii)* $\sup_{\theta\in\Theta} \max_{j=1,\ldots,J_1+J_2} \left|\frac{\hat{\sigma}_{n,j}(\theta)}{\sigma_{P,j}(\theta)} - 1\right| = O_{\mathcal{P}}(n^{-1/2})$.

Assumption 3.3 requires that the normalized population moment is differentiable, the derivative is Lipschitz continuous, and it can be consistently estimated uniformly in $\theta$ and $P$. We require these conditions because we use a linear expansion of the population moments to obtain a first-order approximation to the support function of $\mathcal{C}_n$ and our bootstrap procedure requires an estimator of the population gradient. Note that, while Assumption 3.3 restricts the norm of the gradient, we do not directly assume that $T_n(\theta)$ is bounded from below by a polynomial function of $\theta$ outside a neighborhood of the identification region, which was assumed in some of the existing work (see e.g. Chernozhukov, Hong, and Tamer, 2007). Assumption 3.3 (iii) requires that an estimator of the asymptotic variance is available and it converges at a parametric rate. This condition holds under regularity conditions on higher moments.

A final set of assumptions is on the normalized empirical process. For this, define the variance semimetric $\rho_P$ by

$$\rho_P(\theta, \theta') \equiv \left\|\left\{Var_P\big(\sigma_{P,j}^{-1}(\theta)m_j(X, \theta) - \sigma_{P,j}^{-1}(\theta')m_j(X, \theta')\big)^{1/2}\right\}_{j=1}^{J}\right\|. \tag{3.6}$$

For each $\theta, \theta' \in \Theta$ and $P$, let $Q_P(\theta, \theta')$ denote a $J$-by-$J$ matrix whose $(j, k)$-th element is the covariance between $m_j(X_i, \theta)/\sigma_{P,j}(\theta)$ and $m_k(X_i, \theta'))/\sigma_{P,k}(\theta')$ under $P$.

Assumption 3.4: *(i) For every $P \in \mathcal{P}$, and $j = 1, \cdots, J$, $\{\sigma_{P,j}^{-1}(\theta)m_j(\cdot, \theta) : \mathcal{X} \to \mathbb{R}, \theta \in \Theta\}$ is a measurable class of functions; (ii) The empirical process $\mathbb{G}_{P,n}$ with $j$-th component*

$\mathbb{G}_{P,j,n}$ *is asymptotically* $\rho_P$*-equicontinuous uniformly in* $P \in \mathcal{P}$. *That is, for any* $\epsilon > 0$,

$$\lim_{\delta \downarrow 0} \limsup_{n\to\infty} \sup_{P\in\mathcal{P}} P^*\Big( \sup_{\rho_P(\theta,\theta')<\delta} \|\mathbb{G}_{P,n}(\theta) - \mathbb{G}_{P,n}(\theta')\| > \epsilon \Big) = 0; \tag{3.7}$$

*(iii)* $Q_P$ *satisfies*

$$\lim_{\delta \downarrow 0} \sup_{\|(\theta_1,\theta_1')-(\theta_2,\theta_2')\|<\delta} \sup_{P\in\mathcal{P}} \|Q_P(\theta_1,\theta_1') - Q_P(\theta_2,\theta_2')\| = 0. \tag{3.8}$$

Under this assumption, the class of normalized moment functions is uniformly Donsker. This allows us to show that the first-order linear approximation to $s(p, \mathcal{C}_n(c_n))$ is valid and further establish the validity of our bootstrap procedure.

## 3.3 Main Results

The following theorem is our first main result, which establishes the asymptotic validity of the confidence interval.

THEOREM 3.1: *Suppose Assumptions 3.1-3.4 hold. Let* $0 < \alpha < 1/2$. *Then,*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Theta_I(P)} P(p'\theta \in [-s(-p, \mathcal{C}_n(\hat{c}_n)), s(p, \mathcal{C}_n(\hat{c}_n))]) \geq 1 - \alpha, \tag{3.9}$$

*where* $\hat{c}_n$ *was defined in equation (3.4).*

Our second results establishes that $CI_n$ is always a subset of a confidence interval obtained by projecting an Andrews and Soares (2010) confidence set. We remark, however, that the confidence set proposed by Andrews and Soares (2010) is built to uniformly cover each vector in $\Theta_I$ with a prespecified asymptotic probability, and as such is designed for a different inferential problem than the one considered here. Below we let $c_n^{AS}$ denote the critical value obtained applying Andrews and Soares (2010) with the criterion function in equation (2.3) and with the same choice of $\hat{\zeta}_n$ and $\kappa_n$ as defined in equations (A.5) and (A.7), respectively.

THEOREM 3.2: *Suppose Assumptions 3.1-3.4 hold. Let* $0 < \alpha < 1/2$. *Then for each* $n \in \mathbb{N}$

$$CI_n \subseteq [-s(-p, \mathcal{C}_n(c_n^{AS})), s(p, \mathcal{C}_n(c_n^{AS}))]. \tag{3.10}$$

## 3.4 The local geometry of the identification region and uniform inference

The local geometry of the identification region affects inference for the projection of $\theta$ in various ways. We illustrate, through examples, how our inference method handles some of the key challenges faced by the existing methods.

We begin with a simple example.

EXAMPLE 3.1: Let $\Theta = [-K, K]^2$ for some $K > 0$ and moment functions be given by

$$m_1(x, \theta) = x^{(1)}(\theta_1 - 1)^2 + \theta_2 - x^{(2)} \tag{3.11}$$

$$m_2(x, \theta) = x^{(3)}(\theta_1 + 1)^2 + \theta_2 - x^{(4)}, \tag{3.12}$$

where we assume $X^{(l)}, l = 1, \cdots, 4$ are i.i.d. random variables with mean $\mu_x \geq 0$ and variance $\sigma_x^2$. The parameter of interest is $\theta_2$. So, we let $p = (0, 1)'$.

The projection of $\theta \in \Theta_I$ is maximized at a unique point $\theta^* = 0$. For simplicity, consider constructing a one-sided confidence interval $CI_n = (-\infty, s(p, \mathcal{C}_n(c_n))]$, where $s(p, \mathcal{C}_n(c_n))$ is defined as in (2.4) with $J_1 = 2$ inequality restrictions with the moment functions in (3.11)-(3.12) and no equality restrictions. Then, $\theta^*$ gives the least favorable case for this one-sided confidence interval.

Consider the following linear program at $\theta = \theta^*$:

$$\begin{aligned} \mathcal{Z}_n(p, c, \theta) = \sup_{\lambda \in \mathbb{R}^2} & \; p'\lambda \\ s.t. & \; \mathbb{G}_{P,n}(\theta) + D_P(\theta)\lambda + h_{P,n}(\theta) \leq c \end{aligned} \tag{3.13}$$

where $\mathbb{G}_{P,1,n}(\theta^*) = \sqrt{n}(\bar{X}^{(1)} + \bar{X}_n^{(2)} - 2\mu_x)/\sqrt{2}\sigma_x$ and $\mathbb{G}_{P,2,n}(\theta^*) = \sqrt{n}(\bar{X}^{(3)} + \bar{X}_n^{(4)} - 2\mu_x)/\sqrt{2}\sigma_x$, the gradient matrix $D_P(\theta^*)$ has rows $D_{P,1}(\theta^*)' = (2\mu_x/\sqrt{2}\sigma_x, 1/\sqrt{2}\sigma_x)$ and $D_{P,2}(\theta^*)' = (-2\mu_x/\sqrt{2}\sigma_x, 1/\sqrt{2}\sigma_x)$, and $h_{P,n}(\theta^*) = (0, 0)'$. This program is infeasible in the sense that it uses unknown population objects, in particular, the knowledge that $\theta^*$ is a point at which both population moment inequalities bind, which implies $h_{P,n}(\theta^*) = (0, 0)'$. Though infeasible, it gives useful insights. Figure 1 shows the original nonlinear constraints and linearized constraints around $\theta^*$ perturbed by $\mathbb{G}_{P,n}$. The key idea of our procedure is to find $c_n(\theta^*)$ such that $\mathcal{Z}_n(p, c_n(\theta^*), \theta^*)$, the value of the perturbed linear program, is greater than or equal to 0 with probability $1 - \alpha$, and use it in the original nonlinear problem upon projecting $\mathcal{C}_n(\cdot)$.

In Example 3.1, the optimal value of the linear program in (3.13) has a closed form, which is $\mathcal{Z}_n(p, c, \theta^*) = p' D_P^{-1}(c - \mathbb{G}_{P,n}) = \sqrt{2}\sigma_x(c - W_n)$, where $W_n = (\mathbb{G}_{P,1,n} + \mathbb{G}_{P,2,n})/2$ has a limiting distribution $N(0, 1/2)$ (under a fixed $(\theta^*, P)$). Therefore, by setting $c_n(\theta^*)$ to 1.15, the 95%-quantile of $N(0, 1/2)$, one can let $\mathcal{Z}_n(p, c_n(\theta^*), \theta^*)$ be greater than or equal to 0 with probability 95% asymptotically.[10] This infeasible critical value is the baseline of our method. In practice, the researcher does not know whether a given $\theta$ is on the boundary of the identification region nor the population objects: the distribution of $\mathbb{G}_{P,n}(\theta)$ and $(D_P(\theta), h_{P,n}(\theta))$.

[10] This argument is based on a pointwise asymptotics, which fixes $(\theta^*, P)$ and sends $n$ to $\infty$. This is done only for illustration purposes to obtain a specific value for $c_n(\theta^*)$. Our proof does not use this argument. Note that the critical value calculated under this pointwise asymptotics depends on the covariance matrix of $\mathbb{G}_{P,n}$. For example, if $corr(\mathbb{G}_{P,1,n}, \mathbb{G}_{P,2,n})$=-0.9, it is enough to set $c_n(\theta^*)$ to 0.37.

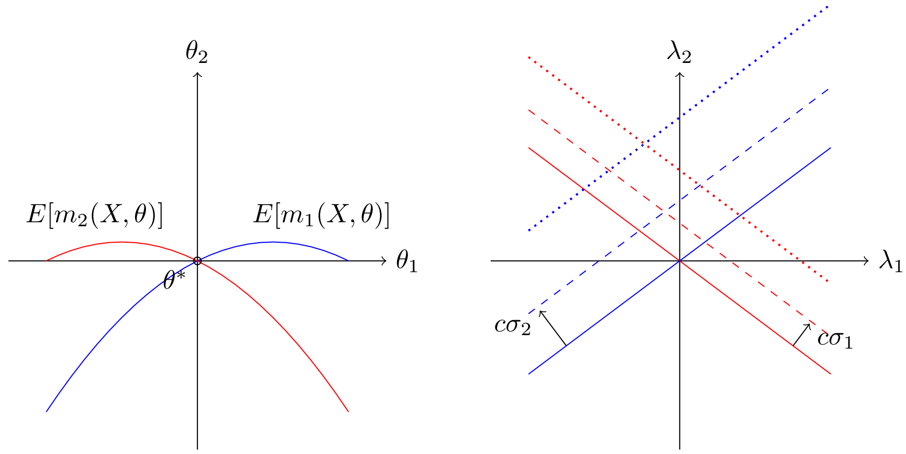


Figure 1: Moment inequalities (left) and linearized constraints (right)

Our bootstrap procedure therefore replaces them with suitable estimators.

An alternative yet closely related way to construct a confidence interval is to use the support function of a sample analog estimator $\mathcal{C}_n(0) \equiv \{\theta \in \Theta : \bar{m}_n(\theta) \leq 0\}$ of the identified set. This approach was taken for example in Pakes, Porter, Ho, and Ishii (2011) (see also Kaido and Santos, 2014, section 4.3). A one-sided confidence interval can be obtained by properly expanding the support function, i.e. $\tilde{CI}_n = (-\infty, s(p, \mathcal{C}_n(0)) + \tau/\sqrt{n}]$, where $\tau$ is the $1-\alpha$ quantile of the limiting distribution of the normalized support function $S_n \equiv \sqrt{n}[s(p, \mathcal{C}_n(0)) - s(p, \Theta_I(P))]$. The normalized support function can be approximated by the linear program in (3.13) with $c = 0$, which is $\mathcal{Z}_n(p, 0, \theta^*) = \sqrt{2}\sigma_x W_n$, and this in turn, suggests the choice $\tau = \sigma_x \times 1.645$. This critical value, however, depends on $\sigma_x$ and is not invariant to scale transformations of moments. This is not desirable as inference methods without scale invariance are known to have poor power properties. (see e.g. Chernozhukov, Kocatulum, and Menzel, 2015). The lack of invariance is due to the fact that the procedure compares the standardized constraints to $c = 0$, which is equivalent to comparing non-standardized constraints to 0. Note that our procedure does not suffer from this issue because, whenever a positive amount of relaxation is necessary, the level $c$ is compared to the standardized moments.

Next, we consider a setting where the projection is maximized at multiple points. For this, we add, to the existing constraints, one more inequality restriction whose moment function is given by

$$m_3(x, \theta) = x^{(5)}\theta_1 + \theta_2 + x^{(6)}, \tag{3.14}$$

where $X^{(5)}$ and $X^{(6)}$ are independent random variables independent from $X^{(1)}, \cdots, X^{(4)}$ with mean $E_P[X^{(5)}] = 0$, $E_P[X^{(6)}] = \mu_x$ and variance $Var_P(X^{(5)}) = Var_P(X^{(6)}) = \sigma_x^2$. (See

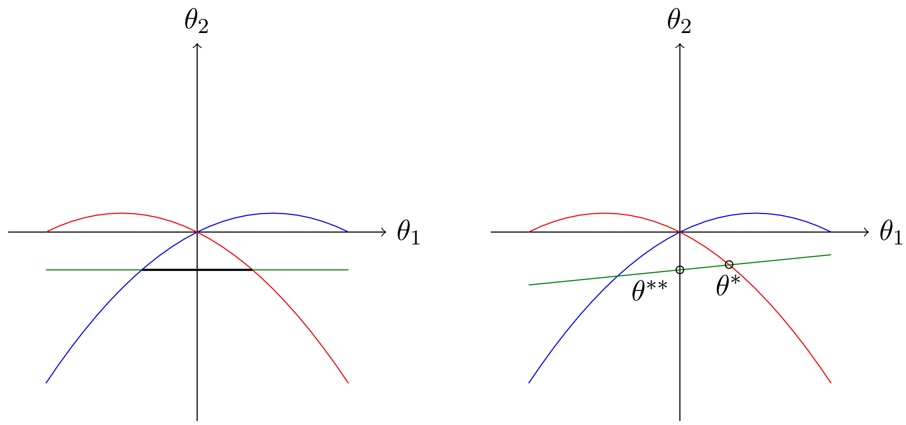


Figure 2: Flat face (left) and a near flat face (right)

Figure 2.)

The projection of $\theta \in \Theta_I$ is then maximized over the following set:

$$H(p, \Theta_I) = \{\theta \in \Theta : \theta_1 \in [1 - \sqrt{2}, -1 + \sqrt{2}], \theta_2 = -\mu_x\}. \tag{3.15}$$

In other words, the identification region has a flat face toward direction $p$. At each $\theta \in H(p, \Theta_I)$, one can study the infeasible linear program. For example, at $\theta^* = (1 - \sqrt{2}, -\mu_x)$, the first and third moment inequalities bind, but not the second one. Then, the approximating linear program in (3.13) holds with $h_{P,n}(\theta^*) = (0, -\sqrt{n}(4-2\sqrt{2})\mu_x, 0)'$. If the magnitude of the second component of $h_{P,n}(\theta^*)$ is large, (or along any sequence $(\theta_n, P_n)$ such that $h_{P_n,2,n}(\theta_n) \to -\infty$), the second moment inequality becomes negligible. Solving for the optimal value using the two remaining constraints then yields $\mathcal{Z}_n(p, c, \theta^*) = \sqrt{2}\sigma_x(c - W_n)$, where $W_n = \mathbb{G}_{P,3,n}(\theta^*)$ approximately follows the standard normal distribution, which suggests that $c_n(\theta^*) = 1.645$, the usual one-sided critical value, can be used. However, if $h_{P,2,n}(\theta^*)$ is close to 0, the second constraint is also relevant. In such cases, our procedure uses the GMS function to replace $h_{P,2,n}(\theta^*)$ with 0 and adds the second inequality as an additional constraint to the linear program. This in turn increases $c_n(\theta^*)$ needed to ensure $\mathcal{Z}_n(p, c_n(\theta^*), \theta^*) \geq 0$ with probability $1 - \alpha$. The same argument applies to every $\theta$ in the support set. For example at $\theta = (0, -\mu_x)$, the third moment inequality is the only one that binds, which again defines another approximating linear program with a different local slackness parameter. Hence, the amount of relaxation needed to ensure the one-sided coverage differs across points in $H(p, \Theta_I)$ due to different values of the slackness parameter.[11] Furthermore, the analysis also extends

[11] On the other hand, if we abstract from the local slackness parameters associated with the slack constraints, the infeasible critical value is common across points in the support set. That is, whenever the magnitude of the slack constraints are $-\infty$, the infeasible critical value is 1.645 at all points in the support set. This is because, in the presence of the third constraint whose gradient is aligned with $p$, the problem reduces to a

to settings where the identification region has a face whose normal vector is nearly aligned with $p$ as shown in Figure 2. We will come back to this case later in this section.

As discussed above, the presence of a flat face or more generally a non-singleton support set does not complicate our inference procedure because we calibrate the level at each $\theta$. On the other hand, these features raise a non-trivial challenge for methods that use test statistics whose limiting distributions depend on $H(p, \Theta_I)$. For example, consider again the method that constructs a confidence interval from the support function of the estimated identified set. If the support set is not a singleton, the distribution of the normalized support function $S_n$ can be shown to be approximated by the supremum of $\mathcal{Z}_n(p, 0, \theta)$ over $H(p, \Theta_I)$. Hence, the support set becomes a nuisance parameter that affects the distribution of the statistic. Uniform size control then becomes challenging. In particular, for a sequence of DGPs $P_n$ along which the support sets are singletons (i.e. $H(p, \Theta_I(P_n)) = \{\theta_n\}$ for all $n$) but non-singleton in the limit, the limiting distribution of the statistic changes in a discontinuous manner. We call such a setting "near flat face". In the present example, one can construct such a sequence $P_n$ by letting $E_{P_n}[X^{(5)}] > 0$ for all $n$ and letting it drift to 0 (see Figure 2). To handle this issue, one needs to either assume away the presence of a flat face (toward direction $p$) or to come up with a way to introduce a conservative distortion. Pakes, Porter, Ho, and Ishii (2011) and Kaido and Santos (2014) (Assumption 4.1) for example take the first approach and assume away flat faces. However, this is not desirable as some of the commonly studied examples in this literature exhibit flat faces.[12] In a related context, the recent work of Bugni, Canay, and Shi (2014) considers testing the hypothesis $H_0 : f(\theta) = \gamma$ and constructing a confidence interval through a test inversion. Their procedure can be employed to make inference for the projection of $\theta$ by taking $f(\theta) = p'\theta$. Using a profiled test statistic $\inf_{\{\theta:p'\theta=\gamma\}} a_n Q_n(\theta)$, where $Q_n$ is a sample criterion function, Bugni, Canay, and Shi (2014) show that, one can make uniformly valid inference by calculating a critical value via bootstrap, while conservatively approximating the local slackness parameters and by constructing an estimator of the parameter set $\Xi_I(\gamma) \equiv \{\theta \in \Theta_I : p'\theta = \gamma\}$. Note that $\Xi_I(\gamma)$ coincides with the support set when $\gamma$ equals $s(p, \Theta_I)$. Although their inference is valid over a class of distributions under which $\Xi_I(\gamma)$ is not necessarily singleton-valued, they require that the population criterion function increases as a polynomial function of the distance from $\theta$ to $\Xi_I(\gamma)$ when $\theta$ deviates from this set along the hyperplane $\{\theta : p'\theta = \gamma\}$.[13] This requirement, however, excludes data generating processes that exhibit near flat faces. For example, in

one-sided testing problem.

[12]For example, Beresteanu and Molinari (2008) show that the identification region for the best linear predictor of an interval-valued outcome variable with discrete covariates has flat faces. See also Freyberger and Horowitz (2013) for a nonparametric IV example with discrete variables.

[13]Without this requirement, their estimator of $\Xi_I(\gamma)$ may include points at which population moment (in)equalities are violated but by not much. At such points, the sample moment inequalities may even realize as slack constraints, and hence replacing the (violated) population local slackness parameter with the GMS function does not necessarily provide conservative approximations. For details, we refer to discussions provided in Bugni, Canay, and Shi (2015) (page 265).

the right panel of Figure 2, consider deviating from $\theta^*$ toward direction $(-1, 0)$. Because of the third constraint tending to a flat face, one can make the population criterion function increase arbitrarily slowly along such a deviation.

Recall that our bootstrap procedure in Section 3.1 imposed the additional constraint $\lambda \in \rho B_d$. Below, we discuss the role of this constraint using Example 3.1.

Recall that $D_P(\theta^*)$ has rows:

$$D_{P,1}(\theta^*)' = (2\mu_x/\sqrt{2}\sigma_x, 1/\sqrt{2}\sigma_x), \quad \text{and} \quad D_{P,2}(\theta^*)' = (-2\mu_x/\sqrt{2}\sigma_x, 1/\sqrt{2}\sigma_x). \tag{3.16}$$

Consider a sequence of DGPs such that $\mu_x \to 0$. As we saw before, under each DGP with $\mu_x > 0$, the infeasible linear program calibrates $c_n(\theta^*) = 1.15$. In the limit, however, the moment inequalities reduce to the following restrictions:

$$\theta_2 - E_P[X^{(2)}] \leq 0 \tag{3.17}$$

$$\theta_2 - E_P[X^{(4)}] \leq 0. \tag{3.18}$$

In other words, $\theta_2$'s upper bound is given by the minimum of the two means: $E_P[X^{(2)}]$ and $E_P[X^{(4)}]$. This structure is also known as "intersection bounds" (Hirano and Porter, 2012). The value of the linear program in (3.13) is then $\mathscr{Z}_n(p, c, \theta^*) = \min\{c - \mathbb{G}_{1,n}, c - \mathbb{G}_{2,n}\}$. This suggests that one would need a two-sided critical value, $c_n(\theta^*) = 1.96$ instead of 1.15 to ensure that $\mathscr{Z}_n(p, c_n(\theta^*), \theta^*) \geq 0$ with probability at least 95%. This is another challenge for the uniform validity of inference. For any setting where the constraints are close to the minimum of the two means, any inference method that does not take into account this feature would have poor size control.

This type of example is the main reason we restrict the localization parameter $\lambda$ into the $\rho$-box. Adding the $\rho$-box constraint to the linear program in (3.13) lowers the value of $\mathscr{Z}_n(p, c, \theta)$ for each $c$, which means that one needs a higher $c_n(\theta)$ to ensure that $\mathscr{Z}_n(p, c_n(\theta), \theta) \geq 0$ with a prescribed probability. Hence, adding the $\rho$-box constraint makes the critical value weakly more conservative. How this modification works in the example above is described in Figure 3. The figure shows the data generating process on the left panel and a realization of a constraint in the bootstrap problem in (3.2). Note that, due to the sampling variation, the estimated gradients $\hat{D}_{1,n}$ and $\hat{D}_{2,n}$ differ slightly from the population gradients. Without the $\rho$-box constraint, the maximum is attained at $\lambda^*$. Since the estimated gradients are fixed across bootstrap replications, $p'\lambda^*$ behaves as approximately normal, and by the previous argument we would end up with $c_n(\theta^*) = 1.15$. With the $\rho$-box, however, the optimum is attained at $\lambda^{**}$ whose projection is the minimum of the projections of two points at which the two constraints intersect with the right boundary of $\rho B_d$. Therefore, our bootstrap procedure mimics the minimum of the two-means problem. This is true whenever the population gradients are close to this situation, and hence restricting $\lambda$ to the $\rho$-box provides a conservative distortion and plays a key role in establishing the uniform validity of our procedure.

The near flat face example in Figure 2 can be handled analogously. For example, for some points such as $\theta^{**}$, the relevant constraint is the third constraint (near flat face). A linearized problem around $\theta^{**}$ then looks akin to the right panel of Figure 3 without the red line. Calculating a bootstrap critical value then yields a one-sided critical value $c_n(\theta) = 1.645$ (approximately) as before.

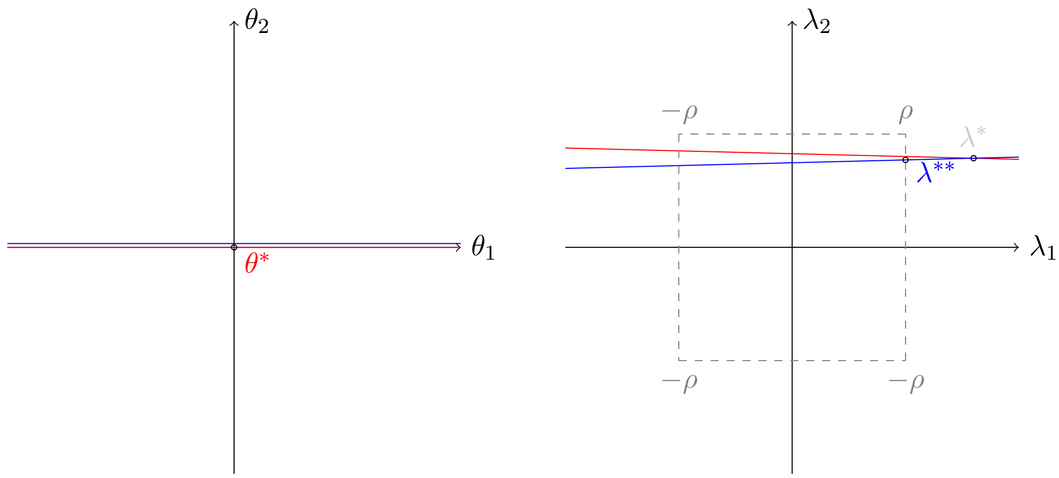


Figure 3: Minimum of two means and a $\rho$-box

## 4 Computational Aspects of the Method

TBA

## 5 Monte Carlo Simulations

TBA

## 6 Conclusions

TBA

# A Definitions, Notations, and Proofs of Main Theorems

## A.1 Objects related to normalized empirical process

For each $j$ and $\theta \in \Theta$, let $\bar{m}_{j,n}(\theta) \equiv \frac{1}{n}\sum_{i=1} m_j(X_i,\theta)$. Define

$$\mathbb{G}_{P,j,n}(\theta) \equiv \sqrt{n}(\bar{m}_{j,n}(\theta) - E_P[m_j(X_i,\theta)])/\sigma_{P,j}(\theta). \tag{A.1}$$

Let $\mathbb{G}_{P,n} \equiv (\mathbb{G}_{P,1,n},\cdots,\mathbb{G}_{P,J,n})$. This is a vector of empirical processes on $\Theta$.

