Inference on the value of a linear program*

Leonard Goff[1] and Eric D. Mbakop[2]

[1]Department of Economics, University of Calgary

[2]Department of Economics, The Ohio State University

June 9, 2025

Abstract

This paper studies inference on the value of linear programs (LPs) when both the objective function and constraints are possibly unknown and must be estimated from data. We show that many inference problems