Under Assumptions A.1-A.4 in Bugni, Canay, and Shi (2015)(Assumptions 3.2-(iv) and 3.4 in the main text), one obtains various desirable properties (Lemma C.1 in Bugni, Canay, and Shi (2014)) including the Donsker property, so that for any $\epsilon,\eta > 0$, there exists $N \in \mathbb{N}$ such that $\sup_{P\in\mathcal{P}} \rho_{BL}(\mathbb{G}_{P,n},\mathbb{G}_P) = o(1)$, where $\mathbb{G}_P$ is a vector of Gaussian processes with covariance kernel $Q_P(\theta,\theta')$ whose $(j,k)$-th element is given by $E_P[(m_j(X_i,\theta) - E_P[m_j(X_i,\theta)])(m_k(X_i,\theta') - E_P[m_k(X_i,\theta')])]/(\sigma_{j,P}(\theta)\sigma_{k,P}(\theta'))$. We then let $\Omega_P(\theta) \equiv Q_P(\theta,\theta)$. One also obtains that $\hat{\sigma}_{n,j}/\sigma_{P,j} \to^p 1$ uniformly in $P$ and $\theta$.

Define $h_{P,j,n}(\theta) \equiv \sqrt{n}E_P[m_j(X_i,\theta)]/\sigma_{P,j}(\theta)$. Let $D_P(\theta)$ be a $J\times d$ matrix whose $j$-th row is $D_{P,j}(\theta)'$. Let $\mathbb{G}^b_{j,n}(\theta)$ be the normalized bootstrap empirical process $\mathbb{G}^b_{j,n}(\theta) = \frac{1}{\sqrt{n}}(\bar{m}^b_{j,n}(\theta) - \bar{m}_{j,n}(\theta))/\hat{\sigma}^b_{j,n}(\theta)$, where for a random sample $\{X^b_1,\cdots,X^b_n\}$ from the empirical distribution, $\bar{m}^b_{j,n}(\theta) \equiv \frac{1}{n}\sum_{i=1} m_j(X^b_i,\theta)$ and $\hat{\sigma}^b_{j,n}(\theta) \equiv (\{\frac{1}{n}\sum_{i=1}^n m(X^b_i,\theta) - \bar{m}^b_{j,n}(\theta)\}^2)^{1/2}$.

Let

$$h^\Delta_n = \sqrt{n}(s(p,\Theta_I(P)) + s(-p,\Theta_I(P))) \tag{A.2}$$

denote the length of the projection of $\Theta_I(P)$ normalized by $\sqrt{n}$. (Recall from the definitions that the projection of any $\theta \in \Theta_I(P)$ to the one-dimensional subspace spanned by the direction $p$ lies in the interval $[-s(-p,\Theta_I(P)), s(p,\Theta_I(P))]$.)

## A.2 Objects related to critical values

Let $B_d \equiv [-1,1]^d$ be the unit-box in $\mathbb{R}^d$. Let $\varphi : \mathbb{R}^J_{[\pm\infty]} \to \mathbb{R}^J_{[-\infty]}$ be defined componentwise by

$$\varphi_j(\xi_j) \equiv \begin{cases} 0 & \text{if } \xi_j \geq -1 \\ -\infty & \text{if } \xi_j < -1 \end{cases}. \tag{A.3}$$

Let $\varphi^* : \mathbb{R}^J_{[\pm\infty]} \to \mathbb{R}^J_{[-\infty]}$ be a function such that

1. $\varphi^*_j(\xi_j) \leq \varphi_j(\xi_j)$ for all $\xi_j \in \mathbb{R}_{[-\infty]}$,
2. $\varphi^*_j(\cdot)$ is continuous,
3. $\varphi^*_j(\xi_j) = 0$ for all $\xi_j \geq 0$ and $\varphi^*_j(-\infty) = -\infty$.

The existence of this function is ensured by Lemma D.8 in Bugni, Canay, and Shi (2015). Define

$$\tilde{\zeta}_{P,j,n}(\theta) \equiv \varphi^*(\kappa_n^{-1}h_{P,j,n}(\theta)), \tag{A.4}$$

$$\hat{\zeta}_{j,n}(\theta) \equiv \varphi(\kappa_n^{-1}\bar{m}_{j,n}(\theta)/\hat{\sigma}_{j,n}(\theta)) \tag{A.5}$$

$$\hat{\zeta}^*_{j,n}(\theta) \equiv \varphi^*(\kappa_n^{-1}\bar{m}_{j,n}(\theta)/\hat{\sigma}_{j,n}(\theta)), \tag{A.6}$$

with

$$\kappa_n = o(n^{-1/2} \tag{A.7}$$

Throughout, we use the following auxiliary linear programming problem and denote its value function by $U : \mathbb{R}^J \times \mathbb{R}^{J\times d} \times \mathbb{R}^J \times \mathbb{R}_+ \times \mathbb{R}_+$.

$$U(\alpha,\beta,\gamma,\delta,\rho) \equiv \sup_{\lambda\in\rho B_d} \langle p,\lambda\rangle$$

$$s.t.\ \alpha_j + \beta'_j\lambda + \gamma_j \leq \delta,\ j = 1,\cdots,J. \tag{A.8}$$

Let

$$\Lambda(\Xi, \nu, \delta) = \{\lambda \in \mathbb{R}^d : \Xi_j \lambda \leq \delta - \nu_j\}, \tag{A.9}$$

with $\nu_j = \alpha_j + \gamma_j$ for $j = 1, \ldots, J$, $\nu_j = \rho$ for $j = J+1, \ldots, J+2d$, yielding a $(J + 2d) \times 1$ vector, and $\Xi$ a $(J + 2d) \times d$ matrix collecting in the first $J$ rows the vectors $\beta_j$, and below it the matrices $I_d$ and $-I_d$.

We then define the following processes:

$$\mathcal{Z}^{\rho}_{h,n}(p, c, \theta) \equiv U(\mathbb{G}_P(\theta), D_P(\theta), h_{P,n}(\theta), c, \rho) \tag{A.10}$$

$$\mathcal{Z}^{\mathbb{G},\rho}_{n}(p, c, \theta) \equiv U(\mathbb{G}_P(\theta), D_P(\theta), \tilde{\zeta}_{P,n}(\theta), c, \rho) \tag{A.11}$$

$$\mathcal{Z}^{b,\rho}_{n}(p, c, \theta) \equiv U(\mathbb{G}^b_n(\theta), \hat{D}_n(\theta), \hat{\zeta}_n(\theta), c, \rho) \tag{A.12}$$

$$\mathcal{Z}^{b*,\rho}_{n}(p, c, \theta) \equiv U(\mathbb{G}^b_n(\theta), \hat{D}_n(\theta), \hat{\zeta}^*_n(\theta), c, \rho). \tag{A.13}$$

We then define the corresponding critical values by

$$c_{h,n}(\theta) \equiv \inf\{c \in \mathbb{R}_+ : P(\mathcal{Z}^{\rho}_{h,n}(p, c, \theta) \geq 0 \geq -\mathcal{Z}^{\rho}_{h,n}(-p, c, \theta)) \geq 1 - \alpha\} \tag{A.14}$$

$$c_{n}(\theta) \equiv \inf\{c \in \mathbb{R}_+ : P(\mathcal{Z}^{\mathbb{G},\rho}_{n}(p, c, \theta) \geq 0 \geq -\mathcal{Z}^{\mathbb{G},\rho}_{n}(-p, c, \theta)) \geq 1 - \alpha\} \tag{A.15}$$

$$\hat{c}_{n}(\theta) \equiv \inf\{c \in \mathbb{R}_+ : P(\mathcal{Z}^{b,\rho}_{n}(p, c, \theta) \geq 0 \geq -\mathcal{Z}^{b,\rho}_{n}(-p, c, \theta)) \geq 1 - \alpha\} \tag{A.16}$$

$$\hat{c}^*_{n}(\theta) \equiv \inf\{c \in \mathbb{R}_+ : P(\mathcal{Z}^{b*,\rho}_{n}(p, c, \theta) \geq 0 \geq -\mathcal{Z}^{b*,\rho}_{n}(-p, c, \theta)) \geq 1 - \alpha\}. \tag{A.17}$$

## A.3 Relevant sets and functions of $\lambda$

For each $\lambda \in \mathbb{R}^d$, define

$$u_{j,n,\theta}(\lambda) \equiv \sqrt{n}\bar{m}_{j,n}(\theta + \lambda/\sqrt{n})/\hat{\sigma}_j(\theta + \lambda/\sqrt{n}) - c_n(\theta + \lambda/\sqrt{n}) \tag{A.18}$$

$$v_{j,n,\theta}(\lambda) \equiv \mathbb{G}_{P,j,n}(\theta) + D_{P,j}(\theta)'\lambda + h_{P,j,n}(\theta) - c_n(\theta). \tag{A.19}$$

$$w_{j,n,\theta}(\lambda) \equiv \mathbb{G}_{P,j}(\theta) + D_{P,j}(\theta)'\lambda + h_{P,j,n}(\theta) - c_n(\theta). \tag{A.20}$$

Below, for each $\theta \in \Theta$ and $\rho > 0$ we let the associated level sets be defined as

$$U_n(\theta) \equiv \big\{\lambda \in \rho B_d : u_{j,n,\theta}(\lambda) \leq 0, \ \forall j = 1, \ldots, J\big\}, \tag{A.21}$$

$$V_n(\theta) \equiv \big\{\lambda \in \rho B_d : v_{j,n,\theta}(\lambda) \leq 0, \ \forall j = 1, \ldots, J\big\}, \tag{A.22}$$

$$W_n(\theta) \equiv \big\{\lambda \in \rho B_d : w_{j,n,\theta}(\lambda) \leq 0, \ \forall j = 1, \ldots, J\big\}. \tag{A.23}$$

Let

$$V_n^{-\delta}(\theta) \equiv \big\{\lambda \in \rho B_d : v_{j,n,\theta}(\lambda) \leq -\delta, \ \forall j = 1, \ldots, J\big\}, \tag{A.24}$$

$$W_n^{-\delta}(\theta) \equiv \big\{\lambda \in \rho B_d : w_{j,n,\theta}(\lambda) \leq -\delta, \ \forall j = 1, \ldots, J\big\}. \tag{A.25}$$

Let $K_P$ be a $(J + 2d) \times d$ matrix collecting the first $J$ rows of the matrix $D_P$, and below it the matrices $I_d$ and $-I_d$.

## A.4 Notation on convergence

For any sequence of random variables $\{X_n\}$ and a positive sequence $a_n$, we write $X_n = o_{\mathcal{P}}(a_n)$ if for any $\epsilon, \eta > 0$, there is $N \in \mathbb{N}$ such that $\sup_{P \in \mathcal{P}} P(|X_n/a_n| > \epsilon) < \eta, \forall n \geq N$. We also write $X_n = O_{\mathcal{P}}(a_n)$ if for any $\eta > 0$, there is a $M \in \mathbb{R}_+$ and $N \in \mathbb{N}$ such that $\sup_{P \in \mathcal{P}} P(|X_n/a_n| > M) < \eta, \forall n \geq N$.

## A.5 Proofs of Main Theorems

Proof of Theorem 3.1: Following Andrews and Guggenberger (2009), we index distributions by a vector of nuisance parameters relevant for the asymptotic size. For this, let $\gamma \equiv (\gamma_1, \gamma_2, \gamma_3)$, where $\gamma_1 = (\gamma_{1,1}, \cdots, \gamma_{1,J})$

with

$$\gamma_{1,j}(\theta) = \sigma_{P,j}^{-1}(\theta) E_P[m_j(X_i, \theta)],\ j = 1, \cdots, J, \tag{A.26}$$

$\gamma_2 = (vech(\Omega_P(\theta)), vec(D_P(\theta)))$, and $\gamma_3 = P$. Let $\{P_{\gamma_n}, \theta_n\} \in \{(P, \theta) : P \in \mathcal{P}, \theta \in \Theta_I(P)\}$ be a sequence such that

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Theta_I(P)} P(p'\theta \in CI_n) = \liminf_{n\to\infty} P_{\gamma_n}(p'\theta_n \in CI_n). \tag{A.27}$$

We then let $\{l_n\}$ be a subsequence of $\{n\}$ such that

$$\liminf_{n\to\infty} P_{\gamma_n}(p'\theta_n \in CI_n) = \lim_{n\to\infty} P_{\gamma_{l_n}}(p'\theta_{l_n} \in CI_{l_n}). \tag{A.28}$$

Recall that

$$CI_{l_n} = [-s(-p, \mathcal{C}_{l_n}(\hat{c}_{l_n})), s(p, \mathcal{C}_{l_n}(\hat{c}_{l_n}))].$$

By Lemma B.1, if $h_{l_n}^{\Delta} \le \kappa_{l_n}^{1/5}$ it suffices to show that $\lim_{n\to\infty} P_{\gamma_{l_n}}\big(s(p, U_{l_n}(\theta_{l_n})) \ge 0 \ge -s(-p, U_{l_n}(\theta_{l_n}))\big) \ge 1-\alpha$. If $h_{l_n}^{\Delta} > \kappa_{l_n}^{1/5}$ it suffices to show that for each $\theta_{l_n}^1 \in H(-p, \Theta_I(P))$ and for each $\theta_{l_n}^2 \in H(p, \Theta_I(P))$, $\min\{\lim_{n\to\infty} P_{\gamma_{l_n}}(s(-p, U_{l_n}(\theta_{l_n}^1)) \ge 0)\ ,\ \lim_{n\to\infty} P_{\gamma_{l_n}}(s(p, U_{l_n}(\theta_{l_n}^2)) \ge 0)\} \ge 1-\alpha$. We argue explicitly for the case that $h_{l_n}^{\Delta} \le \kappa_{l_n}^{1/5}$; the other case is similar.

Define $A_{1n} \equiv \{W_{l_n}(\theta_{l_n}) \ne \emptyset \cap V_{l_n}(\theta_{l_n}) \ne \emptyset \cap U_{l_n}(\theta_{l_n}) \ne \emptyset)\}$. It then follows that

$$\begin{aligned}
&P_{\gamma_{l_n}}\big(s(p, U_{l_n}(\theta_{l_n})) \ge -\epsilon \cap s(-p, U_{l_n}(\theta_{l_n})) \ge -\epsilon\big) \\
&\ge P_{\gamma_{l_n}}\big(\{s(p, U_{l_n}(\theta_{l_n})) \ge -\epsilon\} \cap \{s(-p, U_{l_n}(\theta_{l_n})) \ge -\epsilon\} \cap A_{1n}\big) \\
&\ge P_{\gamma_{l_n}}\big(\big\{s(p, V_{l_n}(\theta_{l_n})) \ge 0\big\} \cap \big\{s(-p, V_{l_n}(\theta_{l_n})) \ge 0\big\} \cap A_{1n}\big) \\
&- P_{\gamma_{l_n}}\big(\big\{\max_{q\in\{p,-p\}} \big|s(q, U_{l_n}(\theta_{l_n})) - s(q, V_{l_n}(\theta_{l_n}))\big| \ge \epsilon\big\} \cap A_{1n}\big) \\
&\ge P_{\gamma_{l_n}}\big(\big\{s(p, V_{l_n}(\theta_{l_n})) \ge 0\big\} \cap \big\{s(-p, V_{l_n}(\theta_{l_n})) \ge 0\big\} \cap A_{1n}\big) - \epsilon
\end{aligned}$$

for $n$ sufficiently large, where the last inequality follows from Lemma B.5, observing that by Lemma B.2 (ii)-(iv), $c_n$ satisfies the conditions of Lemma B.5.

Taking limits as $n \to \infty$ and noting that $\epsilon$ is arbitrary, we have

$$\lim_{n\to\infty} P_{\gamma_{l_n}}\big(s(p, U_{l_n}(\theta_{l_n})) \ge 0 \ge -s(-p, U_{l_n}(\theta_{l_n}))\big) \ge \lim_{n\to\infty} P_{\gamma_{l_n}}\big(\big(s(p, U_{l_n}(\theta_{l_n})) \ge 0 \ge -s(-p, U_{l_n}(\theta_{l_n}))\big) \cap A_{1n}\big). \tag{A.29}$$

Define $A_{2n} \equiv \big\{W_{l_n}(\theta_{l_n}) \ne \emptyset \cap V_{l_n}(\theta_{l_n}) \ne \emptyset\big\}$ and note that $A_{2n} = A_{1n} \cup \big\{W_{l_n}(\theta_{l_n}) \ne \emptyset \cap V_{l_n}(\theta_{l_n}) \ne \emptyset \cap U_{l_n}(\theta_{l_n}) = \emptyset\big\}$. Therefore,

$$\begin{aligned}
&P_{\gamma_{l_n}}\big(\big\{s(p, V_{l_n}(\theta_{l_n})) \ge 0 \ge -s(-p, V_{l_n}(\theta_{l_n}))\big\} \cap A_{1n}\big) \\
&\quad = P_{\gamma_{l_n}}\big(\big\{s(p, V_{l_n}(\theta_{l_n})) \ge 0 \ge -s(-p, V_{l_n}(\theta_{l_n}))\big\} \cap A_{2n}\big) \\
&\quad\quad - P_{\gamma_{l_n}}\big(\big\{s(p, V_{l_n}(\theta_{l_n})) \ge 0 \ge -s(-p, V_{l_n}(\theta_{l_n}))\big\} \cap \big\{W_{l_n}(\theta_{l_n}) \ne \emptyset \cap V_{l_n}(\theta_{l_n}) \ne \emptyset \cap U_{l_n}(\theta_{l_n}) = \emptyset\big\}\big) \\
&\quad \ge P_{\gamma_{l_n}}\big(\big\{s(p, V_{l_n}(\theta_{l_n})) \ge 0 \ge -s(-p, V_{l_n}(\theta_{l_n}))\big\} \cap A_{2n}\big) - \eta,
\end{aligned} \tag{A.30}$$

for $n$ sufficiently large where the second inequality follows from Lemma B.4. Taking limits as $n \to \infty$ and noting that $\eta > 0$ is arbitrary, we have

$$\begin{aligned}
&\lim_{n\to\infty} P_{\gamma_{l_n}}\big(\big\{s(p, V_{l_n}(\theta_{l_n})) \ge 0 \ge -s(-p, V_{l_n}(\theta_{l_n}))\big\} \cap A_{1n}\big) \\
&\ge \lim_{n\to\infty} P_{\gamma_{l_n}}\big(\big\{s(p, V_{l_n}(\theta_{l_n})) \ge 0 \ge -s(-p, V_{l_n}(\theta_{l_n}))\big\} \cap A_{2n}\big).
\end{aligned} \tag{A.31}$$

Recall that for each $q \in \{p, -p\}$, $s(q, V_n(\theta))$ and $\mathcal{Z}_h^{\rho}(q, c_n, \theta)$ are both value functions of linear programs, where the program defining $s(q, V_n(\theta))$ involves the empirical process $\mathbb{G}_{P,j,n}$, while that for $\mathcal{Z}_h^{\rho}(q, c_n, \theta)$ involves the limit Gaussian process $\mathbb{G}_{P,j}$. Since $\mathbb{G}_{P,n}$ converges weakly to $\mathbb{G}_P$ in $\mathcal{C}(\Theta)$, there is a Skorokhod representation $\mathbb{G}_{P,n}^*$ and $\mathbb{G}_P^*$ on some probability space $(\Omega, \mathbf{P})$ such that $\mathbb{G}_{P,n}^* \overset{d}{=} \mathbb{G}_{P,n}$ and $\mathbb{G}_P^* \overset{d}{=} \mathbb{G}_P$ and $\mathbb{G}_{P,n}^* \overset{a.s.}{\to} \mathbb{G}^*$. Hence, for any $n$, replacing $\mathbb{G}_{P,n}^*(\theta)$ with $\mathbb{G}_P^*(\theta)$ in the linear programming problem corresponds to only shifting the constraints while keeping the gradient fixed. By Theorem 14 in Wets (1985), the value function of the linear program is continuous. Therefore, for any $\eta > 0$ and any sequence $\theta_n \in \Theta$, there is $\delta > 0$ such that $\max_j |\mathbb{G}^*_{P,j,n}(\theta_n) - \mathbb{G}^*_{P,j}(\theta_n)| < \delta$ implies $|s(q, V_n(\theta_n)) - \mathcal{Z}^\rho_h(q, c_n, \theta_n)| < \eta$, and $P_{\gamma_n}(\max_j |\mathbb{G}_{P,j,n}(\theta_n) - \mathbb{G}_{P,j}(\theta_n)| < \delta) = \mathbf{P}(\max_j |\mathbb{G}^*_{P,j,n}(\theta_n) - \mathbb{G}^*_{P,j}(\theta_n)| < \delta) \to 1$ as $n \to \infty$. Therefore, for any $\eta > 0$, we have

$$\begin{aligned}
&\lim_{n\to\infty} P_{\gamma_{l_n}}\big(\{s(p, V_{l_n}(\theta_{l_n})) \geq -\eta\} \cap \{s(-p, V_{l_n}(\theta_{l_n})) \geq -\eta\} \cap A_{2n}\big)\\
&\quad \geq \lim_{n\to\infty} P_{\gamma_{l_n}}\big(\{\mathcal{Z}^\rho_h(p, c_{l_n}, \theta_{l_n}) \geq 0 \geq -\mathcal{Z}^\rho_h(-p, c_{l_n}, \theta_{l_n})\} \cap A_{2n}\big)\\
&\quad - P_{\gamma_{l_n}}\big(\{\max_{q\in\{p,-p\}} |s(q, V_n(\theta_{l_n})) - \mathcal{Z}^\rho_h(q, c_{l_n}, \theta_{l_n})| \geq \eta\} \cap A_{2n}\big)\\
&\quad \geq \lim_{n\to\infty} P_{\gamma_{l_n}}\big(\{\mathcal{Z}^\rho_h(p, c_{l_n}, \theta_{l_n}) \geq 0 \geq -\mathcal{Z}^\rho_h(-p, c_{l_n}, \theta_{l_n})\} \cap A_{2n}\big)\\
&\quad - P_{\gamma_{l_n}}\big(\{\max_j |\mathbb{G}_{P,j,l_n}(\theta_{l_n}) - \mathbb{G}_{P,j}(\theta_{l_n})| \geq \delta\} \cap A_{2n}\big)\\
&\quad \geq \lim_{n\to\infty} P_{\gamma_{l_n}}\big(\{\mathcal{Z}^\rho_h(p, c_{l_n}, \theta_{l_n}) \geq 0 \geq -\mathcal{Z}^\rho_h(-p, c_{l_n}, \theta_{l_n})\} \cap \{W_{l_n}(\theta_{l_n}) \neq \emptyset\}\big)\\
&\quad - P_{\gamma_{l_n}}\big(\{W_{l_n}(\theta_{l_n}) \neq \emptyset \cap V_{l_n}(\theta_{l_n}) = \emptyset\}\big)\\
&\quad \geq \lim_{n\to\infty} P_{\gamma_{l_n}}\big(\{\mathcal{Z}^\rho_h(p, c_{l_n}, \theta_{l_n}) \geq 0 \geq -\mathcal{Z}^\rho_h(-p, c_{l_n}, \theta_{l_n})\}\big) - \eta\\
&\quad \geq \lim_{n\to\infty} P_{\gamma_{l_n}}\big(\{\mathcal{Z}^\rho_h(p, c_{h,l_n}, \theta_{l_n}) \geq 0 \geq -\mathcal{Z}^\rho_h(-p, c_{h,l_n}, \theta_{l_n})\}\big) - \eta \geq 1 - \alpha - \eta,
\end{aligned} \tag{A.32}$$

for $n$ sufficiently large where the fourth inequality follows from $\{W_{l_n}(\theta_{l_n}) \neq \emptyset\} = A_{2n} \cup \{W_{l_n}(\theta_{l_n}) \neq \emptyset \cap V_{l_n}(\theta_{l_n}) = \emptyset\}$ and Lemma B.3 and the fifth inequality follows from the fact that $c_n \geq c_{h,n}$ with probability approaching 1 by Lemma B.2 and that $\mathcal{Z}^\rho_h(p, c_{l_n}, \theta_{l_n}) \geq 0 \geq -\mathcal{Z}^\rho_h(-p, c_{l_n}, \theta_{l_n}$ implies $W_{l_n}(\theta_{l_n}) \neq \emptyset$. Since $\eta$ is arbitrary, we then have

$$\lim_{n\to\infty} P_{\gamma_{l_n}}\big(\{s(p, V_{l_n}(\theta_{l_n})) \geq 0 \geq -s(-p, V_{l_n}(\theta_{l_n}))\} \cap A_{2n}\big) \geq 1 - \alpha. \tag{A.33}$$

The conclusion of the lemma now follows from (A.27), (A.28),(A.29), (A.31), and (A.33). □

Proof of Theorem 3.2 To establish the result, observe that for any given $\theta \in \Theta$ the event

$$\max_{j=1,\dots,J}\left\{\frac{\sqrt{n}\bar{m}^*_j(\theta)_+}{\hat{\sigma}_{n,j}} + \hat{\zeta}_{j,n}(\theta)\right\} \leq c \tag{A.34}$$

implies the event

$$\left\{\sup_{\lambda\in\rho B_d} p'\lambda : \max_{j=1,\dots,J}\left\{\frac{\sqrt{n}\bar{m}^*_j(\theta)_+}{\hat{\sigma}_{n,j}} + \hat{D}_{n,j}(\theta)\lambda + \hat{\zeta}_{j,n}(\theta)\right\} \leq c\right\} \geq 0. \tag{A.35}$$

This is so because if $\max_{j=1,\dots,J}\left\{\frac{\sqrt{n}\bar{m}^*_j(\theta)_+}{\hat{\sigma}_{n,j}} + \hat{\zeta}_{j,n}(\theta)\right\} \leq c$, then due to $\hat{\zeta}_{j,n}(\theta) \leq 0$, $\lambda = 0$ is feasible in the outer maximization problem in (A.35), hence the value of (A.35) is greater than or equal to $p'0 = 0$. In turn this yields that $\hat{c}_n \leq c^{AS}_n$.

□

# B Main Lemmas

Fix $\rho > 0$ as discussed in Section 3.4. In all Lemmas below, $\alpha$ is assumed less than 1/2. Throughout, $h^\Delta_n$ is as defined in equation A.2.

Lemma B.1: *Suppose Assumptions 3.1-3.4 hold. Let $\hat{c}_n$ be defined as in (A.16). If $h^\Delta_n \leq \kappa_n^{1/5}$, it follows that*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Theta_I(P)} P(p'\theta \in CI_n) \geq \liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Theta_I(P)} P\Big(s(p, U_n(\theta)) \geq 0 \geq -s(-p, U_n(\theta))\Big). \tag{B.1}$$

*If $h_n^\Delta > \kappa_n^{1/5}$, it follows that*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Theta_I(P)} P(p'\theta \in CI_n)$$
$$\geq \min\left\{ \liminf_{n\to\infty} \inf_{P\in\mathcal{P}} P(s(-p, U_n(\theta_n^1)) \geq 0)\ ,\ \liminf_{n\to\infty} \inf_{P\in\mathcal{P}} P(s(p, U_n(\theta_n^2)) \geq 0) \right\},$$

*for each $\theta_n^1 \in H(-p, \Theta_I(P))$ and $\theta_n^2 \in H(p, \Theta_I(P))$.*

Lemma B.2: *Suppose Assumptions 3.1-3.4 hold. Let $\hat{c}_n$ be defined as in (A.16). Let $\Upsilon(P) = \partial\Theta_I(P)$ if $h_n^\Delta > \kappa_n^{1/5}$, and $\Upsilon(P) = \Theta_I(P)$ if $h_n^\Delta \leq \kappa_n^{1/5}$. Then, there exists a sequence $\{c_n : \Theta \to \mathbb{R}_+, n = 1, 2, \cdots\}$ such that (i) $\hat{c}_n(\theta) \geq c_n(\theta)$ with probability approaching 1 uniformly over $\{(P, \theta) : P \in \mathcal{P}, \theta \in \Upsilon(P)\}$; (ii) for all $P \in \mathcal{P}$, $c_n(\theta) \leq M$ for all $\theta \in \Theta$ for some $M > 0$; (iii) for any $\epsilon, \eta > 0$, there exists $N \in \mathbb{N}$ such that*

$$\sup_{P\in\mathcal{P}} \sup_{\theta\in\Upsilon(P)} P\Big( \sup_{\theta'\in(\theta+n^{-1/2}\rho B_d)\cap\Theta} |c_n(\theta') - c_n(\theta)| > \epsilon\Big) < \eta, \quad \forall n \geq N; \tag{B.2}$$

*(iv) for any $\epsilon > 0$, there exists $N' \in \mathbb{N}$ such that $\inf_{P\in\mathcal{P}} P(c_n(\theta) \geq c_{h,n}^\rho(\theta), \forall\theta \in \Theta) > 1 - \epsilon$ for all $n \geq N'$.*

Lemma B.3: *Suppose Assumptions 3.1-3.4 hold. Let $\Upsilon(P) = H(p, \Theta_I(P))$ if $h_n^\Delta > \kappa_n^{1/5}$, and $\Upsilon(P) = \Theta_I(P)$ if $h_n^\Delta \leq \kappa_n^{1/5}$. Then, (i) for any $\eta$, there exist $\delta > 0$ and $N \in \mathbb{N}$ such that*

$$\sup_{P\in\mathcal{P}} \sup_{\theta\in\Upsilon(P)} P(\{W_n(\theta) \neq \emptyset\} \cap \{(W_n^{-\delta}(\theta))^o = \emptyset\}) < \eta,\ \forall n \geq N \tag{B.3}$$

*(ii) Furthermore, for any $\eta > 0$, there exists a $N \in \mathbb{N}$ such that*

$$\sup_{P\in\mathcal{P}} \sup_{\theta\in\Upsilon(P)} P(\{W_n(\theta) \neq \emptyset\} \cap \{V_n(\theta) = \emptyset\}) < \eta,\ \forall n \geq N. \tag{B.4}$$

Lemma B.4: *Suppose Assumptions 3.1-3.4 hold. Let $\Upsilon(P) = H(p, \Theta_I(P))$ if $h_n^\Delta > \kappa_n^{1/5}$, and $\Upsilon(P) = \Theta_I(P)$ if $h_n^\Delta \leq \kappa_n^{1/5}$. Then, (i) for any $\eta$, there exist $\delta > 0$ and $N \in \mathbb{N}$ such that*

$$\sup_{P\in\mathcal{P}} \sup_{\theta\in\Upsilon(P)} P(\{W_n(\theta) \neq \emptyset\} \cap \{V_n(\theta) \neq \emptyset\} \cap \{(V_n^{-\delta}(\theta))^o = \emptyset\}) < \eta,\ \forall n \geq N. \tag{B.5}$$

*(ii) Furthermore, for any $\eta > 0$, there exists a $N \in \mathbb{N}$ such that*

$$\sup_{P\in\mathcal{P}} \sup_{\theta\in\Upsilon(P)} P(\{W_n(\theta) \neq \emptyset\} \cap \{V_n(\theta) \neq \emptyset\} \cap \{U_n(\theta) = \emptyset\}) < \eta,\ \forall n \geq N. \tag{B.6}$$

Lemma B.5: *Suppose Assumptions 3.1-3.4 hold. Let $\Upsilon(P) = H(p, \Theta_I(P))$ if $h_n^\Delta > \kappa_n^{1/5}$, and $\Upsilon(P) = \Theta_I(P)$ if $h_n^\Delta \leq \kappa_n^{1/5}$. Let $c_n : \Theta \to \mathbb{R}_+$ be such that for all $P \in \mathcal{P}$, $c_n(\theta) \leq M$ for all $\theta \in \Theta$ for some $M > 0$, and for any $\epsilon, \eta > 0$, there exists $N \in \mathbb{N}$ such that*

$$\sup_{P\in\mathcal{P}} \sup_{\theta\in\Upsilon(P)} P^*\Big( \sup_{\theta'\in(\theta+n^{-1/2}\rho B_d)\cap\Theta} |c_n(\theta) - c_n(\theta')| > \epsilon\Big) < \eta, \quad \forall n \geq N. \tag{B.7}$$

*Then, for any $\epsilon, \eta > 0$, there exists $N \in \mathbb{N}$ such that*

$$\sup_{P\in\mathcal{P}} \sup_{\theta\in\Upsilon(P)} P^*\Big(\{|s(p, U_n(\theta)) - s(p, V_n(\theta))| > \epsilon\} \cap \{W_n(\theta) \neq \emptyset \cap V_n(\theta) \neq \emptyset \cap U_n(\theta) \neq \emptyset\}\Big) < \eta, \quad \forall n \geq N. \tag{B.8}$$

LEMMA B.6: *Suppose Assumptions 3.1-3.4 hold. Let* $\Upsilon(P) = H(p, \Theta_I(P))$ *if* $h_n^\Delta > \kappa_n^{1/5}$*, and* $\Upsilon(P) = \Theta_I(P)$ *if* $h_n^\Delta \leq \kappa_n^{1/5}$*. Let* $c_n : \Theta \to \mathbb{R}_+$ *be such that for all* $P \in \mathcal{P}$*,* $c_n(\theta) \leq M$ *for all* $\theta \in \Theta$ *for some* $M > 0$*, and for any* $\epsilon, \eta > 0$*, there exists* $N \in \mathbb{N}$ *such that*

$$\sup_{P\in\mathcal{P}} \sup_{\theta\in\Upsilon(P)} P^*( \sup_{\theta'\in(\theta+n^{-1/2}\rho B_d)\cap\Theta} |c_n(\theta) - c_n(\theta')| > \epsilon) < \eta, \quad \forall n \geq N. \tag{B.9}$$

*Then, for any* $\epsilon, \eta > 0$*, there exists* $N' \in \mathbb{N}$ *such that*

$$\sup_{P\in\mathcal{P}} \sup_{\theta\in\Upsilon(P)} P^*( \sup_{\lambda\in\rho B_d\cap\sqrt{n}(\Theta-\theta)} | \max_{j=1,\cdots,j} u_{j,n,\theta}(\lambda) - \max_{j=1,\cdots,j} v_{j,n,\theta}(\lambda)| \geq \epsilon) < \eta, \ \forall n \geq N'. \tag{B.10}$$

LEMMA B.7: *Let* $c_n$ *be defined as in* (A.15)*. Suppose that Assumptions 3.1-3.4 hold and that for all* $P \in \mathcal{P}$*,* $c_n(\theta) \leq M$ *for all* $\theta \in \Theta$ *for some* $M > 0$*. Let* $\Upsilon(P) = H(p, \Theta_I(P))$ *if* $h_n^\Delta > \kappa_n^{1/5}$*, and* $\Upsilon(P) = \Theta_I(P)$ *if* $h_n^\Delta \leq \kappa_n^{1/5}$*. Then, there exists a* $\underline{c} > 0$ *such that* $\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Upsilon(P)} c_n(\theta) \geq \underline{c}$*.*

LEMMA B.8: *For each* $n$ *and for each* $\theta \in \Theta$*,* $\hat{c}_n(\theta) \geq \hat{c}_n^*(\theta)$*,* $P-a.s.$ *for any* $P \in \mathcal{P}$*.*

LEMMA B.9: *Suppose Assumptions 3.1-3.4 hold. Let* $\Upsilon(P) = \partial\Theta_I(P)$ *if* $h_n^\Delta > \kappa_n^{1/5}$*, and* $\Upsilon(P) = \Theta_I(P)$ *if* $h_n^\Delta \leq \kappa_n^{1/5}$*. Then, for any* $\epsilon > 0$*,*

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \sup_{\theta\in\Upsilon(P)} P\left(|\hat{c}_n^*(\theta) - c_n(\theta)| > \epsilon\right) = 0. \tag{B.11}$$

LEMMA B.10: *Let* $c_n$ *be defined as in* (A.15)*. Let* $\Upsilon(P) = \partial\Theta_I(P)$ *if* $h_n^\Delta > \kappa_n^{1/5}$*, and* $\Upsilon(P) = \Theta_I(P)$ *if* $h_n^\Delta \leq \kappa_n^{1/5}$*. Then, (i) For all* $P \in \mathcal{P}$*,* $c_n(\theta) \leq M$ *for all* $\theta \in \Theta$ *for some* $M > 0$*; (ii) For any* $\epsilon, \eta > 0$*, there exists* $N \in \mathbb{N}$ *such that*

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \sup_{\theta\in\Upsilon(P)} P\Big( \sup_{\theta'\in(\theta+n^{-1/2}\rho B_d)\cap\Theta} |c_n(\theta) - c_n(\theta')| > \epsilon\Big) < \eta, \quad \forall n \geq N. \tag{B.12}$$

LEMMA B.11: *Suppose Assumptions 3.1-3.4 hold. Let* $\Upsilon(P) = H(p, \Theta_I(P))$ *if* $h_n^\Delta > \kappa_n^{1/5}$*, and* $\Upsilon(P) = \Theta_I(P)$ *if* $h_n^\Delta \leq \kappa_n^{1/5}$*. Consider sequences such that for* $\theta_n \in \Upsilon(P_n)$*, for each* $j = 1, \ldots, J_1$*,* $\kappa_n^{-1} h_{P,j,n}(\theta_n) \to \pi_j \in [-\infty, 0]$*. Define* $\mathcal{J}^* \equiv \{j = 1, \cdots, J : \pi_j \in (-\infty, 0]\}$*. Then* $\mathcal{J}^*$ *is non-empty.*

## B.1 Proofs of Lemmas

Proof of Lemma B.1: Consider first the case $h_n^{\Delta} \leq \kappa_n^{1/5}$. Note that for any $n$, the following relations hold:

$$
\begin{aligned}
& p'\theta_n \in CI_n \\
\Leftrightarrow & - s(-p, \mathcal{C}_n(\hat{c}_n)) \leq p'\theta_n \leq s(p, \mathcal{C}_n(\hat{c}_n)) \\
\Leftarrow & \left\{ \begin{array}{l} \sup -p'\vartheta \\ s.t.\, \vartheta \in \Theta, \quad \frac{\sqrt{n}\bar{m}_{n,j}(\vartheta)}{\hat{\sigma}_{n,j}(\vartheta)} \leq c_n(\vartheta),\, j = 1, \dots, J \end{array} \right\} \geq -p'\theta_n \\
\cap & \left\{ \begin{array}{l} \sup p'\vartheta \\ s.t.\, \vartheta \in \Theta, \quad \frac{\sqrt{n}\bar{m}_{n,j}(\vartheta)}{\hat{\sigma}_{n,j}(\vartheta)} \leq c_n(\vartheta),\, j = 1, \dots, J \end{array} \right\} \geq p'\theta_n \\
\Leftrightarrow & \left\{ \begin{array}{c} \sup_{\lambda \in \sqrt{n}(\Theta - \theta_n)} -p'\lambda \\ s.t.\, \frac{\sqrt{n}\ \bar{m}_{n,j}(\theta_n + \lambda/\sqrt{n})}{\hat{\sigma}_{n,j}(\theta_n + \lambda/\sqrt{n})} \leq c_n(\theta_n + \lambda/\sqrt{n}), \quad j = 1, \dots, J \end{array} \right\} \geq 0 \\
\cap & \left\{ \begin{array}{c} \sup_{\lambda \in \sqrt{n}(\Theta - \theta_n)} p'\lambda \\ s.t.\, \frac{\sqrt{n}\ \bar{m}_{n,j}(\theta_n + \lambda/\sqrt{n})}{\hat{\sigma}_{n,j}(\theta_n + \lambda/\sqrt{n})} \leq c_n(\theta_n + \lambda/\sqrt{n}), \quad j = 1, \dots, J \end{array} \right\} \geq 0, \\
\Leftarrow & \left\{ \begin{array}{c} \sup_{\lambda \in \sqrt{n}(\Theta - \theta_n) \cap \rho B_d} -p'\lambda \\ s.t.\, \frac{\sqrt{n}\ \bar{m}_{n,j}(\theta_n + \lambda/\sqrt{n})}{\hat{\sigma}_{n,j}(\theta_n + \lambda/\sqrt{n})} \leq c_n(\theta_n + \lambda/\sqrt{n}), \quad j = 1, \dots, J \end{array} \right\} \geq 0 \\
\cap & \left\{ \begin{array}{c} \sup_{\lambda \in \sqrt{n}(\Theta - \theta_n) \cap \rho B_d} p'\lambda \\ s.t.\, \frac{\sqrt{n}\ \bar{m}_{n,j}(\theta_n + \lambda/\sqrt{n})}{\hat{\sigma}_{n,j}(\theta_n + \lambda/\sqrt{n})} \leq c_n(\theta_n + \lambda/\sqrt{n}), \quad j = 1, \dots, J \end{array} \right\} \geq 0,
\end{aligned}
\tag{B.13}
$$

where the one sided implication in the third row follows from Lemma B.2-(i), and the last one sided implication follows from the addition of the $\rho$-box constraints. In (B.13), $\lambda$ is constrained to be in $\sqrt{n}(\Theta - \theta)$. By Assumption 3.3 (iv) and $\rho/\sqrt{n} \to 0$, for any sequence $\theta_n \in \Theta$,

$$\rho B_d \cap \sqrt{n}(\Theta - \theta_n) = \rho B_d\ , \tag{B.14}$$

for all $n$ sufficiently large. Hence, in what follows, we make the requirement $\lambda \in \sqrt{n}(\Theta - \theta)$ implicit whenever we consider $\rho B_d \cap \sqrt{n}(\Theta - \theta_n)$. It then follows, defining $\gamma_{l_n}, \theta_{l_n}$ as in the proof of Theorem 3.1, that

$$P_{\gamma_{l_n}}(p'\theta_{l_n} \in CI_n) \geq P_{\gamma_{l_n}}\Big(s(p, U_{l_n}(\theta_{l_n})) \geq 0 \geq -s(-p, U_{l_n}(\theta_{l_n}))\Big)$$

Consider now the case that $h_n^{\Delta} > \kappa_n^{1/5}$. Let again $(P_{\gamma_n}, \theta_n) \in \{(P,\theta) : P \in \mathcal{P},\ \theta \in \Theta_I\}$ be a sequence of distributions such that

$$\lim_{n\to\infty} P_{\gamma_n}(p'\theta_n \in CI_n) = \liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Theta_I(P)} P(p'\theta \in CI_n). \tag{B.15}$$

Parametrize the corresponding value of $p'\theta_n$ as

$$
\begin{aligned}
p'\theta_n &= -s(-p, \Theta_I(P_{\gamma_n})) + \varsigma_n \frac{h_n^{\Delta}}{\sqrt{n}} \\
&= s(p, \Theta_I(P_{\gamma_n})) - (1 - \varsigma_n)\frac{h_n^{\Delta}}{\sqrt{n}},
\end{aligned}
$$

where $\varsigma_n \in [0,1]$. To simplify notation, in what follows we omit the subscript $n$ from $\varsigma$, and we write $\Theta_I$ for $\Theta_I(P_{\gamma_n})$. Then we have

$$
\begin{aligned}
& P_{\gamma_n}(p'\theta_n \in CI_n) \\
=& P_{\gamma_n}(-s(-p, \mathcal{C}_n(\hat{c}_n)) \leq p'\theta_n \leq s(p, \mathcal{C}_n(\hat{c}_n))) \\
=& P_{\gamma_n}(-s(-p, \mathcal{C}_n(\hat{c}_n)) \leq -s(-p, \Theta_I) + \varsigma \frac{h_n^{\Delta}}{\sqrt{n}} \leq s(p, \mathcal{C}_n(\hat{c}_n))) \\
=& P_{\gamma_n}(-(s(-p, \mathcal{C}_n(\hat{c}_n)) - s(-p, \Theta_I)) - \varsigma \frac{h_n^{\Delta}}{\sqrt{n}} \leq 0 \leq s(p, \mathcal{C}_n(\hat{c}_n)) - s(p, \Theta_I) + (1-\varsigma)\frac{h_n^{\Delta}}{\sqrt{n}})
\end{aligned}
\tag{B.16}
$$

The event inside equation (B.16) can be expressed as follows

$$
\begin{aligned}
&\Big\{s(-p,\mathcal{C}_n(\hat{c}_n)) \geq s(-p,\Theta_I(P))) - \varsigma\frac{h_n^\Delta}{\sqrt{n}} \cap s(p,\mathcal{C}_n(\hat{c}_n)) \geq s(p,\Theta_I(P)) - (1-\varsigma)\frac{h_n^\Delta}{\sqrt{n}}\Big\}\\
&\Leftarrow \left\{\begin{array}{l}\sup -p'\vartheta\\ s.t.\, \vartheta\in\Theta, \quad \frac{\sqrt{n}\bar{m}_{n,j}(\vartheta)}{\hat{\sigma}_{n,j}(\vartheta)} \leq c_n(\vartheta),\ j=1,\dots,J\end{array}\right\} \geq s(-p,\Theta_I) - \varsigma\frac{h_n^\Delta}{\sqrt{n}}\\
&\cap \left\{\begin{array}{l}\sup p'\vartheta\\ s.t.\, \vartheta\in\Theta, \quad \frac{\sqrt{n}\bar{m}_{n,j}(\vartheta)}{\hat{\sigma}_{n,j}(\vartheta)} \leq c_n(\vartheta),\ j=1,\dots,J\end{array}\right\} \geq s(p,\Theta_I) - (1-\varsigma)\frac{h_n^\Delta}{\sqrt{n}}\\
&\Leftrightarrow \left\{\begin{array}{l}\sup_{(\theta,\lambda)} -p'\lambda\\ s.t.\, \theta\in H(-p,\Theta_I), \quad \frac{\sqrt{n}\ \bar{m}_{n,j}(\theta+\lambda/\sqrt{n})}{\hat{\sigma}_{n,j}(\theta+\lambda/\sqrt{n})} \leq c_n(\theta+\lambda/\sqrt{n}),\ j=1,\dots,J\end{array}\right\} \geq -\varsigma h_n^\Delta\\
&\cap \left\{\begin{array}{l}\sup_{(\theta,\lambda)} p'\lambda\\ s.t.\, \theta\in H(p,\Theta_I), \quad \frac{\sqrt{n}\ \bar{m}_{n,j}(\theta+\lambda/\sqrt{n})}{\hat{\sigma}_{n,j}(\theta+\lambda/\sqrt{n})} \leq c_n(\theta+\lambda/\sqrt{n}),\ j=1,\dots,J\end{array}\right\} \geq -(1-\varsigma)h_n^\Delta, \qquad \text{(B.17)}\\
&\Leftarrow \left\{\begin{array}{l}\sup_{\lambda\in\rho B_d} -p'\lambda\\ \quad s.t. \quad \frac{\sqrt{n}\ \bar{m}_{n,j}(\theta_n^1+\lambda/\sqrt{n})}{\hat{\sigma}_{n,j}(\theta_n^1+\lambda/\sqrt{n})} \leq c_n(\theta_n^1+\lambda/\sqrt{n}),\ j=1,\dots,J\end{array}\right\} \geq -\varsigma h_n^\Delta\\
&\cap \left\{\begin{array}{l}\sup_{\lambda\in\rho B_d} p'\lambda\\ \quad s.t. \quad \frac{\sqrt{n}\ \bar{m}_{n,j}(\theta_n^2+\lambda/\sqrt{n})}{\hat{\sigma}_{n,j}(\theta_n^2+\lambda/\sqrt{n})} \leq c_n(\theta_n^2+\lambda/\sqrt{n}),\ j=1,\dots,J\end{array}\right\} \geq -(1-\varsigma)h_n^\Delta, \qquad \text{(B.18)}
\end{aligned}
$$

where the one sided implication in the second line follows because uniformly in $\theta$, $\hat{c}_n(\theta) \geq c_n(\theta)$ by Lemma B.2, and the last one sided implication follows from the addition of the $\rho$-box constraints, and picking sequences $\theta_n^1 \in H(-p,\Theta_I)$ and $\theta_n^2 \in H(p,\Theta_I)$. Recall that $h_n^\Delta > \kappa_n^{1/5}$, and that the value of each program in B.18 is bounded in direction $p$ by $\sqrt{d}\rho$, and in direction $-p$ by $-\sqrt{d}\rho$. Because $h_n^\Delta$ is diverging to infinity, we have that the above probability is minimized for $\varsigma \in \{0,1\}$, yielding the claim.

□

Proof of Lemma B.2: By Assumptions 3.1-3.4, (i) follows from Lemmas B.8 and B.9. (ii) and (iii) then follow from Lemma B.10. (iv) follows from (A.10)-(A.11), (A.14)-(A.15), and $h_{P,j,n} \leq \tilde{\zeta}_{P,j,n}$ for all $j$. □

Proof of Lemma B.3: We first show (B.3). Let $\theta \in \Upsilon(P)$ be given. By Theorem 22.1 in Rockafellar (1970), a solution to the system of linear inequalities defining $W_n(\theta)$ exists if and only if for all $\mu \in \mathbb{R}_+^{J+2d}$ such that $\mu' K_P(\theta) = 0$, one has $\mu' g_P(\theta) \geq 0$, with $g_P(\theta)$ the $J+2d \times 1$ vector with components

$$g_{P,j}(\theta) \equiv c_n(\theta) - \mathbb{G}_{P,j}(\theta) - h_{P,j,n}(\theta),\ j=1,\dots,J \qquad \text{(B.19)}$$

$$g_{P,j}(\theta) \equiv \rho,\ j=J+1,\dots,J+2d \qquad \text{(B.20)}$$

where $h_{P,j,n}(\theta) = 0$ for $j = J_1+1,\cdots,J$. By Theorem 22.2 in Rockafellar (1970), a solution to the system of strict linear inequalities (which in turn define $(W_n^{-\delta}(\theta))^o$)

$$\mathbb{G}_{P,j}(\theta) + D_{P,j}(\theta)'\lambda + h_{P,j,n}(\theta) < c_n(\theta) - \delta,\ j=1,\dots,J, \qquad \text{(B.21)}$$

$$\lambda_j < \rho - \delta,\ j=J+1,\dots,J+d, \qquad \text{(B.22)}$$

$$-\lambda_j < \rho - \delta,\ j=J+d+1,\dots,J+2d, \qquad \text{(B.23)}$$

exists if and only if for all $\mu \in \mathbb{R}_+^{J+2d}$ such that $\mu \neq 0$ and $\mu' K_P(\theta) = 0$, one has $\mu'(g_P(\theta) - \delta 1_{J+2d}) > 0$, where $1_{J+2d}$ is a $J+2d$-dimensional vector of ones. Define

$$\mathcal{M}(\theta) \equiv \{\mu \in \mathbb{R}_+^{J+2d} : \mu' K_P(\theta) = 0\},\quad \tilde{\mathcal{M}}(\theta) \equiv \{\mu \in \mathbb{R}_+^{J+2d} : \mu \neq 0, \mu' K_P(\theta) = 0\}. \qquad \text{(B.24)}$$

Then, one may write

$$
\begin{aligned}
P(W_n(\theta) \neq \emptyset \cap (W_n^{-\delta}(\theta))^o = \emptyset) &= P(\{\mu' g_P(\theta) \geq 0, \forall \mu \in \mathcal{M}(\theta)\} \cap \{\mu'(g_P(\theta) - \delta 1_{J+2d}) > 0, \forall \mu \in \tilde{\mathcal{M}}(\theta)\}^c) \\
&= P(\{\mu' g_P(\theta) \geq 0, \forall \mu \in \mathcal{M}(\theta)\} \cap \{\mu'(g_P(\theta) - \delta 1_{J+2d}) \leq 0, \exists \mu \in \tilde{\mathcal{M}}(\theta)\}) \\
&= P(\{\mu' g_P(\theta) \geq 0, \forall \mu \in \mathcal{M}(\theta)\} \cap \{\mu' g_P(\theta) \leq \delta \mu' 1_{J+2d}, \exists \mu \in \tilde{\mathcal{M}}(\theta)\}).
\end{aligned} \tag{B.25}
$$

Note that the set $\tilde{\mathcal{M}}(\theta)$ is a (possibly unbounded) non-stochastic convex polyhedron. Hence, by Carathéodory's theorem, there exist $\{\nu^t \in \tilde{\mathcal{M}}(\theta), t = 1, \cdots, T\}$ with $T \leq J+2d+1$ such that any $\mu \in \tilde{\mathcal{M}}(\theta)$ can be represented as

$$
\mu = \sum_{t=1}^{T} a_t \nu_t, \tag{B.26}
$$

where $a_t \geq 0$ and $\sum_{t=1}^{T} a_t = 1$. Hence, if $\mu \in \tilde{\mathcal{M}}(\theta)$ satisfies $\mu' g_P(\theta) \leq \delta \mu' 1_{J+2d}$, we have

$$
\sum_{t=1}^{T} a_t \nu_t' g_P(\theta) \leq \delta \sum_{t=1}^{T} a_t \nu_t' 1_{J+2d}. \tag{B.27}
$$

However, due to $a_t \geq 0, \forall t$, $\sum_{t=1}^{T} a_t = 1$, $\mu \neq 0$, and $\nu^t \in \mathcal{M}(\theta)$, this means $\nu^{t\prime} g_P(\theta) \leq \delta \nu^{t\prime} 1_{J+2d}$ for some $t \in \{1, \ldots, T\}$. Furthermore, since $\nu^t \in \tilde{\mathcal{M}}(\theta) \subset \mathcal{M}(\theta)$, we have $0 \leq \nu^{t\prime} g_P(\theta)$. Therefore,

$$
\begin{aligned}
&P(\{\mu' g_P(\theta) \geq 0, \forall \mu \in \mathcal{M}(\theta)\} \cap \{\mu' g_P(\theta) \leq \delta \mu 1_{J+2d}, \exists \mu \in \tilde{\mathcal{M}}(\theta)\}) \\
&\qquad \leq P(0 \leq \nu_{t\prime} g_P(\theta) \leq \delta \nu_{t\prime} 1_{J+2d}, \exists t \in \{1, \cdots, T\}) \leq \sum_{t=1}^{T} P(0 \leq \frac{\nu^{t\prime} g_P(\theta)}{\nu^{t\prime} 1_{J+2d}} \leq \delta).
\end{aligned} \tag{B.28}
$$

Suppose for the moment that there are no moment equalities and $J = J_1$. Consider first the case that $\nu^t$ assigns positive weight to at least constraints in $\{J+1, \ldots, J+2d\}$. By choosing $\delta < \rho$ we obtain $\frac{\nu^{t\prime} g_P(\theta)}{\nu^{t\prime} 1_{J+2d}} > \delta$.

Consider now the case that $\nu^t$ assigns positive weight also to constraints in $\{1, \ldots, J\}$. By (B.19), for any given $\nu^t \in \tilde{\mathcal{M}}(\theta)$, $\frac{\nu^{t\prime} g_P(\theta)}{\nu^{t\prime} 1_{J+2d}}$ is then a normal random variable with variance $(\nu^{t\prime} 1_{J+2d})^{-2} \mu_k' \Omega_P(\theta) \nu^t$. By Assumption 3.2 (iv), there exists a constant $\omega > 0$ that does not depend on $\theta$ such that the smallest eigenvalue of $\Omega_P(\theta)$ is bounded from below by $\omega$ for all $\theta$. Hence, letting $\|\cdot\|_p$ denote the $p$-norm in $\mathbb{R}^{J+2d}$, we have

$$
\frac{\mu_k' \Omega_P(\theta) \nu^t}{\|\nu^t\|_1^2} \geq \frac{\omega \|\nu^t\|_2^2}{(J+2d)\|\nu^t\|_2^2} \geq \frac{\omega}{J+2d}. \tag{B.29}
$$

Therefore, the variance of the normal random variable in (B.28) is uniformly bounded away from 0, which in turn allows one to find $\delta > 0$ such that $P(0 \leq \frac{\nu^{t\prime} g_P(\theta)}{\nu^{t\prime} 1_{J+2d}} \leq \delta) \leq \eta/K$. This ensures (B.3) when no moment equalities are present.

For the case where moment equalities are present, the same conclusion holds. If $\nu^t$ assigns positive weight only to constraints in $\{J+1, \ldots, J+2d\}$ the same conclusion as before holds. We therefore consider the case that $\nu^t$ assigns positive weight to at least one constraints in $\{1, \ldots, J\}$; to further simplify notation, we assume that $\nu^t$ assigns zero weight to constraints in $\{J+1, \ldots, J+2d\}$. Recall that the moment equalities have indexes $j = J_1+1, \ldots, J_1+J_2$ and are each written as two moment inequalities, therefore yielding a total of $2J_2$ inequalities with:

$$
\begin{array}{lcll}
D_{P,j+J_2}(\theta) & = & -D_{P,j}(\theta) & \text{for } j = J_1+1, \ldots, J_1+J_2, \\
g_{P,j}(\theta) & = & c_n(\theta) - \mathbb{G}_{P,j}(\theta) & \text{for } j = J_1+1, \ldots, J_1+J_2, \\
g_{P,j+J_2}(\theta) & = & c_n(\theta) + \mathbb{G}_{P,j}(\theta) & \text{for } j = J_1+1, \ldots, J_1+J_2.
\end{array} \tag{B.30}
$$

For any $\mu \in \tilde{\mathcal{M}}(\theta)$, (B.30) implies

$$
\sum_{j=J_1+1}^{J_1+2J_2} \mu_j g_{P,j}(\theta) = c_n(\theta) \sum_{j=J_1+1}^{J_1+J_2} (\mu_j + \mu_{j+J_2}) + \sum_{j=J_1+1}^{J_1+J_2} (\mu_j - \mu_{j+J_2}) \mathbb{G}_{P,j}(\theta). \tag{B.31}
$$

For each $j = 1, \cdots, J_1 + J_2$, define

$$\tilde{\nu}_j^t \equiv \begin{cases} \nu_j^t & j = 1, \cdots, J_1 \\ \nu_j^t - \nu_{j+J_2}^t & j = J_1 + 1, \cdots, J_1 + J_2. \end{cases} \tag{B.32}$$

We then let $\tilde{\nu}^t \equiv (\tilde{\nu}_1^t, \cdots, \tilde{\nu}_{J_1+J_2}^t)'$. Then, one may write

$$\nu^{t\prime} g_P(\theta) = \sum_{j=1}^{J_1+J_2} \tilde{\nu}_j^t \mathbb{G}_{P,j}(\theta) + c_n(\theta) \sum_{j=1}^{J} \nu_j^t + \sum_{j=1}^{J_1} \nu_j^t h_{P,j,n}(\theta). \tag{B.33}$$

Suppose $\tilde{\nu}^t \neq 0$. Then, by (B.33), Assumption 3.2 (iv) and arguing as in the case without moment equalities, there exists $\delta > 0$, $P(0 \leq \frac{\nu^{t\prime} g_P(\theta)}{\nu^{t\prime} 1_{J+2d}} \leq \delta) \leq \eta/K$. Now, consider the case $\tilde{\nu}^t = 0$. Since $\nu^t \in \tilde{\mathcal{M}}(\theta)$, this occurs only if $\nu_j^t = 0$ for all $j = 1, \cdots, J_1$ and $\nu_j^t = \mu_{j+J_2}$ for all $j = J_1 + 1, \cdots, J_1 + J_2$, while $\nu_j^t > 0$ for some $j = J_1 + 1, \cdots, J_1 + J_2$. Then, by (B.33) and Lemma B.7, we have

$$\frac{\nu^{t\prime} g_P(\theta)}{\nu^{t\prime} 1_{J+2d}} = \frac{c_n(\theta) \sum_{j=1}^{J} \nu_j^t}{\sum_{j=1}^{J} \nu_j^t} = c_n(\theta) \geq \bar{c} > 0. \tag{B.34}$$

Hence, letting $\delta < \min\{\bar{c}, \rho\}$, we have $P(0 \leq \frac{\nu^{t\prime} g_P(\theta)}{\nu^{t\prime} 1_{J+2d}} \leq \delta) = 0 \leq \eta/K$. Thus, by (B.28), it follows that $P(W_n(\theta) \neq \emptyset \cap (W_n^{-\delta}(\theta))^o = \emptyset) < \eta$ uniformly on $\{(\theta, P) : \theta \in \Upsilon(P), P \in \mathcal{P}\}$. This therefore establishes (B.3).

We now show (B.4). Suppose that $(W_n^{-\delta}(\theta))^o \neq \emptyset$ for $\delta > 0$. By Theorem 2.8.2 in van der Vaart and Wellner (2000), $\mathbb{G}_{P,n}$ weakly converges to $\mathbb{G}_P$ uniformly in $P \in \mathcal{P}$. We take a Skorkhod representations of $\mathbb{G}_P$ and $\mathbb{G}_{P,n}$ and denote them by $\mathbb{G}_P^*$ and $\mathbb{G}_{P,n}^*$. Then, $\mathbb{G}_{P,n}^* \to \mathbb{G}_P^*$ with probability one. Since the constraints are linear, there exists $N \in \mathbb{N}$ that does not depend on $\theta$ such that:

$$\begin{aligned} &\sup_{\lambda \in \rho B_d} |\mathbb{G}_{P,j}^*(\theta) + D_{P,j}(\theta)'\lambda + h_{P,j,n}(\theta) + c_n(\theta) - \mathbb{G}_{P,j,n}^*(\theta) - D_{P,j}(\theta)'\lambda - h_{P,j,n}(\theta) - c_n(\theta)| \\ &\qquad = |\mathbb{G}_{P,j}^*(\theta) - \mathbb{G}_{P,j,n}^*(\theta)| < \delta, \ \forall n \geq N, \ a.s. \end{aligned} \tag{B.35}$$

Therefore, for any $\theta \in (W_n^{-\delta}(\theta))^o$, we have

$$c_n(\theta) > \delta + \mathbb{G}_{P,j}^*(\theta) + D_{P,j}(\theta) + h_{P,j,n}(\theta) \geq \mathbb{G}_{P,j,n}^*(\theta) + D_{P,j}(\theta) + h_{P,j,n}(\theta), \tag{B.36}$$

and hence $(W_n^{-\delta}(\theta))^o \subset V_n(\theta)$ for all $n \geq N$ with probability 1. This leads us to conclude that for any $\eta > 0$,

$$\sup_{P \in \mathcal{P}} \sup_{\theta \in \Upsilon(P)} P((W_n^{-\delta}(\theta))^o \neq \emptyset \cap V_n(\theta) = \emptyset) < \eta/2, \ \forall n \geq N. \tag{B.37}$$

By (B.3) and the triangle inequality, it then follows that

$$\begin{aligned} &\sup_{P \in \mathcal{P}} \sup_{\theta \in \Upsilon(P)} P(W_n(\theta) \neq \emptyset \cap V_n(\theta) = \emptyset) \\ &\qquad \leq \sup_{P \in \mathcal{P}} \sup_{\theta \in \Upsilon(P)} \left\{P(W_n(\theta) \neq \emptyset \cap (W_n^{-\delta}(\theta))^o = \emptyset) + P((W_n^{-\delta}(\theta))^o \neq \emptyset \cap V_n\theta) = \emptyset)\right\} < \eta. \end{aligned} \tag{B.38}$$

for all $n$ sufficiently large. This establishes (B.4). □

Proof of Lemma B.4: For any $\delta, \eta > 0$, there exists $N \in \mathbb{N}$ such that

$$\begin{aligned} &\sup_{P \in \mathcal{P}} \sup_{\theta \in \Upsilon(P)} P(\{W_n(\theta) \neq \emptyset\} \cap \{V_n(\theta) \neq \emptyset\} \cap \{(V_n^{-\delta}(\theta))^o = \emptyset\}) \\ &\quad \leq \sup_{P \in \mathcal{P}} \sup_{\theta \in \Upsilon(P)} P(\{W_n(\theta) \neq \emptyset\} \cap \{(W_n^{-\delta/2}(\theta))^o = \emptyset\} \cap \{V_n(\theta) \neq \emptyset\} \cap \{(V_n^{-\delta}(\theta))^o = \emptyset\}) \\ &\quad + \sup_{P \in \mathcal{P}} \sup_{\theta \in \Upsilon(P)} P(\{W_n(\theta) \neq \emptyset\} \cap \{(W_n^{-\delta/2}(\theta))^o \neq \emptyset\} \cap \{V_n(\theta) \neq \emptyset\} \cap \{(V_n^{-\delta}(\theta))^o = \emptyset\}) \\ &\quad \leq \sup_{P \in \mathcal{P}} \sup_{\theta \in \Upsilon(P)} P(\{W_n(\theta) \neq \emptyset\} \cap \{(W_n^{-\delta/2}(\theta))^o \neq \emptyset\} \cap \{V_n(\theta) \neq \emptyset\} \cap \{(V_n^{-\delta}(\theta))^o = \emptyset\}) + \eta/2, \end{aligned} \tag{B.39}$$

for all $n \geq N$, where the last inequality follows from Lemma B.3 (i) and

$$P(\{W_n(\theta) \neq \emptyset\} \cap \{(W_n^{-\delta/2}(\theta))^o = \emptyset\} \cap \{V_n(\theta) \neq \emptyset\} \cap \{(V_n^{-\delta}(\theta))^o = \emptyset\})$$
$$\leq P(\{W_n(\theta) \neq \emptyset\} \cap \{(W_n^{-\delta/2}(\theta))^o = \emptyset\}). \quad \text{(B.40)}$$

Furthermore,

$$P(\{W_n(\theta) \neq \emptyset\} \cap \{(W_n^{-\delta/2}(\theta))^o \neq \emptyset\} \cap \{V_n(\theta) \neq \emptyset\} \cap \{(V_n^{-\delta}(\theta))^o = \emptyset\}$$
$$\leq P(\{(W_n^{-\delta/2}(\theta))^o \neq \emptyset\} \cap \{(V_n^{-\delta}(\theta))^o = \emptyset\}) = \mathbf{P}(\{(W_n^{*,-\delta/2}(\theta))^o \neq \emptyset\} \cap \{(V_n^{*,-\delta}(\theta))^o = \emptyset\}), \quad \text{(B.41)}$$

where $W_n^*$ and $V_n^*$ replace $\mathbb{G}_P$ and $\mathbb{G}_{P,n}$ in (A.25)-(A.24) by their Skorokhod representations $\mathbb{G}_P^*$ and $\mathbb{G}_{P,n}^*$. Arguing as in (B.37), it follows that there exists $N' \in \mathbb{N}$ that does not depend on $\theta$ such that $(V_n^{*,-\delta}(\theta))^o \subset (W_n^{-\delta/2}(\theta))^o, \forall n \geq N', \mathbf{P} - a.s.$ Hence,

$$\mathbf{P}(\{(W_n^{-\delta/2}(\theta))^o \neq \emptyset\} \cap \{(V_n^{-\delta}(\theta))^o = \emptyset\}) < \eta/2, \ \forall n \geq N'. \quad \text{(B.42)}$$

The first conclusion of the lemma then follows from (B.39)-(B.42).

For the second claim, note that for any $\delta, \eta > 0$, there exists $N \in \mathbb{N}$ such that

$$\sup_{P \in \mathcal{P}} \sup_{\theta \in \Upsilon(P)} P(\{W_n(\theta) \neq \emptyset\} \cap \{V_n(\theta) \neq \emptyset\} \cap \{U_n(\theta) = \emptyset\})$$
$$\leq \sup_{P \in \mathcal{P}} \sup_{\theta \in \Upsilon(P)} P(\{W_n(\theta) \neq \emptyset\} \cap \{V_n(\theta) \neq \emptyset\} \cap \{(V_n^{-\delta}(\theta))^o \neq \emptyset\} \cap \{U_n(\theta) = \emptyset\})$$
$$+ \sup_{P \in \mathcal{P}} \sup_{\theta \in \Upsilon(P)} P(\{W_n(\theta) \neq \emptyset\} \cap \{V_n(\theta) \neq \emptyset\} \cap \{(V_n^{-\delta}(\theta))^o = \emptyset\} \cap \{U_n(\theta) = \emptyset\})$$
$$\leq \sup_{P \in \mathcal{P}} \sup_{\theta \in \Upsilon(P)} P(\{(V_n^{-\delta}(\theta))^o \neq \emptyset\} \cap \{U_n(\theta) = \emptyset\}) + \eta/2 \quad \text{(B.43)}$$

for all $n \geq N$, where the last inequality follows from the first claim of the lemma shown above. By (A.18)-(A.19) and (A.21)-(A.22), one may write

$$(V_n^{-\delta}(\theta))^o \equiv \{\lambda \in \rho_n B_d : v_{j,n,\theta}(\lambda) < -\delta, \ j = 1, \cdots, J\} \quad \text{(B.44)}$$
$$U_n(\theta) \equiv \big\{\lambda \in \rho B_d : u_{j,n,\theta}(\lambda) \leq 0, \ j = 1, \cdots, J\big\}. \quad \text{(B.45)}$$

Define the event

$$A_n(\theta, P) \equiv \{(V_n^{-\delta}(\theta))^o \subset U_n(\theta)\}. \quad \text{(B.46)}$$

Then, by Lemma B.6, for any $\eta > 0$ there exists a $N' \in \mathbb{N}$ such that

$$\inf_{P \in \mathcal{P}} \inf_{\theta \in \Upsilon(P)} P(A_n(\theta, P)) \geq 1 - \eta/2, \ \forall n \geq N'. \quad \text{(B.47)}$$

Note further that $\{(V_n^{-\delta}(\theta))^o \neq \emptyset\} \cap \{U_n(\theta) = \emptyset\} \cap A_n(\theta, P) = \emptyset$ by the definition of $A_n(\theta, P)$. Hence,

$$\sup_{P \in \mathcal{P}} \sup_{\theta \in \Upsilon(P)} P(\{(V_n^{-\delta}(\theta))^o \neq \emptyset\} \cap \{U_n(\theta) = \emptyset\})$$
$$\leq \sup_{P \in \mathcal{P}} \sup_{\theta \in \Upsilon(P)} \{P(\{V_n^{-\delta}(\theta))^o \neq \emptyset\} \cap \{U_n(\theta) = \emptyset\} \cap A_n(\theta, P)) + P(A_n(\theta, P)^c)\} \leq \eta/2, \ \forall n \geq N', \quad \text{(B.48)}$$

where the last inequality follows from (B.47). The second claim of the lemma the follows from (B.43) and (B.48). □

Proof of Lemma B.5: For each $\theta$, let

$$\zeta_n(\theta) \equiv \sup_{\lambda \in \rho B_d \cap \sqrt{n}(\Theta - \theta)} \Big| \max_{j=1,\cdots,J} \{u_{j,n,\theta}(\lambda)\} - \max_{j=1,\cdots,J} \{v_{j,n,\theta}(\lambda)\} \Big|. \quad \text{(B.49)}$$

By Lemma B.6, for any $\epsilon, \eta > 0$, there exists $N' \in \mathbb{N}$ such that

$$\sup_{P \in \mathcal{P}} \sup_{\theta \in \Upsilon(P)} P(\zeta_n(\theta) \geq \epsilon) < \eta, \ \forall n \geq N'. \tag{B.50}$$

Next, suppose that for some $\delta > 0$, $(W_n^{-\delta}(\theta))^o \neq \emptyset$ so that $(V_n(\theta))^o \neq \emptyset$. Then, we have

$$\begin{aligned} V_n(\theta) &= cl((V_n(\theta))^o) \\ &= cl(\{\lambda \in \rho B_d^o : \mathbb{G}_{P,j,n}(\theta) + D_{P,j}(\theta)'\lambda + h_{P,j,n}(\theta) < c(\theta), \ j = 1, \dots, J\}), \end{aligned} \tag{B.51}$$

since $V_n(\theta)$ and $(V_n(\theta))^o$ are polyhedral sets.

Define the function

$$\phi_n(\varepsilon) = |s(p, (V_n(\theta)) - s(p, \{\lambda \in \rho B_d : v_{j,n,\theta}(\lambda) \leq \varepsilon \|D_{P,j}(\theta)\|, j = 1, \dots, J\})|. \tag{B.52}$$

For the moment, suppose that $\{\lambda \in \rho B_d : v_{j,n,\theta}(\lambda) \leq \varepsilon \|D_{P,j}(\theta)\|, j = 1, \dots, J\} \neq \emptyset$. By Lemma C.1 (whose assumptions are verified in Appendix C.1 and C.2) and Theorem C.2, uniformly in $P \in \mathcal{P}$ with probability at least $1 - \eta/2$, there exists a $M_{\eta/2}$ and $N \in \mathbb{N}$ such that for all $n \geq N$,

$$\phi_n(\varepsilon) \leq (J + 2d) M_{\eta/2} |\varepsilon|. \tag{B.53}$$

Arguing similarly to the proof of Theorem 2.1 in Molchanov (1998), we have that

$$U_n(\theta) \cap \rho B_d \subseteq \left\{\lambda \in \rho B_d : v_{j,n,\theta}(\lambda) \leq \zeta_n(\theta) \frac{\|D_{P,j}(\theta)\|}{\underline{M}}, \ j = 1, \cdots, J\right\}, \tag{B.54}$$

because for each $j = 1, \cdots, J$, $\{u_{j,n,\theta}(\lambda)\} \leq 0$ implies $v_{j,n,\theta}(\lambda) - \zeta_n(\theta)\|D_{P,j}(\theta)\|/\underline{M} \leq 0$ by Assumption 3.3-(i), and therefore

$$s(p, U_n(\theta)) \leq s(p, V_n(\theta)) + \phi_n(\zeta_n(\theta)/\underline{M}). \tag{B.55}$$

If for some $\delta > 0$, $(V_n^{-\delta}(\theta))^o$ is non-empty, we have that for $n$ sufficiently large, $\{\lambda \in \rho B_d : v_{j,n,\theta}(\lambda) \leq -\zeta_n(\theta)\|D_{P,j}(\theta)\|/\underline{M}, \ j = 1, \cdots, J\}$ is non-empty. Hence,

$$\begin{aligned} s(p, V_n(\theta)) &\leq s(p, \{\lambda \in \rho B_d : v_{j,n,\theta}(\lambda) \leq -\zeta_n(\theta)\|D_{P,j}(\theta)\|/\underline{M}, \ j = 1, \cdots, J\}) + \phi_n(\zeta_n(\theta)/\underline{M}) \\ &\leq s(p, U_n(\theta)) + \phi_n(\zeta_n(\theta)/\underline{M}), \end{aligned} \tag{B.56}$$

where the first inequality follows from the definition of $\phi_n$, and the second inequality follows from the definition of $\zeta_n$, because for each $j = 1, \dots, J$, $v_{j,n,\theta}(\lambda) + \zeta_n(\theta)/\underline{M} \leq 0$ implies $u_{j,n,\theta}(\lambda) \leq 0$. Hence, we have that uniformly in $P \in \mathcal{P}$ with probability at least $1 - \eta/2$, there exists a $M_{\eta/2}$ and $N \in \mathbb{N}$ such that for all $n \geq N$,

$$|s(p, V_n(\theta)) - s(p, U_n(\theta))| \leq \phi_n(\zeta_n(\theta)/\underline{M}) \leq \frac{(J + 2d) M_{\eta/2}}{\underline{M}} \zeta_n(\theta). \tag{B.57}$$

Finally, for any $\epsilon > 0$, there exists $N_1 \in \mathbb{N}$ such that $N_1 \geq N$ that does not depend on $P \in \mathcal{P}$ or $\theta \in \Upsilon(P)$ such that

$$\begin{aligned} &P^*\Big(\{|s(p, U_n(\theta)) - s(p, V_n(\theta))| > \epsilon\} \cap \{W_n(\theta) \neq \emptyset \cap V_n(\theta) \neq \emptyset \cap U_n(\theta) \neq \emptyset\}\Big) \\ &\leq P^*\Big(\{|s(p, U_n(\theta)) - s(p, V_n(\theta))| > \epsilon\} \qquad\qquad (B.58) \\ &\qquad \cap \{W_n(\theta) \neq \emptyset \cap V_n(\theta) \neq \emptyset \cap U_n(\theta) \neq \emptyset\} \cap \{(V_n^{-\delta}(\theta))^o \neq \emptyset\}\Big) \\ &\quad + P^*(\{W_n(\theta) \neq \emptyset\} \cap \{(V_n^{-\delta}(\theta))^o = \emptyset\}) \leq \eta, \ \forall n \geq N. \qquad\qquad (B.59) \end{aligned}$$

where the last inequality follows from (B.50), (B.57), and Lemma B.3. This establishes the claim of the lemma. □

Proof of Lemma B.6: The function $u_{j,n,\theta}(\lambda)$ can be written as

$$u_{j,n,\theta}(\lambda)=\sqrt{n}\frac{(\bar{m}_j(X_i,\theta+\lambda/\sqrt{n})-E_P[m_j(X_i,\theta+\lambda/\sqrt{n})])}{\hat{\sigma}_j(\theta+\lambda/\sqrt{n})}+\sqrt{n}\frac{E_P[m_j(X_i,\theta+\lambda/\sqrt{n})]}{\hat{\sigma}_j(\theta+\lambda/\sqrt{n})}-c_n(\theta+\lambda/\sqrt{n})$$
$$=\{\mathbb{G}_{P,j,n}(\theta+\lambda/\sqrt{n})+D_{P,j}(\bar{\theta})'\lambda+h_{P,j,n}(\theta)\}(1+\eta_{j,n}(\theta+\lambda/\sqrt{n}))-c_n(\theta+\lambda/\sqrt{n}),\tag{B.60}$$

where $\eta_{j,n}(\theta)\equiv\sigma_{P,j}(\theta)/\hat{\sigma}_j(\theta)-1$. The second equality follows from the mean value theorem, and $\bar{\theta}$ represents a mean value between $\theta$ and $\theta+\lambda/\sqrt{n}$, which can differ across components of the gradient.

Let $(P_{a_n},\theta_{a_n})\in\{(P,\theta):P\in\mathcal{P},\theta\in\Upsilon(P)\}$ be a subsequence of distributions such that

$$\lim_n P^*_{a_n}(\sup_{\lambda\in\rho_{a_n}B_d\cap\sqrt{a}_n(\Theta-\theta_{a_n})}|\max_{j=1,\cdots,J}u_{j,a_n,\theta_{a_n}}(\lambda)-\max_{j=1,\cdots,J}v_{j,n,\theta_{a_n}}(\lambda)|\ge\epsilon)$$
$$=\limsup_{n\to\infty}\sup_{P\in\mathcal{P}}\sup_{\theta\in\Upsilon(P)}P^*(\sup_{\lambda\in\rho B_d\cap\sqrt{n}(\Theta-\theta)}|\max_{j=1,\cdots,J}u_{j,n,\theta}(\lambda)-\max_{j=1,\cdots,J}v_{j,n,\theta}(\lambda)|\ge\epsilon).\tag{B.61}$$

By passing to a further subsequence $\{m_n\}$, for each $j$, we have either (i) $\kappa^{-1}_{m_n}h_{P,j,m_n}(\theta_{m_n})\to\pi_j\in(-\infty,0]$ or (ii) $\kappa^{-1}_{m_n}h_{P,j,m_n}(\theta_{m_n})\to\pi_j=-\infty$. Define $\mathcal{J}^*\equiv\{j=1,\cdots,J:\pi_j\in(-\infty,0]\}$. By Lemma B.11, $\mathcal{J}^*$ is non-empty.

Define the event

$$A_n\equiv\{\max_{j\in J^*}u_{j,n,\theta_n}(\lambda)=\max_{j=1,\cdots,J}u_{j,n,\theta_n}(\lambda),\ \text{and}\ \max_{j\in\mathcal{J}^*}v_{j,n,\theta_n}(\lambda)=\max_{j=1,\cdots,J}v_{j,n,\theta_n}(\lambda),\ \forall\lambda\in\rho B_d\cap\sqrt{n}(\Theta-\theta_n)\}.\tag{B.62}$$

One may then write

$$\lim_{n\to\infty}P^*_{m_n}(\sup_{\lambda\in\rho B_d\cap\sqrt{m}_n(\Theta-\theta_{m_n})}|\max_{j=1,\cdots,J}u_{j,m_n,\theta_{m_n}}(\lambda)-\max_{j=1,\cdots,J}v_{j,n,\theta_{m_n}}(\lambda)|\ge\epsilon)$$
$$\le\lim_{n\to\infty}P^*_{m_n}(\sup_{\lambda\in\rho B_d\cap\sqrt{m}_n(\Theta-\theta_{m_n})}|\max_{j\in\mathcal{J}^*}u_{j,m_n,\theta_{m_n}}(\lambda)-\max_{j\in\mathcal{J}^*}v_{j,n,\theta_{m_n}}(\lambda)|\ge\epsilon)+P^*_{m_n}(A^c_{m_n}).\tag{B.63}$$

Note that $u_{j,n,\theta_n}(\lambda)$ can be written as in (B.60), and $v_{j,n,\theta}(\lambda)\equiv\mathbb{G}_{P,j,n}(\theta)+\nabla_\theta D_{P,j}(\theta)'\lambda+h_{P,j,n}(\theta)-c_n(\theta)$. Hence, $u_{j,n,\theta_n}(\lambda)=O_{\mathcal{P}}(1)+O(1)+O(h_{P,j,n})$ and $v_{j,n,\theta_n}(\lambda)=O_{\mathcal{P}}(1)+O(1)+O(h_{P,j,n})$ uniformly in $\lambda\in\rho B_d\cap\sqrt{n}(\Theta-\theta_n)$, where we used that $c_n(\theta)\le M$. However, $h_{P,j,n}\to-\infty$ for all $j\notin\mathcal{J}^*$ at a rate faster than $\kappa_n$, which in turn implies $P_{m_n}(A^c_{m_n})\to0$. Therefore, for the conclusion of the lemma, it suffices to show $\lim_{n\to\infty}P^*_{m_n}(\sup_{\lambda\in\rho B_d\cap\sqrt{m}_n(\Theta-\theta_{m_n})}|\max_{j\in\mathcal{J}^*}u_{j,m_n,\theta_{m_n}}(\lambda)-\max_{j\in\mathcal{J}^*}v_{j,m_n,\theta_{m_n}}(\lambda)|\ge\epsilon)=0$.

For each $\lambda\in\mathbb{R}^d$, define $r_{j,n,\theta}(\lambda)\equiv u_{j,n,\theta}(\lambda)-v_{j,n,\theta}(\lambda)$. By the triangle and Cauchy-Schwarz inequalities, for any $\theta\in\Theta$ and $\lambda\in\rho B_d\cap\sqrt{n}(\Theta-\theta)$, we have

$$|r_{j,n,\theta}(\lambda)|\le|\mathbb{G}_{P,j,n}(\theta+\lambda/\sqrt{n})-\mathbb{G}_{P,j,n}(\theta)|+\|D_{P,j}(\bar{\theta})-D_{P,j}(\theta)\|\|\lambda\|+|c_n(\theta+\lambda/\sqrt{n})-c_n(\theta)|$$
$$+|\mathbb{G}_{P,j,n}(\theta+\lambda/\sqrt{n})+D_{P,j}(\bar{\theta})'\lambda+h_{P,j,n}(\theta)|\eta_{j,n}(\theta+\lambda/\sqrt{n})$$
$$\le|\mathbb{G}_{P,j,n}(\theta+\lambda/\sqrt{n})-\mathbb{G}_{P,j,n}(\theta)|+o_{\mathcal{P}}(1)+o_{\mathcal{P}}(1)+\{O_{\mathcal{P}}(1)+O(1)+|h_{P,j,n}(\theta)|\}O_{\mathcal{P}}(n^{-1/2}),\tag{B.64}$$

where the last inequality follows from $\|D_{P,j}(\bar{\theta})-D_{P,j}(\theta)]\|=o_{\mathcal{P}}(1)$ by the Lipschitz continuity of $D_{P,j}$ (Assumption 3.3-(ii)) and $\bar{\theta}$ being a mean value between $\theta$ and $\theta+\lambda/\sqrt{n}$, $\|\lambda\|\le\rho$, the equicontinuity assumption on $c_n$, $\|D_{P,j}(\theta)\|$ being uniformly bounded (Assumption 3.3-(i)), and $\sup_{\theta\in\Theta}|\eta_{j,n}(\theta)|=O_{\mathcal{P}}(n^{-1/2})$ by Assumption 3.3-(iii).

By (B.64), the uniform stochastic equicontinuity of $\{\mathbb{G}_{P,j,n}\}$ (Assumption 3.4), we have

$$\sup_{\lambda\in\rho B_d\cap\sqrt{m}_n(\Theta-\theta_{m_n})}|\max_{j\in\mathcal{J}^*}u_{j,m_n,\theta_{m_n}}(\lambda)-\max_{j\in\mathcal{J}^*}v_{j,n,\theta_{m_n}}(\lambda)|$$
$$\le\sup_{\lambda\in\rho B_d\cap\sqrt{m}_n(\Theta-\theta_{m_n})}\max_{j\in\mathcal{J}^*}|r_{j,n,\theta_{m_n}}(\lambda)|=o_{\mathcal{P}}(1)+\max_{j\in\mathcal{J}^*}|h_{P_{m_n},j,m_n}(\theta_{m_n})|O_{\mathcal{P}}(n^{-1/2})=o_{\mathcal{P}}(1),\tag{B.65}$$

where the last equality follows from $h_{P_{m_n},j,m_n}(\theta_{m_n})=O(\kappa_n)$ for all $j\in\mathcal{J}^*$ and $\kappa_n/n^{1/2}\to0$. The conclusion of the lemma then follows from (B.61), (B.63), and (B.65). □

Proof of Lemma B.7:

Recall that $c_n(\theta)$ is defined as the smallest $c$ such that $\Pr(\mathcal{Z}_n^{\mathbb{G}}(p,c,\theta) \geq 0 \geq -\mathcal{Z}_n^{\mathbb{G}}(-p,c,\theta)) \geq 1-\alpha$. Observe that $c_n(\theta)$ is lower bounded by the smallest $c$ such that $\Pr(\mathcal{Z}_n^{\mathbb{G}}(p,c,\theta) \geq 0) \geq 1-\alpha$, where

$$\mathcal{Z}_n^{\mathbb{G}}(p,c,\theta) \quad = \quad \sup\left\{p'\lambda : D_{P,j}(\theta)'\lambda + \tilde{\zeta}_{n,j}(\theta) \leq c - \mathbb{G}_{P,j}(\theta), j = 1,...,J\right\}. \tag{B.66}$$

We work with this one-sided $c$ below. With abuse of notation, we denote this one sided $c$ by $c_n(\theta)$.

Here, $\{\mathbb{G}_{P,j}(\theta), j = 1,\cdots,J\}$ jointly follows $N(0,\Omega_P(\theta))$. Note that $c_n$ depends on $P$ through $\Omega_P$, $D_P$, and $\tilde{\zeta}_{n,j}(\theta) \equiv \varphi^*(\kappa_n^{-1}\sqrt{n}E_P(m_j(\theta))/\sigma_{P,j}(\theta))$. To show its dependence on $P$, we write it as $c_{n,P}(\theta)$ below. Let

$$\tilde{c} \equiv \liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Upsilon(P)} c_{n,P}(\theta). \tag{B.67}$$

In what follows, let $\{a_n\}$ be a subsequence of $\{n\}$ such that

$$\lim_{a_n\to\infty} c_{a_n,P_{a_n}}(\theta_{a_n}) = \tilde{c}. \tag{B.68}$$

This subsequence exists because $c_n(\theta) \leq M$. Let $\{l_n\}$ be a further subsequence such that for some $(\Omega, D, h, \tilde{\zeta}) \in \mathbb{R}^{J\times J} \times \mathbb{R}^{d\times J} \times (\mathbb{R}_- \cup \{-\infty\})^J \times (\mathbb{R}_- \cup \{-\infty\})^J$,

$$\Omega_{P_{l_n}}(\theta_{l_n}) \to \Omega,\ D_{P_{l_n}}(\theta_{l_n}) \to D,\ h_{l_n}(\theta_{l_n}) \to h, \text{ and } \tilde{\zeta}_{l_n}(\theta_{l_n}) \to \tilde{\zeta}. \tag{B.69}$$

Define

$$\mathcal{Z}^{\mathbb{W}}(p,c) \quad \equiv \quad \sup\left\{p'\lambda \in \rho B_d : D_j'\lambda + \tilde{\zeta}_j \leq c - \mathbb{W}_j, j = 1,...,J\right\}, \tag{B.70}$$

where $\mathbb{W}_j, j = 1,\cdots,J$ jointly follow $N(0,\Omega)$.

We proceed below by taking the following steps.

Step 1: The claim of this step is that $\min\{c \geq 0 : \Pr(\mathcal{Z}^{\mathbb{W}}(p,c) \geq 0) \geq 1-\alpha\} \leq \tilde{c}$.

For each $n$, let $\mathbb{W}_{l_n} \sim N(0, \Omega_{P_{l_n}}(\theta_{l_n}))$. Then, one may write

$$\mathcal{Z}_{l_n}^{\mathbb{G}}(p,c,\theta_{l_n}) \stackrel{d}{=} \mathcal{Z}_{l_n}^{\mathbb{W}_{l_n}}(p,c,\theta_{l_n}) \equiv \sup\left\{p'\lambda \in \rho B_d : D_{P_{l_n},j}(\theta_{l_n})'\lambda + \tilde{\zeta}_{l_n,j}(\theta_{l_n}) \leq c - \mathbb{W}_{l_n}, j = 1,...,J\right\}. \tag{B.71}$$

By (B.69), the characteristic function of $\mathbb{W}_{l_n}$ converges pointwise to that of $\mathbb{W}$. By Lévy's continuity theorem, it then follows that $\mathbb{W}_{l_n} \stackrel{d}{\to} \mathbb{W}$. Take Skorokhod representations $\mathbb{W}_{l_n}^*, \mathbb{W}^*$ of $\mathbb{W}_{l_n}$ and $\mathbb{W}$ such that $\mathbb{W}_{l_n}^* \stackrel{a.s.}{\to} \mathbb{W}^*$. Let $c_{l_n}$ be an arbitrary sequence such that $c_{l_n} \to c^*$. Then, by (B.69) and Lemma C.1 (whose assumptions are verified in Appendix C.1 and C.2), it follows that $\mathcal{Z}_{l_n}^{\mathbb{W}_{l_n}^*}(p, c_{l_n}, \theta_{l_n}) \stackrel{a.s.}{\to} \mathcal{Z}^{\mathbb{W}^*}(p,c^*)$. This and (B.71) in turn imply that

$$\mathcal{Z}_{l_n}^{\mathbb{G}}(p, c_{l_n}, \theta_{l_n}) \stackrel{d}{\to} \mathcal{Z}^{\mathbb{W}}(p,c^*). \tag{B.72}$$

Therefore, we have

$$\limsup_{l_n\to\infty} \Pr(\mathcal{Z}_{l_n}^{\mathbb{G}}(p,c_{l_n},\theta_{l_n}) \geq 0) \leq \Pr(\mathcal{Z}^{\mathbb{W}}(p,c^*) \geq 0). \tag{B.73}$$

Now take $c_{l_n} = c_{l_n,P_{l_n}}(\theta_{l_n})$ in (B.73). Then, by (B.68) and (B.73), we have

$$1-\alpha \leq \Pr(\mathcal{Z}^{\mathbb{W}}(p,\tilde{c}) \geq 0). \tag{B.74}$$

Hence, $\min\{c \geq 0 : \Pr(\mathcal{Z}^{\mathbb{W}}(p,c) \geq 0) \geq 1-\alpha\} \leq \tilde{c}$.

Step 2: The claim of this step is that there exists $\underline{c} > 0$ such that $\min\{c \geq 0 : \Pr(\mathcal{Z}^{\mathbb{W}}(p,c) \geq 0) \geq 1-\alpha\} \geq \underline{c}$. We further divide the argument in cases. Let $K$ be the $(J+2d)\times d$ matrix collecting in the first $J$ rows the matrix $D$, and below it the matrices $I_d$ and $-I_d$. We liberally use the fact that dropping constraints estimates $\mathcal{Z}^{\mathbb{W}}(p,c)$ from above and therefore $c^{\mathbb{W}} = \min\{c \geq 0 : P(\mathcal{Z}^{\mathbb{W}}(p,c) \geq 0) \geq 1-\alpha\}$ from below.

Case 1: $\theta_{l_n} \in H(p, \Theta_I(P_{l_n}))$.

Because $\theta_{l_n} \in H(p, \Theta_I(P_{l_n}))$ for all $n$, there exists a sequence of non-empty sets $\mathcal{J}(\theta_{l_n}) \subseteq \{1, \cdots, J\}$ such that $\mathcal{J}(\theta_{l_n}) \equiv \{j \in \{1, \cdots, J_1 + 2J_2\} : \tilde{\zeta}_{l_n,j}(\theta_{l_n}) = 0\}$. Note that $\mathcal{J}(\theta_{l_n})$ is a subset of a finite set for all $n$. Hence, there is a further subsequence $\{k_n\}$ of $\{l_n\}$ and $\mathcal{J}^* \subset \{1, \cdots, J\}$ such that $\mathcal{J}(\theta_{k_n}) = \mathcal{J}^*$ for all $n$. Hence, $\tilde{\zeta}_{k_n,j}(\theta_{k_n}) = 0$ for all $n$ and $j \in \mathcal{J}^*$. Then, by (B.69), it follows that $\tilde{\zeta}_j = 0$ for all $j \in \mathcal{J}^*$.

Let $\Lambda(K^*, g^{*\mathbb{W}}, c)$ denote the constraint set of the problem defining $\mathscr{Z}^{\mathbb{W}}(p, c)$ corresponding to $j \in \mathcal{J}^*$, with $g_j^{*\mathbb{W}} = -\mathbb{W}_j - \tilde{\zeta}_j$ for $j \in \mathcal{J}^* \cap \{1, \ldots, J\}$, and $g_j^{*\mathbb{W}} = \rho$ for $j \in \mathcal{J}^* \cap \{J+1, \ldots, J+2d\}$, with $\Lambda(\cdot, \cdot, \cdot)$ is defined in equation A.9. For any equality, at most one of the corresponding inequalities constrains the value of $\max_{\lambda \in \Lambda(K^*, 0, 0)} p'\lambda$. Remove one inequality that does not constrain this value. Observe that after this simplification, $\Lambda(K^*, 0, 0)$ cannot lose dimensionality due to containing equalities but only due to inequalities intersecting on a subspace.

Suppose first that $\Lambda(K^*, 0, 0)$ has a non-empty interior.

Consider the cone $\{\lambda : D_j\lambda \leq 0, j \in \mathcal{J}^*\}$, which is a superset of $\Lambda(K^*, 0, 0)$. By the lower bound on gradients, $\lambda = 0$ solves $\max_{\lambda \in \Lambda(K^*, 0, 0)} p'\lambda$. This requires that a Karush-Kuhn-Tucker (KKT) condition applies at $\lambda = 0$. Define a set $\mathcal{J} \subseteq \mathcal{J}^*$ such that

$$p = \sum_{j \in \mathcal{J}} D_j'\mu_j, \quad \text{and } \mu_j > 0, \forall j \in \mathcal{J}, \tag{B.75}$$

and no strict subset of $\mathcal{J}$ satisfies (B.75). By the definition of $\mathcal{J}$, $\{D_j, j \in \mathcal{J}\}$ are linearly independent.

Consider first the following special case: $\tilde{\zeta}_j = 0$ for all $j$ in $\mathcal{J}$ and $\tilde{\zeta}_j = -\infty$ otherwise. Let $D_{\mathcal{J}}$ be a $\#\mathcal{J} \times d$ matrix that stacks $\{D_j', j \in \mathcal{J}\}$. Similarly, let $\Omega_{\mathcal{J}}$ be the correlation matrix of $\{\mathbb{W}_j, j \in \mathcal{J}\}$. In this special case, one may then write

$$\mathscr{Z}^{\mathbb{W}}(p, c) \quad = \quad \sup \left\{p'\lambda : D_j'\lambda \leq c - \mathbb{W}_j, j \in \mathcal{J}\right\}, \tag{B.76}$$

By KKT, the optimal value in (B.76) is bounded, and the problem above is solved (not necessarily uniquely) by $\lambda^*(c) = D_{\mathcal{J}}'(D_{\mathcal{J}}D_{\mathcal{J}}')^{-1}(c \cdot \mathbf{1}_{\mathcal{J}} - \mathbb{W}_{\mathcal{J}})$, where $1_{\mathcal{J}}$ is a $\#\mathcal{J} \times 1$ vector of ones. Thus, by (B.75), the problem's optimal value is given by $\mathscr{Z}^{\mathbb{W}}(p, c) = \mu_{\mathcal{J}}'(c \cdot \mathbf{1}_{\mathcal{J}} - \mathbb{W}_{\mathcal{J}})$, where $\mu_{\mathcal{J}}$ is a $\#\mathcal{J} \times 1$ vector that stacks $\mu_j, j \in \mathcal{J}$. This object is distributed as $N(c\mu_{\mathcal{J}}'\mathbf{1}_{\mathcal{J}}, \mu_{\mathcal{J}}'\Omega_{\mathcal{J}}\mu_{\mathcal{J}})$. Hence, letting $\rho \equiv \mu_{\mathcal{J}} / ||\mu_{\mathcal{J}}||$, we have

$$\begin{aligned}
&\min\{c \geq 0 : \Pr(\mathscr{Z}^{\mathbb{W}}(p, c) \geq 0) \geq 1 - \alpha\} \\
&\quad = \Phi^{-1}(1-\alpha) \times \sqrt{\mu_{\mathcal{J}}\Omega_{\mathcal{J}}\mu_{\mathcal{J}}'} / \left(\mu_{\mathcal{J}}'\mathbf{1}\right) = \Phi^{-1}(1-\alpha) \times \sqrt{\rho\Omega_{\mathcal{J}}\rho'} / \left(\rho'\mathbf{1}\right) \geq \Phi^{-1}(1-\alpha) \times \lambda_{\min}(\Omega_{\mathcal{J}})^{1/2}/d.
\end{aligned} \tag{B.77}$$

By Assumption 3.2 (iv), there exists a constant $\epsilon > 0$ that does not depend on $\theta$ such that the smallest eigenvalue of $\Omega_P(\theta)$ is bounded from below by $\epsilon$ for all $\theta$ and $P \in \mathcal{P}$. Define $\underline{c} \equiv \Phi^{-1}(1-\alpha) \times \epsilon^{1/2}/d$. Then, the conclusion of this step follows for the special case.

In general, $\mathscr{Z}^{\mathbb{W}}$ is given by

$$\begin{aligned}
\mathscr{Z}^{\mathbb{W}}(p, c) = \sup \Big\{p'\lambda : &D_j'\lambda \leq c - \mathbb{W}_j, \ j \in \mathcal{J}, \\
&D_j'\lambda + \tilde{\zeta}_j \leq c - \mathbb{W}_j, \ j \notin \mathcal{J}\Big\},
\end{aligned} \tag{B.78}$$

where $\tilde{\zeta}_j \in \mathbb{R}_- \cup \{-\infty\}$ for $j \notin \mathcal{J}$. Note that adding the constraints with $j \notin \mathcal{J}$ weakly decreases $\mathscr{Z}^{\mathbb{W}}(p, c)$ and therefore weakly increases $\min\{c \geq 0 : \Pr(\mathscr{Z}^{\mathbb{W}}(p, c) \geq 0) \geq 1 - \alpha\}$. Hence, $\underline{c}$ is also a valid lower bound for the general case.

By (B.67) and Steps 1-2, it then follows that

$$\liminf_{n \to \infty} \inf_{P \in \mathcal{P}} \inf_{\theta \in H(p, \Theta_I(P))} c_{n,P}(\theta) \geq \underline{c} > 0. \tag{B.79}$$

Suppose now that $\Lambda(K^*, 0, 0)$ has no interior, i.e. is lower dimensional.

Remove constraints until left with an irreducible set $\widetilde{C}$ containing $\widetilde{d} \leq d$ constraints s.t. $\left\{D^{\widetilde{C}}\lambda \leq 0\right\}$ has dimension strictly less than $\widetilde{d}$. (Irreducibility means that for any index set $C^* \subset \widetilde{C}$, the cone $\left\{D^{C^*}\lambda \leq 0\right\}$ is of dimension $\widetilde{d}$.) Without loss of generality set $\widetilde{C} = \{1, ..., \widetilde{d}\}$. For $\left\{D^{[1:\widetilde{d}]}\lambda \leq 0\right\}$ to be lower dimensional, the hyperplane $\left\{D^{[\widetilde{d}]}\lambda = 0\right\}$ must separate the halfspace $\left\{D^{[\widetilde{d}]}\lambda \leq 0\right\}$ from the cone $\left\{D^{[1:\widetilde{d}-1]}\lambda \leq 0\right\}$. Hence, any element of $\left\{D^{[1:\widetilde{d}-1]}\lambda = 0\right\}$ minimizes $D^{[\widetilde{d}]}\lambda$ subject to $D^{[1:\widetilde{d}-1]}\lambda \leq 0$. As this minimum is attained, it is characterized by a KKT condition, thus $D^{[\widetilde{d}]} = \mu'\left(-D^{[1:\widetilde{d}-1]}\right) = -\mu' D^{[1:\widetilde{d}-1]}$ for some LM vector $\mu \in \mathbb{R}_+^{\widetilde{d}-1}$. If there existed such a $\mu$ with one or more zero components, then the corresponding constraints could be dropped without changing the solution to the program and hence the intersection of halfspaces, contradicting irreducibility. Hence, no $\mu$ can have zero elements, but that means that $\mu$ is unique and that $D^{[1:\widetilde{d}-1]}$ has full row rank. This, in turn, means that for any $g$, $\left\{D^{[1:\widetilde{d}-1]}\lambda = g^{[1:\widetilde{d}-1]}\right\}$ is nonempty and (by sufficiency of KKT in linear programs) any element of it minimizes $D^{[\widetilde{d}]}\lambda$ subject to $D^{[1:\widetilde{d}-1]}\lambda \leq g^{[1:\widetilde{d}-1]}$. This, in turn, means that $\left\{D^{[1:\widetilde{d}]}\lambda \leq g^{[1:\widetilde{d}]}\right\} \neq \emptyset$ iff $D^{[\widetilde{d}]}\lambda^* \leq g^{[\widetilde{d}]}$, where $\lambda^* = D^{[1:\widetilde{d}-1]\prime}\left(D^{[1:\widetilde{d}-1]}D^{[1:\widetilde{d}-1]\prime}\right)^{-1} g^{[1:\widetilde{d}-1]}$. This condition can be written as

$$\begin{aligned}
D^{[\widetilde{d}]}D^{[1:\widetilde{d}-1]\prime}\left(D^{[1:\widetilde{d}-1]}D^{[1:\widetilde{d}-1]\prime}\right)^{-1} g^{[1:\widetilde{d}-1]} - g^{[\widetilde{d}]} &\leq 0 \\
-\mu' g^{[1:\widetilde{d}-1]} - g^{[\widetilde{d}]} &\leq 0 \iff \\
\iff \left[\begin{array}{ccc} -\mu', & -1, & \underset{[1\times(J^*-\widetilde{d})]}{0} \end{array}\right] g &\leq 0,
\end{aligned}$$

where we substituted for $D^{[\widetilde{d}]} = -\mu D^{[1:\widetilde{d}-1]}$ and simplified. This condition is linear in $g$ and just fulfilled at $g = 0$, thus its probability under the multivariate normal distribution of $g$ is exactly $1/2$. But then, $\Pr(\mathcal{Z}^{\mathbb{W}}(p, 0) \geq 0) \leq \Pr(\Lambda(K^*, g^{*\mathbb{W}}, 0) \neq \emptyset) \leq 1/2$.

Next, consider $\Lambda(K^*, g^{*\mathbb{W}}, c)$ for some $c \geq 0$. The arguments involving Lagrange multipliers are unchanged, and so a necessary condition for $\Lambda(K^*, g^{*\mathbb{W}}, c)$ to be nonempty is that

$$\begin{aligned}
-\mu'\left(g^{*\mathbb{W}[1:\widetilde{d}-1]} + c \cdot 1_{\widetilde{d}-1}\right) &\leq g^{*\mathbb{W}[\widetilde{d}]} + c \\
\iff c &\geq -\frac{[\mu', 1]g^{*\mathbb{W}[1:\widetilde{d}]}}{[\mu', 1]1_{\widetilde{d}}},
\end{aligned}$$

thus we can write

$$\Pr(\mathcal{Z}^{\mathbb{W}}(p, c) \geq 0) \leq \Pr(\Lambda(K^*, g^{*\mathbb{W}}, c) \neq \emptyset) \leq \Pr\left(c \geq -\frac{[\mu', 1]g^{*\mathbb{W}[1:\widetilde{d}]}}{[\mu', 1]1_{\widetilde{d}}}\right) = \Pr\left(\frac{[\mu', 1]g^{*\mathbb{W}[1:\widetilde{d}]}}{[\mu', 1]1_{\widetilde{d}}} \leq c\right).$$

where the last step used that $[\mu', 1]g^{*\mathbb{W}[1:\widetilde{d}]}$ is distributed symmetrically about zero. The r.h. probability is the c.d.f. of a normal r.v. centered at 0, so we immediately have $c^{\mathbb{W}} \geq 0$ for $\alpha = 1/2$ and $c^{\mathbb{W}} > 0$ for $\alpha < 1/2$. The relevant case for bounding the c.d.f. is, therefore, $c \geq 0$, in which case the c.d.f. is maximized by minimizing the r.v.'s variance. Write

$$\frac{[\mu', 1]g^{*\mathbb{W}[1:\widetilde{d}]}}{[\mu', 1]1_{\widetilde{d}}} = \frac{\frac{[\mu', 1]}{\|[\mu', 1]\|}g^{*\mathbb{W}[1:\widetilde{d}]}}{\frac{[\mu', 1]}{\|[\mu', 1]\|}1_{\widetilde{d}}}.$$

The numerator is distributed $N(0, \varsigma^2)$, where $\varsigma^2 \geq \omega$ by Assumption 3.2-(v). Also, $\left\|\frac{[\mu', 1]}{\|[\mu', 1]\|}1_{\widetilde{d}}\right\| \leq 1 \cdot \left\|1_{\widetilde{d}}\right\| = \sqrt{\widetilde{d}} \leq \sqrt{d}$. Using these bounds to bound the variance from below, one gets

$$c \geq \omega^{1/2} d^{-1/2} \Phi^{-1}(1 - \alpha).$$

This establishes the claim of the lemma for this case.

Case 2: $\theta_{l_n} \notin H(p, \Theta_I(P_{l_n}))$ and there exists a direction (=vector of unit length) $r$ s.t. $D_j(\theta_{l_n})r \leq -{l_n}^{-1/4}$ for all $j \in \mathcal{J}^*$.

Let $T = \{\lambda / \|\lambda\| : \lambda \in \Lambda(K^*, 0, 0)\}$ collect first-order feasible directions in $\Lambda(K^*, 0, 0)$. Fix $q \in T$. For any $j = 1, \ldots J$ and scalar $\alpha > 0$, the mean value theorem implies

$$\frac{E_P(m_j(\theta_{l_n} + \alpha q))}{\sigma_{P,j}(\theta_{l_n} + \alpha q)} = \frac{E_P(m_j(\theta_{l_n}))}{\sigma_{P,j}(\theta_{l_n})} + \alpha D_j(\bar{\theta}_{l_n}) q$$

for some $\bar{\theta}_{l_n}$ that is componentwise between $\theta_{l_n}$ and $\theta_{l_n} + \alpha q$.

If $j \notin \mathcal{J}^*$, then $E_P(m_j(\theta_{l_n}))/\sigma_{P,j}(\theta_{l_n}) < -\kappa_{l_n} {l_n}^{-1/2}$ and therefore

$$\frac{E_P(m_j(\theta_{l_n} + \alpha q))}{\sigma_{P,j}(\theta_{l_n} + \alpha q)} \leq -\kappa_{l_n} {l_n}^{-1/2} + \alpha D_j(\bar{\theta_{l_n}}) q \leq -\kappa_{l_n} {l_n}^{-1/2} + \alpha \bar{M},$$

where we used Assumption 3.3-(i). This quantity is nonpositive for $\alpha \leq \kappa_{l_n} {l_n}^{-1/2} \bar{M}^{-1}$.

For any $j \in \mathcal{J}^*$, $q \in T$ implies $D_j(\theta_{l_n}) q \leq 0$ and therefore

$$\frac{E_P(m_j(\theta_{l_n} + \alpha q))}{\sigma_{P,j}(\theta_{l_n} + \alpha q)} \leq = \underbrace{\frac{E_P(m_j(\theta_{l_n}))}{\sigma_{P,j}(\theta_{l_n})}}_{\leq 0} + \underbrace{\alpha D_j(\theta_{l_n}) q}_{\leq 0} + \alpha \left(D_j(\bar{\theta}_{l_n}) - D_j(\theta_{l_n})\right) q \leq M\alpha^2$$

using that $\|q\| = 1$ and that $D_j(\theta)$ is Lipschitz continuous with Lipschitz constant $M$ by Assumption 3.3-(i). Next, the mean value theorem yields

$$\frac{E_P(m_j(\theta_{l_n} + \alpha q + \beta r))}{\sigma_{P,j}(\theta_{l_n} + \alpha q + \beta r)} = \frac{E_P(m_j(\theta + \alpha q))}{\sigma_{P,j}(\theta_{l_n} + \alpha q)} + \beta D_j(\bar{\theta}_{l_n}) r,$$

where $\bar{\theta}_{l_n}$ lies componentwise between $\theta_{l_n} + \alpha q$ and $\theta_{l_n} + \alpha q + \beta r$. Hence,

$$\begin{aligned} \frac{E_P(m_j(\theta_{l_n} + \alpha q + \beta r))}{\sigma_{P,j}(\theta_{l_n} + \alpha q + \beta r)} &\leq M\alpha^2 + \beta D_j(\theta_{l_n}) r + \beta \left(D_j(\bar{\theta}_{l_n}) - D_j(\theta_{l_n})\right) r \\ &\leq M\alpha^2 - {l_n}^{1/4} \beta + \beta(\beta + \alpha) M. \end{aligned}$$

Let $\alpha = \kappa_{l_n} {l_n}^{-1/2} \bar{M}^{-1}$ and $\beta = 3\kappa_{l_n}^2 {l_n}^{-3/4} M \bar{M}^{-2}$. For $l_n$ large enough (namely $l_n \kappa_{l_n}^{-4} \geq 81 M^4$, a threshold that does not depend on moving parameters) this implies $\beta \leq \alpha$ and therefore

$$\begin{aligned} \frac{E_P(m_j(\theta_{l_n} + \alpha q + \beta r))}{\sigma_{P,j}(\theta_{l_n} + \alpha q + \beta r)} &\leq 3M\alpha^2 - {l_n}^{-1/4} \beta \\ &= 3M \left(\kappa_{l_n} {l_n}^{-1/2} \bar{M}^{-1}\right)^2 - {l_n}^{-1/4} 3\kappa_{l_n}^2 {l_n}^{-3/4} M \bar{M}^{-2} = 0. \end{aligned}$$

We conclude that for $l_n$ large enough, $q \in T$ implies $\theta_{l_n} + \kappa_{l_n} {l_n}^{-1/2} \bar{M}^{-1} q \in \Theta_I(P)$. Thus, if $p'q = O(\kappa_{l_n}^{-3/4})$, then one can write

$$s(p, \Theta_I(P)) - s(-p, \Theta_I(P)) \geq p' \left(\theta_{l_n} + \kappa_{l_n} {l_n}^{-1/2} \bar{M}^{-1} q\right) - p'\theta_{l_n} = \kappa_{l_n} {l_n}^{-1/2} \bar{M}^{-1} p'q = O(\kappa_{l_n}^{1/4} n^{-1/2})$$

so that the projection is long, thereby violating the assumptions of the Lemma. This obviously extends to larger $p'q$.

Next, if $\max\{p'q : q \in T\} = o(\kappa_{l_n}^{-3/4})$, then the bound from case 1 applies asymptotically. To see this, let $r$ denote the angle of the simple rotation $R$ that rotates $p$ into $\widehat{p}$, the projection of $p$ subject to unit length into the normal cone corresponding to the tangent cone $T$, denoted $N$. Then $\widehat{p}$ is on $\partial N$ and $T$ therefore contains a direction $\widehat{q} \perp \widehat{p}$. By our current assumption we then have $p'\widehat{q} = o(\kappa_{l_n}^{-3/4})$, hence $\sin r = 1 - p'\widehat{p} = o(\kappa_{l_n}^{-3/4})$, hence $\cos r = 1 - o(\kappa_{l_n}^{-3/4})$.

Now fix any $\Lambda(K^*, g^{*\mathbb{W}}, c)$. We show that $\max_{\lambda \in \Lambda(K^*, g^{*\mathbb{W}}, c)} p'\lambda = \max_{\lambda \in \Lambda(K^*, g^{*\mathbb{W}}, c)} \widehat{p}'\lambda + o(\kappa_{l_n}^{-3/4})$. Because case 1 would apply if the direction of projection were $\widehat{p}$, it then applies asymptotically to $p$. Initially assume $\Theta \subset \mathbb{R}^2$, fix any $\lambda \in \Lambda(K^*, g^{*\mathbb{W}}, c)$ and write

$$\widehat{p}'\lambda = (Rp)'\lambda = p' \begin{bmatrix} \cos r & \sin r \\ -\sin r & \cos r \end{bmatrix} \lambda = p' \left(I_2 + o(\kappa_{l_n}^{-3/4})\right) \lambda = p'\lambda + o(\kappa_{l_n}^{-3/4}).$$

This concludes the proof because the difference of maxima is bounded above by the maximum of differences.

For higher-dimensional parameter spaces, assume w.l.o.g. (because it can always be ensured by orthonormal base change) that $p = (1, 0, ..., 0)$ and that $R = \begin{bmatrix} \cos r & \sin r & 0 \\ -\sin r & \cos r & 0 \\ 0 & 0 & I_{d-2} \end{bmatrix}$, so that the above algebra goes through with minimal modification.

Case 3: $\theta_{l_n} \notin H(p, \Theta_I(P_{l_n}))$ and for any direction (vector of unit length) $r$ one has $D_j(\theta_{l_n})r > -{l_n}^{-1/4}$ for some $j \in \mathcal{J}^*$.

Let $y = \inf\{x : \Lambda(K^*, 0, x) \neq \emptyset\}$, then $\Lambda(K^*, 0, y)$ is nonempty but has no interior. For any $\lambda \in \Lambda(K^*, 0, 0)$, there exists $j \in \mathcal{J}^*$ s.t. $D_j\lambda / \|\lambda\| > -{l_n}^{-1/4}$, implying that $D_j\lambda > -\|\lambda\| \, {l_n}^{-1/4} \geq -\rho\sqrt{d}{l_n}^{-1/4}$. It follows that $\Lambda(K^*, 0, -\rho\sqrt{d}{l_n}^{-1/4}) = \emptyset$ and therefore that $y > -\rho\sqrt{d}{l_n}^{-1/4}$. Next,

$$\Lambda(K^*, g^{*\mathbb{W}}, c - \rho\sqrt{d}{l_n}^{-1/4}) = \Lambda(K^*, g^{*\mathbb{W}} - \rho\sqrt{d}{l_n}^{-1/4}1_{J^*}, c).$$

Assumption 3.2-(v) ensures that

$$\Pr\left(\Lambda(K^*, g^{*\mathbb{W}}, c) \neq \emptyset, \Lambda(K^*, g^{*\mathbb{W}} - \rho\sqrt{d}{l_n}^{-1/4}1_{J^*}, c) = \emptyset\right) \to 0$$

and therefore the bound from case 1 applies.

□

Proof of Lemma B.8: By definition, $\hat{\zeta}_{n,j} \equiv \varphi(\kappa_n^{-1}\sqrt{n}\bar{m}_{j,n}(\theta))/\hat{\sigma}_{n,j}(\theta))$, and therefore $\hat{\zeta}^*_{n,j}(\theta) \leq \hat{\zeta}_{n,j}(\theta)$ for each $n$ and $\theta$. This implies that the constraint set in (A.8) used to define $\mathcal{Z}_n^{b*,\rho}(p, c, \theta)$ in (A.13) is tighter than the constraint set used to define (A.8) in (A.12), and therefore $\mathcal{Z}_n^{b*,\rho}(p, c, \theta) \geq \mathcal{Z}_n^{b,\rho}(p, c, \theta)$, which in turn yields $\hat{c}_n(\theta) \geq \hat{c}^*_n(\theta)$ for each $n$ and $\theta$. □

Proof of Lemma B.9: Let $(\theta_n, P_{\gamma_n}) \in \{(\theta, P), P \in \mathcal{P}, \theta \in \Upsilon(P)\}$ be a sequence such that

$$\limsup_{n\to\infty} P_{\gamma_n}\left(|\hat{c}^*_n(\theta_n) - c_n(\theta_n)| > \epsilon\right) = \limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \sup_{\theta\in\Upsilon(P)} P\left(|\hat{c}^*_n(\theta) - c_n(\theta)| > \epsilon\right). \tag{B.80}$$

Let $\{a_n\}$ be a subsequence of $\{n\}$ such that

$$\lim_{n\to\infty} P_{\gamma_{a_n}}\left(|\hat{c}^*_{a_n}(\theta_{a_n}) - c_{a_n}(\theta_{a_n})| > \epsilon\right) = \limsup_{n\to\infty} P_{\gamma_n}\left(|\hat{c}^*_n(\theta_n) - c_n(\theta_n)| > \epsilon\right). \tag{B.81}$$

Observe that by (A.11)-(A.12), $\mathcal{Z}_n^{\mathbb{G},\rho}(p, c, \theta)$ and $\mathcal{Z}_n^{b*,\rho}(p, c, \theta)$ differ only in the constraint set, with the bootstrap process and the estimated gradient replaced by their limits, and $\hat{\zeta}^*_n(\theta)$ replaced by $\tilde{\zeta}_{P,n}(\theta)$. Below, we argue that the convergence of the bootstrap process, the estimated gradient, and the smooth GMS term imply the desired result.

By passing to a further subsequence $\{l_n\}$ of $\{a_n\}$, one may assume

$$\kappa_{l_n}^{-1} h_{P_{\gamma_{l_n}}, l_n, j}(\theta_{l_n}) \to \pi_j \in [-\infty, 0], \; j = 1, \cdots, J, \tag{B.82}$$

and therefore by continuity of $\varphi^*$, $\tilde{\zeta}_{P,n,j}(\theta) \to \varphi^*(\pi_j)$ for $j = 1, \cdots, J$. Define $\mathcal{J}^* \equiv \{j = 1, \cdots, J : \pi_j \in (-\infty, 0]\}$. Note that $\mathcal{J}^*$ is non-empty by Lemma B.11. Because $\sqrt{n}(\bar{m}^b_{n,j}(\theta_n) - E_P[m_j(X_i, \theta_n)]) = O_{\mathcal{P}}(1)$ and $\varphi^*$ is continuous, it follows that for any $j \in \mathcal{J}^*$,

$$\begin{aligned}\hat{\zeta}^*_{n,j}(\theta_{l_n}) &= \varphi^*\left(\kappa_{l_n}^{-1}\frac{\sqrt{l_n}\bar{m}^b_{l_n,j}(\theta_{l_n})}{\hat{\sigma}_{l_n,j}(\theta_{l_n})}\right) \\ &= \varphi^*\left(\kappa_{l_n}^{-1}\frac{\sqrt{l_n}(\bar{m}_{l_n,j} - E_{P_{\gamma_{l_n}}}[m_j(X_i, \theta_{l_n})]}{\hat{\sigma}_{l_n,j}(\theta_{l_n})} + \kappa_{l_n}^{-1}h_{P_{\gamma_{l_n}}, l_n, j}(\theta_{l_n})\frac{\sigma_{P_{\gamma_{l_n}}, j}(\theta_{l_n})}{\hat{\sigma}_{l_n,j}(\theta_{l_n})}\right) = \tilde{\zeta}_{P_{\gamma_{l_n}}, l_n, j}(\theta_{l_n}) + o_{\mathcal{P}}(1).\end{aligned} \tag{B.83}$$

Define the event

$$A_{l_n} \equiv \{\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(p,c,\theta_{l_n}) = U(\mathbb{G}_{P_{l_n}}^{\mathcal{J}^*}(\theta_{l_n}), D_{P_{l_n}}^{\mathcal{J}^*}(\theta_{l_n}), \tilde{\zeta}_{P_{l_n},l_n}^{\mathcal{J}^*}(\theta_{l_n}), c, \rho),$$
$$\text{and } \mathcal{Z}_{l_n}^{b,\rho}(p,c,\theta_{l_n}) = U(\mathbb{G}_{P_{l_n}}^{b,\mathcal{J}^*}(\theta_{l_n}), \hat{D}_{P_{l_n}}^{\mathcal{J}^*}(\theta_{l_n}), \hat{\zeta}_{l_n}^{*,\mathcal{J}^*}(\theta_{l_n}), c, \rho)\}. \quad \text{(B.84)}$$

Arguing as in (B.62), one can show $P_{\gamma_{l_n}}(A_{l_n}) \to 1$.

By Lemma D.2.8 in Bugni, Canay, and Shi (2015), $\mathbb{G}_n^b \overset{d}{\to} \mathbb{G}_P$ in $l^\infty(\Theta)$ uniformly in $P$ conditional on $\{X_1, \cdots, X_n\}$. In what follows, we take $(\mathbb{G}_n^{b*}, \mathbb{G}_P^*)$ to be the almost sure representation of $(\mathbb{G}_n^b, \mathbb{G}_P)$ such that $\mathbb{G}_n^{b^*}$'s distribution equals the distribution of $\mathbb{G}_n^b$ conditional on $\{X_1, \cdots, X_n\}$, $\mathbb{G}_P^* \overset{d}{=} \mathbb{G}_P$, and $\mathbb{G}^{b^*} \overset{a.s.}{\to} \mathbb{G}_P^*$. Below, we assume that $\mathcal{Z}_n^{b*,\rho}$ and $\mathcal{Z}_n^{\mathbb{G}^*,\rho}$ are defined for $(\mathbb{G}_n^{b*}, \mathbb{G}_P^*)$. Then, for any $\epsilon > 0$,

$$\begin{aligned} & P_{\gamma_{l_n}}(\mathcal{Z}_{l_n}^{b*}(p,c,\theta_{l_n}) \ge -\varepsilon \cap \mathcal{Z}_{l_n}^{b*}(-p,c,\theta_{l_n}) \ge -\varepsilon \cap A_{l_n}) \\ & \quad \ge P_{\gamma_{nu_n}}\left(\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(p,c,\theta_{l_n}) \ge 0 \ge -\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(p,c,\theta_{l_n}) \cap A_{l_n}\right) \\ & \quad - P_{\gamma_{l_n}}\left(\max_{q \in p,-p} |\mathcal{Z}_{l_n}^{b*,\rho}(q,c,\theta_{l_n}) - \mathcal{Z}_n^{\mathbb{G},\rho}(q,c,\theta_{l_n})| \ge \epsilon \cap A_{l_n}\right) \\ & \quad \ge P_{\gamma_{l_n}}\left(\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(p,c,\theta_{l_n}) \ge 0 \ge -\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(-p,c,\theta_{l_n}) \cap A_{l_n}\right) - \epsilon \end{aligned} \quad \text{(B.85)}$$

for all $n$ sufficiently large, where the second inequality follows from $\mathbb{G}^{b^*} \overset{a.s.}{\to} \mathbb{G}_P^*$, $\hat{D}_n \overset{p}{\to} D_P$ uniformly in $P$ by Assumption 3.3 (iii), (B.83), and Lemma C.1 (whose assumptions are verified in Appendix C.1 and C.2). Noting that $P_{\gamma_{l_n}}(A_{l_n}) \to 1$, it then follows that for any $\epsilon > 0$

$$P_{\gamma_{l_n}}\left(\mathcal{Z}_{l_n}^{b*,\rho}(p,c,\theta_{l_n}) \ge 0 \ge -\mathcal{Z}_{l_n}^{b*,\rho}(-p,c,\theta_{l_n})\right) \ge P_{\gamma_{l_n}}\left(\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(p,c,\theta_{l_n}) \ge 0 \ge -\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(-p,c,\theta_{l_n})\right) - \epsilon\ , \quad \text{(B.86)}$$

for all $n$ sufficiently large. Similarly, reversing the roles of $\mathcal{Z}_n^{b*,\rho}(p,c,\theta_{l_n})$ and $\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(p,c,\theta_{l_n})$, we obtain

$$\lim_{n\to\infty} \left| P_{\gamma_{l_n}}\left(\mathcal{Z}_{l_n}^{b*,\rho}(p,c,\theta_{l_n}) \ge 0 \ge -\mathcal{Z}_{l_n}^{b*,\rho}(-p,c,\theta_{l_n})\right) - P_{\gamma_{l_n}}\left(\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(p,c,\theta_{l_n}) \ge 0 \ge -\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(-p,c,\theta_{l_n})\right)\right| = 0\ . \quad \text{(B.87)}$$

The conclusion of the lemma then follows from (B.80)-(B.81), $G_n(x) \equiv P_{\gamma_{l_n}}(\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(p,x,\theta_{l_n}) \ge 0 \ge -\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(-p,x,\theta_{l_n}))$ being continuous (uniformly in $n$) and strictly decreasing by Assumption 3.2 (iv) and arguing similarly to Lemma 1.2.1 in Politis, Romano, and Wolf (1999). □

Proof of Lemma B.10: (i) To establish the first result, observe that for given $\theta$ the event

$$\max_{j=1,\dots,J}\{\mathbb{G}_{P,j}(\theta)\} \le c \quad \text{(B.88)}$$

implies the event

$$\left\{\sup_{\lambda \in \rho B_d} \langle p, \lambda\rangle : \max_{j=1,\dots,J}\left\{\mathbb{G}_{P,j}(\theta) + D_{P,j}(\theta)'\lambda + \tilde{\zeta}_{P,j,n}(\theta)\right\} \le c\right\} \ge 0. \quad \text{(B.89)}$$

This is so because if $\max_{j=1,\dots,J}\{\mathbb{G}_{P,j}(\theta)\} \le c$, then due to $\tilde{\zeta}_{P,j,n}(\theta) \le 0$, $\lambda = 0$ is feasible in the outer maximization problem in (B.89), hence the value of (B.89) is greater than or equal to $\langle p, 0\rangle = 0$. In turn this yields the desired result for

$$\begin{aligned} M &= z_{1-\alpha/J} \\ &\ge \sup_{P\in\mathcal{P}} \sup_{\theta\in\Theta}\left\{\inf\left\{c : P\left(\max_{j=1,\dots,J}\{\mathbb{G}_{P,j}(\theta)\} \le c\right) \ge 1-\alpha\right\}\right\}, \end{aligned} \quad \text{(B.90)}$$

where $z_\tau$ is the $\tau$ quantile of the standard normal distribution, and the second line follows from Bonferroni's inequality applied to the maximum order statistic.

(ii) The proof is similar to that of Lemma B.9. Let Let $(\theta_n, P_{\gamma_n}) \in \{(\theta, P), P \in \mathcal{P}, \theta \in \Upsilon(P)\}$ be a sequence such that

$$\limsup_{n\to\infty} P_{\gamma_n}\Big(\sup_{\theta'\in(\theta_n+n^{-1/2}\rho_n B_d)\cap\Upsilon(P_{\gamma_n})}|c_n(\theta_n)-c_n(\theta')|>\epsilon\Big)$$
$$=\limsup_{n\to\infty}\sup_{P\in\mathcal{P}}\sup_{\theta\in\Upsilon(P)} P\Big(\sup_{\theta'\in(\theta+n^{-1/2}\rho_n B_d)\cap\Upsilon(P)}|c_n(\theta)-c_n(\theta')|>\epsilon\Big). \tag{B.91}$$

Let $\{a_n\}$ be a subsequence of $\{n\}$ such that

$$\lim_{n\to\infty} P_{\gamma_{a_n}}\left(\sup_{\theta'\in(\theta_{a_n}+a_n^{-1/2}\rho_{a_n} B_d)\cap\Upsilon(P_{\gamma_{a_n}})}|c_{a_n}(\theta_{a_n})-c_{a_n}(\theta')|>\epsilon\right)$$
$$=\limsup_{n\to\infty} P_{\gamma_n}\Big(\sup_{\theta'\in(\theta_n+n^{-1/2}\rho_n B_d)\cap\Upsilon(P_{\gamma_n})}|c_n(\theta_n)-c_n(\theta')|>\epsilon\Big). \tag{B.92}$$

Arguing as in (B.82) and passing to a further subsequence $\{l_n\}$, one has from Lemma B.11 a non-empty set of $\mathcal{J}^*\subseteq\{1,\cdots,J\}$ such that

$$\kappa_{l_n}^{-1}h_{P_{\gamma_{l_n}},j,l_n}(\theta_{l_n})\to\pi_j\in(-\infty,0],\ \forall j\in\mathcal{J}^*. \tag{B.93}$$

This in turn implies that $\kappa_{l_n}^{-1}h_{P_{\gamma_{l_n}},j,l_n}(\theta')\to\pi_j$ uniformly in $\theta'\in(\theta_{l_n}+l_n^{-1/2}\rho B_d)\cap\Upsilon(P_{\gamma_{l_n}})$. This is because, by the mean value theorem, one has

$$\kappa_{l_n}^{-1}h_{P_{\gamma_{l_n}},j,l_n}(\theta')=\kappa_{l_n}^{-1}h_{P_{\gamma_{l_n}},j,l_n}(\theta_{l_n})+\kappa_{l_n}^{-1}D_{P_{l_n},j}(\bar\theta_{l_n})'(\theta'-\theta_{l_n}), \tag{B.94}$$

for some mean value $\bar\theta_{l_n}$ between $\theta_{l_n}$ and $\theta'$, and $\kappa_{l_n}^{-1}D_{P_{l_n},j}(\bar\theta_{l_n})'(\theta'-\theta_{l_n})=o(1)$ due to Assumption 3.3 (i), $\|\theta'-\theta_{l_n}\|=O(1)$. Define

$$A_{l_n}\equiv\{\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(p,c,\theta_{l_n})=U(\mathbb{G}_{P_{l_n}}^{\mathcal{J}^*}(\theta_{l_n}),D_{P_{l_n}}^{\mathcal{J}^*}(\theta_{l_n}),\tilde\zeta_{P_{l_n},l_n}^{\mathcal{J}^*}(\theta_{l_n}),c,\rho))\}. \tag{B.95}$$

An argument similar to the one in the proof of Lemma B.6 then ensures $P_{\gamma_{l_n}}(A_{l_n})\to 1$. Arguing similarly to (B.85)-(B.87), one then obtains for any $\theta'_{l_n}\in(\theta_{l_n}+l_n^{-1/2}\rho B_d)\cap\Upsilon(P_{\gamma_{l_n}})$

$$\lim_{n\to\infty}\left|P_{\gamma_{l_n}}\left(\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(p,c,\theta_{l_n})\ge 0\ge-\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(-p,c,\theta_{l_n})\right)-P_{\gamma_{l_n}}\left(\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(p,c,\theta'_{l_n})\ge 0\ge-\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(-p,c,\theta'_{l_n})\right)\right|=0\ . \tag{B.96}$$

By Assumption 3.2 (iv) one may then show that $G_n(x)\equiv P_{\gamma_{l_n}}(\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(p,x,\theta_{l_n})\ge 0\ge-\mathcal{Z}_{l_n}^{\mathbb{G},\rho}(-p,x,\theta_{l_n}))$ is continuous (uniformly in $n$) and strictly decreasing. The conclusion of the Lemma then follows from (B.96) and arguing as in Lemma 1.2.1 in Politis, Romano, and Wolf (1999).

Proof of Lemma B.11: When $h_n^\Delta>\kappa_n^{1/5}$, $\theta_n\in H(p,\Theta_I(P_n))$ for all $n$ implying there exists a $j\in\{1,\cdots,J\}$ and a sub-sequence such that $h_{P,j,m_n}(\theta_{m_n})=0$ for all $n$.

When $h_n^\Delta\le\kappa_n^{1/5}$, $\theta_n\in\Theta_I(P_n)$. For $\theta_n\in\partial\Theta_I(P_n)$, the previous argument applies. Suppose $\theta_n\in\Theta_I^o$, which in turn implies that there are no moment equalities. Let $\tilde\theta_n\equiv\theta_n+\alpha p\in\partial\Theta_I(P_n)$ denote the closest point to $\theta_n$ in direction $p$ on the boundary of $\Theta_I(P_n)$. By a mean value expansion,

$$\frac{E(m_j(\theta_n))}{\sigma_j(\theta_n)}=\frac{E(m_j(\tilde\theta_n))}{\sigma_j(\tilde\theta_n)}+D_j(\bar\theta_n)(\tilde\theta_n-\theta_n),$$

for $\bar\theta_n$ a value between $\theta_n$ and $\tilde\theta_n$. Because $\tilde\theta_n\in\partial\Theta_I(P_n)$, it follows that there exists at least one $j\in\{1,\dots,J\}$ such that $E(m_j(\tilde\theta_n))/\sigma_j(\tilde\theta_n)=0$, and therefore for that same $j$,

$$\left|\frac{E(m_j(\theta_n))}{\sigma_j(\theta_n)}\right|\le\|D_j(\bar\theta_n)\|\|\tilde\theta_n-\theta_n\|=O(\kappa_n^{1/5}n^{-1/2}),$$

where the last equality follows from the assumption that $h_n^\Delta\le\kappa_n^{1/5}$, the fact that $D_j$ is Lipschitz in $\theta$, and the fact that $D_j(\tilde\theta_n)$ is uniformly bounded. It then follows that for this same $j$, $\kappa_n^{-1}h_{P,j,n}(\theta)\to 0$, establishing the claim.

□

# C Uniform continuity of the linear program value and uniform bound on Lagrange Multipliers

For each $(D, e, \rho) \in \mathbb{R}^{J\times d} \times \mathbb{R}^J \times \mathbb{R}_+$, define the value function $V(D, e, \rho)$ of the following linear programming problem $LP(D, e, \rho)$:

$$V(D,e,\rho) \equiv \sup_{\lambda\in\mathbb{R}^d} \langle p, \lambda\rangle$$

$$s.t. \quad D\lambda \le e, \tag{C.1}$$

$$\lambda \in \rho B^d. \tag{C.2}$$

Let $\mathcal{S}(D, e, \rho)$ and $\mathcal{M}(D, e, \rho)$ be the solution set and the set of Lagrange multipliers (solutions to the dual problem) to $LP(D, e, \rho)$ respectively. Let the Lagrangian be defined by

$$L(\lambda, D, e, \rho) \equiv -\langle p, \lambda\rangle + \sum_{j=1}^{J} \mu_j(D'_j\lambda - e_j) + \sum_{k=1}^{d} \mu_{J+k}(\iota'_k\lambda - \rho) + \sum_{k=1}^{d} \mu_{J+d+k}(-\rho - \iota'_k\lambda), \tag{C.3}$$

where $\iota_k$ is a $d$-dimensional vector whose $k$-th component is 1 and other components are 0s.

Below, let $w \equiv (w_D, w_e) \in \mathbb{R}^{Jd+J}$ denote a direction of deviation of $(D', e')$ from $(D, e)$. For any $t \ge 0$ and $w \in \mathbb{R}^{Jd+J}$, let $\Phi_w(t) \equiv \{\lambda \in \rho B_d : (D + tw_D)\lambda \le (e + tw_e)\}$ be the feasibility set under a small perturbation toward $w$. We verify the key assumptions of Lemma C.1 for the linear programs that we use, in Appendix C.1 and C.2.

Lemma C.1: *Let $\mathcal{T}$ be the set of $(D, e)$s that satisfies the following conditions. There exists a positive constants $M > 0$ such that (a) Slater's condition holds; (b) $\mu_j \le M$ for all $j = 1, \cdots, J$, $\mu \in \mathcal{M}(D, e, \rho)$.*

*Suppose $(D, e) \in \mathcal{T}$ and $(D', e') \in \mathbb{R}^{Jd+J}$ is such that $\Phi_w(t) \ne \emptyset$ and $(D + tw_D, e + tw_e) \in \mathcal{T}$ for all $t$ sufficiently small with $w = (D' - D, e' - e)$. Then*

$$|V(D', e', \rho) - V(D, e, \rho)| \le M\|D' - D\|\rho + M\|e' - e\| \tag{C.4}$$

*Proof.* We first verify conditions (i)-(iv) of Theorem 4.4 in Bonnans and Shapiro (1998) (BS henceforth). Note that $LP(D, e, \rho)$ is convex. $\mathcal{S}(D, e, \rho)$ is non-empty and compact by $(D, e) \in \mathcal{T}$ and $\rho B_d$ being compact. The directional regularity condition in BS is satisfied because Slater's condition holds. By the argument in the proof of Theorem 3.2 in Shapiro (1995), condition (iv) of Theorem 4.4 in BS is satisfied if Slater's condition holds and if there exist $\bar\alpha < V(D, e, \rho)$, $t^* > 0$, and a compact set $S$ such that

$$\{\lambda : \langle p, \lambda\rangle \ge \bar\alpha, \lambda \in \Phi_w(t)\} \subset S\ , \tag{C.5}$$

for all $t \in [0, t^*]$. By our assumption on $(D', e')$, the above condition is satisfied with $S = \rho B_d$ and $\bar\alpha = -d^{1/2}\rho$ for some $t^* > 0$. Hence, by Theorem 4.4 in BS and their subsequent remark, $V(D, e, \rho)$ is Hadamard directionally differentiable with the directional derivative

$$V'(D,e,\rho)[w] = \inf_{\lambda\in\mathcal{S}(D,e,\rho)} \sup_{\mu\in\mathcal{M}(D,e,\rho)} \nabla_{(D,e)}L(D,e,\rho)'w = \inf_{\lambda\in\mathcal{S}(D,e,\rho)} \sup_{\mu\in\mathcal{M}(D,e,\rho)} \sum_{j=1}^{J} \mu_j(w'_{D_j}\lambda - w_{e_j})\ . \tag{C.6}$$

By the Hadamard directional differentiability of $V$ and $(D, e)$ being in a finite dimensional space, it follows that for any $t_0 > 0$,

$$\inf_{t\in[0,t_0]} V'(D+tw_D, e+tw_e, \rho)[w] \le V(D+t_0w_D, e+t_0w_e,\rho) - V(D,e,\rho)$$
$$\le \sup_{t\in[0,t_0]} V'(D+tw_D, e+tw_e,\rho)[w]\ , \quad (C.7)$$

(see e.g. Proposition 6 in Demyanov, 2009). Letting $w_D = D' - D$ and $w_e = e' - e$, (C.6) and (C.7) imply

$$\begin{aligned}|V(D',e',\rho)-&V(D,e,\rho)| \\ &\le \sup_{t\in[0,1]}\Big|\inf_{\lambda\in\mathcal{S}(D+tw_D,e+tw_e,\rho)}\ \sup_{\mu\in\mathcal{M}(D+tw_D,e+tw_e,\rho)}\sum_{j=1}^J \mu_j(w'_{D_j}\lambda - w_{e_j})\Big| \\ &\le M\sum_{j=1}^J\|w_{D_j}\|\sup_{\lambda\in\rho B_d}\|\lambda\| + M\sum_{j=1}^J|w_{e_j}| \\ &\le M\|w_D\|\rho + M\|w_e\|, \end{aligned} \tag{C.8}$$

where the second inequality follows from $S(D+tw_D, e+tw_e,\rho)\subset \rho B_d, \forall t\in[0,1]$ and Assumption (b), and the last inequality used $\|x\|_2 \le \|x\|_1$ for any $x\in\mathbb{R}^J$. This establishes the claim of the lemma. □

## C.1 Slater condition

Let $D_{J_1}$ be a $J_1\times d$ matrix whose rows are transposed gradients $D'_j$ for $j = 1,\cdots,J_1$. Let $D_{J_2}$ be similarly defined. Below we use the matrix norm $\|A\|_{op}\equiv \sup_{\|x\|=1}\|Ax\|$.

Let $Q$ denote the law of $g \equiv (g'_{J_1}, g'_{J_2})'$, a random vector in $\in\mathbb{R}^{J_1+J_2}$. Let $\mathcal{Q}$ be the set of distributions under consideration. Let $e = (g_{J_1} - h_{J_1} + c, g_{J_2} + c, -g_{J_2} + c)$ where $h_{J_1}\le 0$ and $c \ge 0$. Let $\rho > 0$ be given.

$LP(D,e,\rho)$ is said to satisfy the Slater condition if (i) $c = 0$ and there exists $\lambda\in\rho B^o_d$ such that $D_{J_1}\lambda < e_{J_1}$ and $D_{J_2}\lambda = e_{J_2}$ or (ii) $c > 0$ and there exists $\lambda \in \rho B^o_d$ such that $D\lambda < e$. Below, we let $L$ be the event that $LP(D,e,\rho)$ satisfies the Slater condition.

Lemma C.2: *For any $\eta > 0$, there exists $M_\eta > 0$ such that*

$$P\big(L\cap\{\sup_{\mu\in\mathcal{M}(D,e,\rho)}\|\mu\|\le M_{\eta/2}\}\cap\{\mathcal{S}(D,e,\rho)\ne\emptyset\}\big)\ge 1-\eta \tag{C.9}$$

*uniformly in $Q\in\mathcal{Q}$.*

*Proof.* First, we show the result for $c = 0$. In this setting, $LP(D,e,\rho)$ involves $J_1$ affine inequalities, $J_2$ affine equalities, and the constraint $\lambda\in\rho B^d$. Below, I assume that

$$\sup_{\mu\in\mathcal{M}(D,e,\rho)}\|\mu\|\le M_{\eta/2},\ \mathcal{S}(D,e,\rho)\ne\emptyset. \tag{C.10}$$

By Theorem C.1, this event has at least $1-\eta/2$ probability under any $Q\in\mathcal{Q}$. Take any pair $(\lambda^*,\mu)\in\mathcal{S}(D,e,\rho)\times\mathcal{M}(D,e,\rho)$. Let $C(\lambda^*)\equiv\{j=1,\cdots,J_1+J_2+2d : K_j\lambda - e_j = 0\}$, $\mathcal{J}_1(\lambda^*)\equiv\{j=1,\cdots,J_1 : D'_j\lambda^* = e_j\}$, $\mathcal{J}_d(\lambda^*)\equiv\{k=1,\cdots,d:\lambda^*_k=\rho\}$, $\mathcal{J}_{-d}(\lambda^*)\equiv\{k=1,\cdots,d : -\lambda^*_k=\rho\}$. Denote the cardinality of each of these sets by $J_1(\lambda^*)$, $J_d(\lambda^*)$, and $J_{-d}(\lambda^*)$ respectively. Let $0<\rho'<\rho$ and define the following linear programming problem:

$$(D)\quad \min_{\mu\in\mathbb{R}^{J_1}_+\times\mathbb{R}^{J_2}\times\mathbb{R}^{2d}_+} - C_{\eta/2}\sum_{j\in\mathcal{J}_1(\lambda^*)}\mu_j - \sum_{k=1}^{J_2} e_{J_1+k}\mu_{J_1+k} + \rho'\sum_{l\in\mathcal{J}_d(\lambda^*)}\mu_{J_1+J_2+l} + \rho'\sum_{l\in\mathcal{J}_{-d}(\lambda^*)}\mu_{J_1+J_2+d+l}$$
$$s.t. \sum_{j=1}^{J_1}\mu_j D_j + \sum_{k=1}^{J_2}\mu_{J_1+k}D_{J_1+k} + \sum_{l=1}^{2d}(\mu_{J_1+J_2+l} - \mu_{J_1+J_2+d+l})\iota_l = p$$
$$\mu_j = 0, \forall j\notin C(\lambda^*).$$

By Theorem 28.3 in Rockafellar (1970), any $\mu \in \mathcal{M}(D,e,\rho)$ satisfies the KKT condition. Therefore, $\mu \in \mathcal{M}(D,e,\rho)$ is a feasible solution to $(D)$. By Lemma C.4 and (C.10), the optimal value achieved in $(D)$ is also uniformly bounded (with probability at least $1-\eta/2$). Note that $(D)$ is a dual problem to the following primal problem:

$$
\begin{aligned}
(P) \quad & \sup \langle p, \lambda \rangle \\
& s.t. D_j' \lambda \le -C_{\eta/2}, \ j \in \mathcal{J}_1(\lambda^*) \\
& \quad D_j' \lambda \le 0, \ j \in \{1, \cdots, J_1\} \setminus \mathcal{J}_1(\lambda^*) \\
& \quad D_j' \lambda = e_j, \ j = J_1+1, \cdots, J_2 \\
& \quad \lambda_k \le \rho', \ k \in \mathcal{J}_d(\lambda^*) \\
& \quad -\lambda_k \le \rho', \ k \in \mathcal{J}_{-d}(\lambda^*) \\
& \quad \lambda_k = 0, \ k \notin \mathcal{J}_d(\lambda^*) \cup \mathcal{J}_{-d}(\lambda^*).
\end{aligned}
$$

Since both $(P)$ and $(D)$ are LPs and $(D)$ is feasible, the strong duality holds (see e.g. Boyd and Vandenberghe, 2004, p.227). Since $(D)$ is feasible and uniformly bounded, $(P)$ is feasible, which in turn ensures that there exists a feasible solution $\tilde{\lambda}$ to $(P)$. Consider a convex combination $\lambda_\alpha \equiv \alpha\tilde{\lambda} + (1-\alpha)\lambda^*$ for some $\alpha \in (0,1)$. Then, for any $j \in \mathcal{J}_1(\lambda^*)$, $C_{\eta/2} + e_j > 0$ implies

$$D_j'\lambda_\alpha = \alpha D_j'\tilde{\lambda} + (1-\alpha)D_j'\lambda^* \le -\alpha C_{\eta/2} + (1-\alpha)e_j = e_j - \alpha(C_{\eta/2} + e_j) < e_j. \tag{C.11}$$

Furthermore, for any $j \in \{1, \cdots, J_1\} \setminus \mathcal{J}_1(\lambda^*)$, one has

$$D_j'\lambda_\alpha = \alpha D_j'\tilde{\lambda} + (1-\alpha)D_j'\lambda^* \le 0 + (1-\alpha)D_j'\lambda^* < (1-\alpha)e_j < e_j, \tag{C.12}$$

where the first weak inequality follows from $D_j'\tilde{\lambda} \le 0$ for any $j \in \{1, \cdots, J_1\} \setminus \mathcal{J}_1(\lambda^*)$ by $\tilde{\lambda}$ being a feasible solution to $(P)$, and the second inequality follows from $D_j'\lambda^* < e_j$ because of the $j$-th inequality being slack at $\lambda^*$.

Next, since $D_{J_2}\tilde{\lambda} = D_{J_2}\lambda^* = e_{J_2}$, it follows that

$$D_{J_2}\lambda_\alpha = \alpha D_{J_2}\tilde{\lambda} + (1-\alpha)D_{J_2}\lambda^* = e_{J_2}. \tag{C.13}$$

Further, the $k$-th component of $\lambda_\alpha$ satisfies

$$\lambda_{\alpha,k} = \alpha\tilde{\lambda}_k + (1-\alpha)\lambda_k^* \le \alpha\rho' + (1-\alpha)\rho < \rho, \ \forall k \in \mathcal{J}_d(\lambda^*) \tag{C.14}$$

$$\lambda_{\alpha,k} = \alpha\tilde{\lambda}_k + (1-\alpha)\lambda_k^* \ge -\alpha\rho' - (1-\alpha)\rho > -\rho, \ \forall k \in \mathcal{J}_{-d}(\lambda^*) \tag{C.15}$$

$$\lambda_{\alpha,k} = (1-\alpha)\lambda_k^* \in (-\rho, \rho), \ k \notin \mathcal{J}_d(\lambda^*) \cup \mathcal{J}_{-d}(\lambda^*) \tag{C.16}$$

Note that for any $j \in \{1, \cdots, J_1\}$, $C_{\eta/2} + e_j = C_{\eta/2} + g_j - h_j > 0$ holds with probability at least $1 - \eta/2$ by $h_j \le 0$ and Lemma C.4. Therefore, (C.11)-(C.16) hold with probability at least $1 - \eta/2$. Combine this with the probability of (C.10), which is also $1 - \eta/2$. Then, the claim of the lemma follows for the case in which $c = 0$.

We now consider the case in which $c > 0$. The argument is similar to the one used above. A key difference is that the affine equalities in the previous case are treated as two opposing inequalities in the current case. Again, take any pair $(\lambda^*, \mu) \in \mathcal{S}(D,e,\rho) \times \mathcal{M}(D,e,\rho)$. Let $\mathcal{J}_2(\lambda^*) \equiv \{k = 1, \cdots, J_2 : D_{J_1+k} = e_{J_1+k}\}$ and $\mathcal{J}_{-2}(\lambda^*) \equiv \{k = 1, \cdots, J_2 : D_{J_1+J_2+k} = e_{J_1+J_2+k}\}$ be the set of binding constraints among the inequality constraints that are generated from the equality constraints. Define

$$
\begin{aligned}
(D) \quad \min_{\mu\in\mathbb{R}_+^{J_1}\times\mathbb{R}^{J_2}\times\mathbb{R}_+^{2d}} & -C_{\eta/2}\sum_{j\in\mathcal{J}_1(\lambda^*)}\mu_j - C_{\eta/2}\sum_{k\in\mathcal{J}_2(\lambda^*)}\mu_{J_1+k} - C_{\eta/2}\sum_{k\in\mathcal{J}_{-2}(\lambda^*)}\mu_{J_1+J_2+k}\\
& +\rho'\sum_{l\in\mathcal{J}_d(\lambda^*)}\mu_{J_1+J_2+l} + \rho'\sum_{l\in\mathcal{J}_{-d}(\lambda^*)}\mu_{J_1+J_2+d+l}\\
s.t. & \sum_{j=1}^{J_1}\mu_j D_j + \sum_{k=1}^{J_2}\mu_{J_1+k}D_{J_1+k} + \sum_{l=1}^{2d}(\mu_{J_1+J_2+l}-\mu_{J_1+J_2+d+l})\iota_l = p\\
& \mu_j = 0, \forall j\notin C(\lambda^*).
\end{aligned}
$$

This is a dual problem to the following primal problem:

$$
\begin{aligned}
(P) \quad & \sup\langle p,\lambda\rangle\\
& s.t. D_j'\lambda \le -C_{\eta/2},\ j\in\mathcal{J}_1(\lambda^*)\cup\mathcal{J}_2(\lambda^*)\cup\mathcal{J}_{-2}(\lambda^*)\\
& D_j'\lambda\le 0,\ j\in\{1,\cdots,J\}\setminus(\mathcal{J}_1(\lambda^*)\cup\mathcal{J}_2(\lambda^*)\cup\mathcal{J}_{-2}(\lambda^*))\\
& \lambda_k\le\rho',\ k\in\mathcal{J}_d(\lambda^*)\\
& -\lambda_k\le\rho',\ k\in\mathcal{J}_{-d}(\lambda^*)\\
& \lambda_k = 0,\ k\notin\mathcal{J}_d(\lambda^*)\cup\mathcal{J}_{-d}(\lambda^*).
\end{aligned}
$$

Arguing as in the previous case, there is a feasible solution $\tilde{\lambda}$ to $(P)$. Define $\lambda_\alpha\equiv\alpha\tilde{\lambda}+(1-\alpha)\lambda^*$ for some $\alpha\in(0,1)$. Then, arguing as in (C.11), one has

$$D_j'\lambda_\alpha\le e_j-\alpha(C_{\eta/2}+e_j)<e_j,\ \forall j\in\mathcal{J}_1(\lambda^*)\cup\mathcal{J}_2(\lambda^*)\cup\mathcal{J}_{-2}(\lambda^*), \tag{C.17}$$

provided that $C_{\eta/2}+e_j>0$. Note that $e_j\ge g_j$ for all $j=1,\cdots,J_1+J_2$. Hence, $C_{\eta/2}+e_j>0$ occurs with probability at least $1-\eta$ by Lemma C.4. Furthermore, arguing as in (C.12), for any $j\in\{1,\cdots,J\}\setminus(\mathcal{J}_1(\lambda^*)\cup\mathcal{J}_2(\lambda^*)\cup\mathcal{J}_{-2}(\lambda^*))$, one has

$$D_j'\lambda_\alpha=\alpha D_j'\tilde{\lambda}+(1-\alpha)D_j'\lambda^*\le 0+(1-\alpha)D_j'\lambda^*<(1-\alpha)e_j<e_j. \tag{C.18}$$

In addition, (C.14)-(C.16) hold by the same argument as before. Therefore, (C.14)-(C.18) hold with probability at least $1-\eta/2$. Combine this with the probability of (C.10), which is also $1-\eta/2$. Then, the claim of the lemma follows for the case in which $c>0$. □

## C.2 Uniform bound on Lagrange Multipliers

THEOREM C.1: *Suppose Assumptions 3.1, 3.2, 3.3 hold and that $\Lambda(K_P,g_P,c_n)\neq\emptyset$, with this set defined in equation A.9. Then for any $\eta>0$ there exists a $M_\eta<\infty$ such that*

$$\sup_{P\in\mathcal{P}}P\left(\sup_{\mu\in\mathcal{M}(D_P,e_P,\rho)}\|\mu\|>M_\eta,\Lambda(K_P,g_P,c_n)\neq\emptyset\right)<\eta. \tag{C.19}$$

THEOREM C.2: *Suppose Assumptions 3.1, 3.2, 3.3, 3.4 hold and that $\Lambda(K_P,g_P,c_n)\neq\emptyset$, with this set defined in equation A.9. Then for any $\eta>0$ there exists a $M_\eta<\infty$ and $N\in\mathbb{N}$ such that*

$$\sup_{P\in\mathcal{P}}P\left(\sup_{\mu\in\mathcal{M}(D_n,e_n,\rho)}\|\mu\|>M_\eta,\Lambda(K_n,g_n,c_n)\neq\emptyset\right)<\eta, \tag{C.20}$$

*for all $n\ge N$.*

THEOREM C.3: *Suppose Assumptions 3.1, 3.2, 3.3, 3.4 hold. Then for any $\bar{\eta}>0$ there exists a $N\in\mathbb{N}$ such that*

$$\sup_{P\in\mathcal{P}}P(\Lambda(K_P,g_P,c_n)\neq\emptyset,\Lambda(K_n,g_n,c_n)=\emptyset)<\bar{\eta}, \tag{C.21}$$

*for all $n \geq N$.*

***Proof of Theorem C.1.*** Let $\mathcal{B}(D_P, e_P, \rho)$ denote the set of basic solutions,

$$\mathcal{B}(D_P, e_P, \rho) \equiv \{\lambda^C \in \mathbb{R}^d : \lambda^C = (K_P^{C\prime})^{-1} g^C, |det(K_P^C)| > 0, \exists C \subset \{1, \ldots, J+2d\}, |C| = d\}.$$

Note that the cardinality of the set $\mathcal{B}(D_P, e_P, \rho)$ is finite. Define the collection of constraint indexes that form a basic optimal solution by

$$\mathbb{C}(D_P, e_P, \rho) \equiv \{C \subset \{1, \ldots, J+2d\} : \lambda^C \in \mathcal{B}(D_P, e_P, \rho) \cap \mathcal{S}(D_P, e_P, \rho)\}.$$

The set $\mathcal{M}(D_P, e_n, \rho)$ is a random closed set in $\mathbb{R}^{J+2d}$, and the set $\mathbb{C}(D_P, e_P, \rho)$ is a random closed set taking its values in the subsets of $\{1, \ldots, J+2d\}$, see Molchanov (2005, Definition 1.1.1). Let

$$\mathbb{M}(D_P, e_n, \rho) \equiv \cup_{C \in \mathbb{C}(D_P, e_P, \rho)} \{\mu^C : \mu^C = (K_P^{C\prime})^{-1} p\}.$$

In general, $\mathbb{M}(D_P, e_n, \rho) \subseteq \mathcal{M}(D_P, e_n, \rho)$. However, we argue that for all $\epsilon > 0$,

$$\sup_{P \in \mathcal{P}} P\left( \left| \max_{\mu \in \mathbb{M}(D_P, e_n, \rho)} \|\mu\| - \sup_{\mu \in \mathcal{M}(D_P, e_P, \rho)} \|\mu\| \right| > \epsilon \cap \Lambda(K_P, g_P, c_n) \neq \emptyset \right) < \epsilon. \tag{C.22}$$

To see why this is the case, observe that each $\mu \in \mathcal{M}(D_P, e_P, \rho)$ solves the equation $p = K_P' \mu$, and $\mu$ can have less than $d$ non-zero entries, exactly $d$ non-zero entries, and more than $d$ non-zero entries.

The first case corresponds to an optimal solution with less than $d$ active inequalities. Let $A$ denote the set of active inequalities at this optimal solution, with $|A|$ denoting its cardinality. Then $p = K_P^{A\prime} \mu^A$. Consider now an optimal basic solution $C \in \mathbb{C}$ with $A \subset C$. Then for that basic solution, $p = K_P^{C\prime} \mu^C$ and $\mu^C$ is uniquely determined by this equality because $K_P^{C\prime}$ is invertible by Lemma C.6. Without loss of generality, assume that $K_P^A$ corresponds to the first $|A|$ rows in $K_P^C$. It then follows that $\mu^C$ equals $\mu^A$ followed by $d - |A|$ zeros, and therefore $\|\mu^A\| = \|\mu^C\|$.

The case in which $\mu \in \mathcal{M}(D_P, e_P, \rho)$ has more than $d$ non-zero entries corresponds to the case that there are more than $d$ active inequalities satisfied at a given solution to the linear system. By Lemma C.7, the probability of this event can be made uniformly smaller than $\epsilon$.

Finally, we consider the case that $\mu \in \mathcal{M}(D_P, e_P, \rho)$ has exactly $d$ non-zero entries. Denote the indexes of the non-zero entries of $\mu$ by $C$. We now argue that $|det(K_P^C)| > 0$. Suppose by contradiction that $det(K_P^C) = 0$. By Lemma C.6, the probability that $det(K_P^C) = 0$ and there exists a $\lambda \in \mathbb{R}^d$ such that $K_P^C \lambda = g_P^C$, is uniformly equal to zero. This in turn implies that $\mu \notin \mathcal{M}(D_P, e_P, \rho)$. Hence (C.22) holds.

Fix an arbitrary collection $C \subset \{1, \ldots, J+2d\}$ of cardinality $d$ such that $|det(K_P^C)| > 0$ . Let $\mu^C = (K_P^{C\prime})^{-1} p$ and let $\lambda^C = (K_P^{C\prime})^{-1} g^C$. By definition,

$$p' \lambda^C = p' (K_P^C)^{-1} g^C = \mu^{C\prime} g^C.$$

Remark that for given $K_P$, $\mu^C$ is non-stochastic. Suppose $\|\mu^C\| > M_\eta$. Then

$$\begin{aligned} \inf_{P \in \mathcal{P}} P(\{C \notin \mathbb{C}\} \cap \Lambda(K_P, g_P, c_n) \neq \emptyset) &\geq \inf_{P \in \mathcal{P}} P(\{p' \lambda^C \in \rho B_d\} \cap \Lambda(K_P, g_P, c_n) \neq \emptyset) \\ &\geq \inf_{P \in \mathcal{P}} P(\{p' \lambda^C \in \rho B_d\} \cap \Lambda(K_P, g_P, c_n) \neq \emptyset \cap |det(K_P^C)| > 0) \end{aligned}$$

where the first inequality follows because if $\lambda^C$ is outside the $\rho$ box, it cannot be optimal. The result then follows from Lemma C.5, observing that $\mathbb{C}$ has finite cardinality.

□

***Proof of Theorem C.2.*** Let $\mathbb{C}(D_n, e_n, \rho)$ and $\mathcal{M}(D_n, e_n, \rho)$ be defined analogously to $\mathbb{C}(D_P, e_P, \rho)$ and $\mathcal{M}(D_P, e_P, \rho)$. By Lemma C.8, we have that the analogs of Lemmas C.5, C.6, and C.7 hold when $K_P$ and $g_P$ are replaced by $K_n$ and $g_n$. Hence the result follows. □

***Proof of Theorem C.3.*** For any index set $C$ such that $|C| = d$ and $K_P^C \lambda^C = g_P^C$, Lemma C.6 yields that

$$\sup_{P \in \mathcal{P}} P(\{\lambda^C \in \Lambda(K_P, g_P, c_n)\} \cap det(K_P^C) = 0) = 0,$$

and for any $\eta > 0$ there exists $\bar{\alpha}_\eta$ such that

$$\sup_{P \in \mathcal{P}} P(0 < |\alpha_d^C| < \bar{\alpha}_\eta \cap \{\lambda^C \in \rho B^d\}) < \eta/4,$$

where $\lambda^C = (K_P^C)^{-1} g_P^C$ and $\alpha_d^C$ denotes the smallest eigenvalue of $K_P^{C\prime} K_P^C$. Recall that

$$|det(K_P^C)| \geq \left(\sqrt{\alpha_d^C}\right)^d.$$

If $|det(K_n^C)| > 0$, let $\lambda_n^C = (K_n^C)^{-1} g_n^C$. Because $K_n \to^p K_P$ uniformly, and the determinant is a continuous function of its argument, we have that for any $\eta > 0$ there exists a $N \in \mathbb{N}$ such that

$$\sup_{P \in \mathcal{P}} P(|det(K_P^C)| \geq \left(\sqrt{\bar{\alpha}_\eta}\right)^d \cap |det(K_n^C)| = 0) < \eta/4$$

for all $n \geq N$. Let $\tilde{\mathbb{C}} = \{C \in \{1, \ldots, J + 2d\} : |C| = d\}$. It then follows that

$$\begin{aligned}
&\sup_{P \in \mathcal{P}} P(\Lambda(K_P, g_P, c_n) \neq \emptyset, \Lambda(K_n, g_n, c_n) = \emptyset) \\
&\leq \sup_{P \in \mathcal{P}} P(\cup_{C \in \tilde{\mathbb{C}}} \{\lambda^C \in \Lambda(K_P, g_P, c_n)\} \cap \{\Lambda(K_n, g_n, c_n) = \emptyset\}) \\
&\leq \sup_{P \in \mathcal{P}} P(\cup_{C \in \tilde{\mathbb{C}}} \{\lambda^C \in \Lambda(K_P, g_P, c_n)\} \cap |det(K_P^C)| > \left(\sqrt{\bar{\alpha}_\eta}\right)^d \cap \{\Lambda(K_n, g_n, c_n) = \emptyset\}) + \eta/4 \\
&\leq \sup_{P \in \mathcal{P}} P\Big( \cup_{C \in \tilde{\mathbb{C}}} \{\lambda^C \in \Lambda(K_P, g_P, c_n)\} \cap |det(K_P^C)| > \left(\sqrt{\bar{\alpha}_\eta}\right)^d \\
&\quad \cap |det(K_n^C)| > 0 \cap \{\Lambda(K_n, g_n, c_n) = \emptyset\}\Big) + 2\eta/4 \\
&\leq \sup_{P \in \mathcal{P}} P\Big( \cup_{C \in \tilde{\mathbb{C}}} \{\lambda^C \in \Lambda(K_P, g_P, c_n)\} \cap |det(K_P^C)| > \left(\sqrt{\bar{\alpha}_\eta}\right)^d \\
&\quad \cap |det(K_n^C)| > 0 \cap \{\lambda_n^C \notin \Lambda(K_n, g_n, c_n)\}\Big) + 2\eta/4 \\
&\leq \sup_{P \in \mathcal{P}} P\Big( \cup_{C \in \tilde{\mathbb{C}}} \{\lambda^C \in \Lambda(K_P, g_P, c_n)\} \cap |det(K_P^C)| > \left(\sqrt{\bar{\alpha}_\eta}\right)^d \\
&\quad \cap |det(K_n^C)| > 0 \cap \{\exists l \in \{1, \ldots, J + 2d\} : K_n^{[l]} \lambda_n^C > g_n^{[l]})\}) + 2\eta/4 \\
&\leq \sum_{C \in \tilde{\mathbb{C}}} \sup_{P \in \mathcal{P}} P\Big(\{\lambda^C \in \Lambda(K_P, g_P, c_n)\} \cap |det(K_P^C)| > \left(\sqrt{\bar{\alpha}_\eta}\right)^d \\
&\quad \cap |det(K_n^C)| > 0 \cap \{\exists l \in \{1, \ldots, J + 2d\} : K_n^{[l]} \lambda_n^C > g_n^{[l]})\}) + 2\eta/4.
\end{aligned}$$

Observe that whenever $\lambda^C \in \Lambda(K_P, g_P, c_n)$, we have that for all $l \in \{1, \ldots, J + 2d\}$, $K_P^{[l]} \lambda^C \leq g_P^{[l]}$. Using this fact and Lemma C.7, we have that for all $l \in \{1, \ldots, J + 2d\} : l \notin C$ and for all $\eta > 0$ there is an $\epsilon_\eta$ such that

$$\sup_{P \in \mathcal{P}} P\left(0 \geq K_P^{[l]} \left(K_P^{C\prime}\right)^{-1} g_P^C - g_P^{[l]} \geq -\epsilon_\eta \cap A_1\right) < \eta/4,$$

with $A_1$ denoting the event that $\left\{|det(K_P^C)| > (\sqrt{\bar{\alpha}_\eta})^d\right\}$. Next, observe that by assumption, uniformly over $\mathcal{P}$, $K_n \to^p K_P$ and $g_n \to^{a.s.} g_P$, so that for all $\eta > 0$ and the same $\epsilon_\eta$ there is $N \in \mathbb{N}$ such that

$$\sup_{P \in \mathcal{P}} P\left(\left\{\left|\left(K_P^{[l]} \left(K_P^{C\prime}\right)^{-1} g_P^C - g_P^{[l]}\right) - \left(K_n^{[l]} \left(K_n^{C\prime}\right)^{-1} g_n^C - g_n^{[l]}\right)\right| > \epsilon_\eta/2\right\} \cap A_1 \cap A_2\right) < \eta/4$$

for all $n \geq N$, and where $A_2$ denotes the event $|det(K_n^C)| > 0$. This in turn implies that

$$\begin{aligned}
&\sup_{P\in\mathcal{P}} P\Big(\{\lambda^C \in \Lambda(K_P, g_P, c_n)\} \cap |det(K_P^C)| > \left(\sqrt{\bar{\alpha}_\eta}\right)^d \\
&\cap |det(K_n^C)| > 0 \cap \{\exists l \in \{1,\dots,J+2d\} : K_n^{[l]}\lambda_n^C > g_n^{[l]})\}\Big) + 2\eta/4 \\
&\leq \sup_{P\in\mathcal{P}} P\Big(\{\lambda^C = (K_P^C)^{-1} \cap \{\forall\, l \in \{1,\dots,J+2d\} : l \notin C,\, K_P^{[l]}\left(K_P^{C\prime}\right)^{-1} g_P^C - g_P^{[l]} < -\epsilon_\eta\}\} \\
&\cap |det(K_P^C)| > \left(\sqrt{\bar{\alpha}_\eta}\right)^d \cap |det(K_n^C)| > 0 \cap \{\exists l \in \{1,\dots,J+2d\} : K_n^{[l]}\lambda_n^C > g_n^{[l]})\}\Big) + 3\eta/4 \\
&\leq \eta.
\end{aligned}$$

Observing that the cardinality of $\tilde{\mathbb{C}}$ and the cardinality of $\{1,\dots,J+2d\}$ are both finite, the result follows by choosing $\eta$ appropriately. □

## C.3 Auxiliary Lemmas

LEMMA C.3: *Suppose Assumption 3.1, 3.2, 3.3 hold. Let $C$ denote a subset of $\{1,\dots,J+2d\}$ of cardinality $d$. Suppose $C = C_1 \cup C_2$, with $C_1 \subset \{J+1,\dots,J+2d\}$ an index set of cardinality $d_1$ and $C_2 \subset \{1,\dots,J\}$ an index set of cardinality $d_2$ and $d_1 + d_2 = d$, $d_1 \geq 1, d_2 \geq 1$. Suppose $det(K_P^C) \neq 0$. Let $\mu^C = (K_P^{C\prime})^{-1}p$. Then if $\|\mu^C\| > M_\eta > 2$, it follows that $\|\mu^{[d_1+1:d]}\| > (M_\eta - 2)/2d\bar{M}$.*

*Proof.* Without loss of generality,[14] we can write

$$K_P^C = \begin{bmatrix} I_{d_1} & O_{d_1\times d_2} \\ D^{C_2}_{P,d_2\times d_1} & D^{C_2}_{P,d_2\times d_2} \end{bmatrix}, \quad K_P^{C\prime} = \begin{bmatrix} I_{d_1} & D^{C_2\prime}_{P,d_2\times d_1} \\ O_{d_2\times d_1} & D^{C_2\prime}_{P,d_2\times d_2} \end{bmatrix},$$

where $O_{d_2\times d_1}$ is a $d_2 \times d_1$ matrix of zeros, and $[D^{C_2}_{P,d_1\times d_2}\ D^{C_2}_{P,d_1\times d_1}] = D_P^{C_2}$. Let $p = [p_1;\ p_2]$, with $p_1$ the $d_1 \times 1$ vector with the first $d_1$ components of $p$, and $p_2$ the remaining $d_2 \times 1$ components. Using the algebra for matrix blockwise inversion we have that

$$\left(K_P^{C\prime}\right)^{-1} p = \begin{bmatrix} p_1 - D^{C_2\prime}_{P,d_2\times d_1}\left(D^{C_2\prime}_{P,d_2\times d_2}\right)^{-1} p_2 \\ \left(D^{C_2\prime}_{P,d_2\times d_2}\right)^{-1} p_2 \end{bmatrix}.$$

We first argue that $p_2 \neq O_{d_2\times 1}$. Suppose not. Then, recalling that $\|p\| = 1$, we have $\|\mu^C\| = 1$, contradicting the assumption.

Next observe that by triangle inequality and elementary operations,

$$M_\eta < \|\mu^C\| \leq 2\max\left\{\left\|p_1 - D^{C_2\prime}_{P,d_2\times d_1}\left(D^{C_2\prime}_{P,d_2\times d_2}\right)^{-1} p_2\right\|, \left\|\left(D^{C_2\prime}_{P,d_2\times d_2}\right)^{-1} p_2\right\|\right\},$$

which in turn implies

$$\frac{M_\eta}{2} < 1 + d\bar{M}\left\|\mu^{[d_1+1:d]}\right\|. \tag{C.23}$$

□

LEMMA C.4: *Under Assumption 3.2-(v), for any $\epsilon \geq 0$ and $r \in \mathbb{R}$, and for any vector $k \in \mathbb{R}^{J+2d}$ such that $\|k^{[1,\dots,J]}\| = 1$, it follows that*

$$\inf_{P\in\mathcal{P}} P(\|k' g_P - r\| > \epsilon) \geq 2\Phi(-\epsilon/\sqrt{\omega}). \tag{C.24}$$

[14] The use of the $I_{d_1}$ is without loss of generality in the sense that having one or more of its rows multiplied by $-1$ will not change the argument.

*Proof.* Let $N(a,b^2)$ denote a random variable distributed normal, with mean $a$ and variance $b^2$. Then:

$$\begin{aligned}
& \inf_{P\in\mathcal{P}} P\left(\left\|k'g_P - r\right\| > \epsilon\right) \\
= & \inf_{P\in\mathcal{P}} P\left(\left\|N\left(k'E_P(g) - r, (k^{[1:J]})'\Omega_P k^{[1:J]}\right)\right\| > \epsilon\right) \\
\geq & \inf_{P\in\mathcal{P}} P\left(\left\|N\left(0, (k^{[1:J]})'\Omega_P k^{[1:J]}\right)\right\| > \epsilon\right) \\
\geq & \inf_{P\in\mathcal{P}} P\left(\left\|N\left(0,\omega\right)\right\| > \varepsilon\right) \\
= & 2\Phi\left(-\epsilon/\sqrt{\omega}\right),
\end{aligned}$$

where the first inequality uses that for given $b^2$ and $\epsilon$, $P\left(\left\|N\left(a,b^2\right)\right\| > \epsilon\right)$ is minimized at $a=0$. □

Lemma C.5: *Suppose Assumptions 3.1, 3.2, 3.3 hold. Fix any $q : \|q\| = 1$. Let $\lambda^C$ be such that $K_P^C\lambda^C = g_P^C$, and let $\tilde{\mu}^C$ be such that $K_P^{C\prime}\tilde{\mu}^C = q$. For any $0 < \eta < 1$ there is a $M_\eta$ such that if $\|\tilde{\mu}^C\| > M_\eta$ then*

$$\inf_{P\in\mathcal{P}} P(\lambda^C \notin \rho B_d) > 1 - \eta. \tag{C.25}$$

*Proof.* By Lemma C.6-(i), $\lambda^C$ exists with positive probability only if $|det(K_P^C)| > 0$, and therefore

$$\inf_{P\in\mathcal{P}} P(\lambda^C \notin \rho B_d) \geq \inf_{P\in\mathcal{P}} P(\{\lambda^C \notin \rho B_d\} \cap |det(K_P^C)| > 0). \tag{C.26}$$

Initially let $C$ pick only stochastic constraints, in which case it is w.l.o.g. to set $C = \{1,...,d\}$. We then have

$$\begin{aligned}
& \inf_{P\in\mathcal{P}} P\left(\lambda^C \notin \rho B_d \cap |det(K_P^C)| > 0\right) \\
\geq & \inf_{P\in\mathcal{P}} P\left(\left\|q'\lambda^C\right\| > \rho \cap |det(K_P^C)| > 0\right) \\
= & \inf_{P\in\mathcal{P}} P\left(\left\|\tilde{\mu}^{C\prime} g^C\right\| > \rho \cap |det(K_P^C)| > 0\right) \\
= & \inf_{P\in\mathcal{P}} P\left(\left\|\left[\tilde{\mu}^{C\prime}, \underset{1\times J+d}{0}\right] g\right\| > \rho \cap |det(K_P^C)| > 0\right) \\
= & \inf_{P\in\mathcal{P}} P\left(\left\|\left[\frac{\tilde{\mu}^{C\prime}}{\|\tilde{\mu}^C\|}, \underset{1\times J+d}{0}\right] g\right\| > \frac{\rho}{\|\tilde{\mu}^C\|} \cap |det(K_P^C)| > 0\right) \\
\geq & 2\Phi\left(-\rho/\left(M_\eta\sqrt{d\omega}\right)\right) \geq 1 - \eta.
\end{aligned}$$

If $C$ picks some stochastic constraints and some non-stochastic ones, then define $\bar{\mu}$ to be the $(J+2d)$-vector that agrees with $\tilde{\mu}^C$ in components picked by $C$ and is zero otherwise. The above algebra then applies with $\left[\frac{\tilde{\mu}^{C\prime}}{\|\tilde{\mu}^C\|}, \underset{1\times J+d}{0}\right]$ replaced by $\bar{\mu}/\left\|\bar{\mu}^{[1:J]}\right\|$. The fact that $\left\|\bar{\mu}^{[1:J]}\right\| > (M_\eta - 2)/2d\bar{M}$ follows because the vector $\bar{\mu}^{[1:J]}$ contains the entries of $\tilde{\mu}^C$ corresponding to stochastic constraints, and by Lemma C.3, if $\left\|\tilde{\mu}^C\right\| > M_\eta$ we have that $\left\|\tilde{\mu}^{[d_1+1:d]}\right\| > (M_\eta - 2)/2d\bar{M}$, with $\tilde{\mu}^{[d_1+1:d]}$ denoting the entries of $\tilde{\mu}$ corresponding to the stochastic constraints in $C$. Hence, $M_\eta$ can be chosen also in this case so that the result holds. Finally, $C$ cannot pick only nonstochastic constraints, because in that case $\|\tilde{\mu}\| = 1$ and the assumption is violated. □

Lemma C.6: *Suppose Assumptions 3.1, 3.2, 3.3 hold. Fix any index set $C$ of cardinality $d$. Then*

*(i)*

$$\sup_{P\in\mathcal{P}} P(det(K_P^C) = 0 \cap \{\exists\lambda \in \mathbb{R}^d : K_P^C\lambda = g_P^C\}) = 0,$$

*(ii) for any $\eta > 0$ there exists $\bar{\alpha}_\eta > 0$ such that*

$$\sup_{P\in\mathcal{P}} P(0 < |\alpha_d^C| < \bar{\alpha}_\eta \cap \{\lambda^C \in \rho B^d\}) < \eta,$$

*where $\lambda^C = (K_P^C)^{-1}g_P^C$ and $\alpha_d^C$ denotes the smallest eigenvalue of $K_P^{C\prime}K_P^C$.*

*Proof.* *(i)* If $K_P^C$ is singular, there exists $h^C \in \mathbb{R}^d$, $h^C \neq 0$, s.t. $h^{C\prime} K^C = 0$. Then $K^C \lambda = g^C$ implies $h^{C\prime} g^C = 0$. Let $h$ denote the $(J+2d)$-vector that agrees with $h^C$ on components corresponding to the index set $C$ and contains zeros otherwise. If $\left\| h^{[1:J]} \right\| = 0$, then the submatrix of $K^C$ corresponding to nonstochastic constraints is singular. This is only possible if $C$ picks opposing faces of the $\rho$-cube, in which case the conclusion is obvious. If $\left\| h^{[1:J]} \right\| > 0$, let $\tilde{h} = h / \left\| h^{[1:J]} \right\|$. Then $h^{C\prime} g^C = 0$ iff $\tilde{h}' g = 0$, but by Lemma C.4 for all $\epsilon \geq 0$,

$$\inf_{P \in \mathcal{P}} P\left( \left\| \tilde{h}' g \right\| > \epsilon \right) \geq 2\Phi\left( -\epsilon / \sqrt{\omega} \right),$$

yielding the desired result for $\epsilon = 0$.

*(ii)* Let $q^C$ denote the eigenvector associated with $\alpha_d^C$ (recall that because $K^{C\prime} K^C$ is symmetric, $\left\| q^C \right\| = 1$). Then we have

$$\begin{aligned} 1 &= \left\| q^C \right\| = \left\| q^{C\prime} q^C \right\| = \left\| (K^C q^C)' (K^{C\prime})^{-1} q^C \right\| \\ &\leq \left\| (K^C q^C) \right\| \left\| (K^{C\prime})^{-1} q^C \right\|, \end{aligned}$$

and therefore, denoting $\tilde{\mu} = (K^{C\prime})^{-1} q^C$,

$$\|\tilde{\mu}\|^2 = \left\| (K^{C\prime})^{-1} q^C \right\|^2 \geq \frac{1}{\|(K^C q^C)\|^2} = \frac{1}{\alpha_d^C} > \frac{1}{\bar{\alpha}_\eta}.$$

It then follows from Lemma C.5 that we can choose $\bar{\alpha}_\eta$ so that the claim holds. ☐

Lemma C.7: *Suppose Assumptions 3.1, 3.2, 3.3 hold. Let $D$ and $C$ denote two distinct subsets of $\{1, \ldots, J+2d\}$ of cardinality $d$, and let $\lambda^C$ and $\lambda^D$ be such that $K_P^C \lambda^C = g_P^C$ and $K_P^D \lambda^D = g_P^D$. Then for all $\epsilon \geq 0$,*

$$\inf_{P \in \mathcal{P}} P\left( \|\lambda^C - \lambda^D\| > \epsilon \right) > 2\Phi(-\epsilon \bar{M} / \sqrt{\omega}).$$

*Proof.* Pick any two basic solutions indexed by $C$ and $D$. By Lemma C.6-(i),

$$\begin{aligned} \sup_{P \in \mathcal{P}} P\Big( & \{det(K_P^C) = 0 \cap \{\exists \lambda \in \mathbb{R}^d : K_P^C \lambda = g_P^C\}\} \\ & \cup \{det(K_P^D) = 0 \cap \{\exists \lambda \in \mathbb{R}^d : K_P^D \lambda = g_P^D\}\}\Big) = 0. \end{aligned}$$

Let $A_1$ denote the event that $\{|det(K_P^C)| > 0 \cap |det(K_P^D)| > 0\}$. Initially assume that both basic solutions pick only stochastic constraints, then it is w.l.o.g. to assume that $C = \{1, ..., d\}$ and that $d+1 \in D \backslash C$. Next, $\lambda^D \in \{\lambda : K_P^{[d+1]} \lambda = g_P^{[d+1]}\}$ implies that

$$\left\| \lambda^C - \lambda^D \right\| \geq d\left( \lambda^C, \{\lambda : K_P^{[d+1]} \lambda = g_P^{[d+1]}\} \right) = \frac{|K_P^{[d+1]} \lambda^C - g_P^{[d+1]}|}{\left\| K_P^{[d+1]} \right\|}.$$

Hence,

$$\begin{aligned} & \inf_{P \in \mathcal{P}} P\left( \left\| \lambda^C - \lambda^D \right\| > \epsilon \right) \\ \geq & \inf_{P \in \mathcal{P}} P\left( \left| K_P^{[d+1]} \left( K_P^{C\prime} \right)^{-1} g_P^C - g_P^{[d+1]} \right| / \left\| K_P^{[d+1]} \right\| > \epsilon \cap A_1 \right) \\ = & \inf_{P \in \mathcal{P}} P\left( \left| \underbrace{\left[ K_P^{[d+1]} \left( K_P^{C\prime} \right)^{-1}, -1, \underset{1 \times J+d-1}{0} \right]}_{=: k_P} g_P \right| > \epsilon \left\| K_P^{[d+1]} \right\| \cap A_1 \right) \\ \geq & \inf_{P \in \mathcal{P}} P\left( |k_P' g_P| > \epsilon \bar{M} \cap A_1 \right) \\ \geq & \inf_{P \in \mathcal{P}} P\left( |k_P' g_P| / \left\| k_P^{[1:J]} \right\| > \epsilon \bar{M} \cap A_1 \right) \\ \geq & 2\Phi\left( -\epsilon \bar{M} / \sqrt{\omega} \right), \end{aligned}$$

with $\bar{M} \geq \|K_P^{[d+1]}| \geq \underline{M}$ by Assumptions 3.2-3.3. Here, we used that $k_P^{[d+1]} = -1$, hence $\left\|k_P^{[1:J]}\right\| > 1$, and invoked Lemma C.4.

If neither $C$ nor $D$ pick any stochastic constraint, then $\lambda^C$ and $\lambda^D$ are distinct corners of the $\rho$-cube and $\left\|\lambda^C - \lambda^D\right\| \geq \rho$ with probability one. Finally, assume w.l.o.g. that $D$ but potentially not $C$ contains a stochastic constraint. Without further loss of generality, assume that $1 \in D/C$. Let $k_P$ denote the $(J+2d)$-vector that agrees with $K_P^{[1]} \left(K_P^{C\prime}\right)^{-1}$ in components corresponding to elements of $C$, that has first component $(-1)$ and that otherwise equals zero. Then the above algebra applies using the new definition of $k_P$. □

Lemma C.8: *Suppose Assumptions 3.1, 3.2, 3.3, 3.4 hold. Then for any $\epsilon \geq 0$, for any $\delta > 0$, for any $r \in \mathbb{R}$, and for any vector $k \in \mathbb{R}^{J+2d}$ such that $\|k^{[1,\dots,J]}\| = 1$, it follows that there exists $N$ such that $n \geq N$ implies*

$$\inf_{P \in \mathcal{P}} P\left(\left\|k' g_n - r\right\| > \epsilon\right) \geq 2\Phi\left(-\left(\epsilon + \delta\right)/\sqrt{\omega}\right) - \delta.$$

*Proof.*

$$\begin{aligned} \inf_{P \in \mathcal{P}} P\left(\left\|k' g_n - r\right\| > \epsilon\right) &= \inf_{P \in \mathcal{P}} P\left(\left\|k'\left(g_P + \left(g_n - g_P\right)\right) - r\right\| > \epsilon\right) \\ \geq \quad & \inf_{P \in \mathcal{P}} P\left(\left\|k' g_P - r\right\| > \epsilon + \delta\right) - P\left(\left\|g_n - g_P\right\| > \delta\right). \end{aligned}$$

Choosing $N$ such that $\sup_{P \in \mathcal{P}} P\left(\|g_n - g_P\| > \delta\right) \leq \delta$ and applying Lemma C.4, yields the result. □

# References


Andrews, D. W. K., and P. J. Barwick (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80(6), 28052826.

Andrews, D. W. K., S. T. Berry, and P. Jia (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," mimeo.

Andrews, D. W. K., and X. Cheng (2012): "Estimation and Inference with Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80, 2153–2211.

Andrews, D. W. K., and P. Guggenberger (2009): "Validity of Subsampling and 'Plug-In Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25(3), 669–709.

Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157.

Beresteanu, A., and F. Molinari (2008): "Asymptotic properties for a class of partially identified models," *Econometrica*, 76, 763–814.

Bonnans, J. F., and A. Shapiro (1998): "Optimization problems with perturbations: A guided tour," *SIAM review*, 40(2), 228–264.

Boyd, S., and L. Vandenberghe (2004): *Convex Optimization*. Cambridge University Press, Cambridge.

Bugni, F. A. (2009): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78(2), 735–753.

Bugni, F. A., I. A. Canay, and X. Shi (2014): "Inference for Functions of Partially Identified Parameters in Moment Inequality Models," Working Paper CWP22/14, CeMMAP.

——— (2015): "Specification tests for partially identified models defined by moment inequalities," *Journal of Econometrics*, 185(1), 259–282.

Canay, I. (2010): "EL inference for partially identified models: large deviations optimality and bootstrap validity," *Journal of Econometrics*, 156(2), 408–425.

Chen, X., E. Tamer, and A. Torgovitsky (2011): "Sensitivity Analysis in Semiparametric Likelihood Models," Working paper.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets In Econometric Models," *Econometrica*, 75, 1243–1284.

Chernozhukov, V., E. Kocatulum, and K. Menzel (2015): "Inference on Sets in Finance," *Quantitative Economics*, forthcoming.

Chesher, A., A. M. Rosen, and K. Smolinski (2012): "An instrumental variable model of multiple discrete choice," *Quantitative Economics*, forthcoming.

Demyanov, V. F. (2009): "Dini and Hadamard Derivatives in Optimization," in *Encyclopedia of Optimization*, pp. 710–716. Springer.

Freyberger, J., and J. Horowitz (2013): "Identification and shape restrictions in nonparametric instrumental variables estimation," .

Hirano, K., and J. Porter (2012): “Impossibility results for nondifferentiable functionals,” *Econometrica*, 80(4), 1769–1790.

Imbens, G. W., and C. F. Manski (2004): “Confidence Intervals for Partially Identified Parameters,” *Econometrica*, 72, 1845–1857.

Kaido, H. (2012): “A Dual Approach to Inference for Partially Identified Econometric Models,” Working Paper.

Kaido, H., and A. Santos (2014): “Asymptotically efficient estimation of models defined by convex moment inequalities,” *Econometrica*, 82(1), 387–413.

Kitagawa, T. (2012): “Inference and Decision for Set Identified Parameters Using Posterior Lower Probabilities,” CeMMAP Working Paper.

Kline, B., and E. Tamer (2015): “Bayesian inference in a class of partially identified models,” Working Paper.

Leeb, H., and B. M. Pötscher (2005): “Model Selection and Inference: Facts and Fiction,” *Econometric Theory*, 21, 21–59.

Mikusheva, A. (2007): “Uniform inference in autoregressive models,” *Econometrica*, 75, 1411–1452.

Molchanov, I. (2005): *Theory of Random Sets*. Springer, London.

Molchanov, I. S. (1998): “A Limit Theorem for Solutions of Inequalities,” *Scandinavian Journal of Statistics*, 25, 235–242.

Molinari, F. (2008): “Partial Identification of Probability Distributions with Misclassified Data,” *Journal of Econometrics*, 144(1), 81–117.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2011): “Moment Inequalities and Their Application,” Discussion Paper, Harvard University.

Politis, D., J. Romano, and M. Wolf (1999): *Subsampling*. Springer-Verlag, Berlin.

Rockafellar, R. T. (1970): *Convex Analysis*. Princeton University Press, Princeton.

Romano, J. P., and A. M. Shaikh (2008): “Inference for Identifiable Parameters in Partially Identified Econometric Models,” *Journal of Statistical Planning and Inference*, 138, 2786–2807.

——— (2010): “Inference for the Identified Set in Partially Identified Econometric Models,” *Econometrica*, 78(1), 169–211.

Rosen, A. M. (2008): “Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities,” *Journal of Econometrics*, 146(1), 107 – 117.

Shapiro, A. (1995): “Directional differentiability of the optimal value function in convex semi-infinite programming,” *Mathematical programming*, 70(1-3), 149–157.

Stoye, J. (2009): “More on Confidence Intervals for Partially Identified Parameters,” *Econometrica*, 77, 1299–1315.

van der Vaart, A., and J. Wellner (2000): *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag, Berlin.

Wan, Y. (2013): "An Integration-based Approach to Moment Inequality Models," Working Paper.

Wets, R. J.-B. (1985): "On the Continuity of the Value of a Linear Program and of Related Polyhedral-valued Multifunctions," *Mathematical Programming Study*, 24, 14–29.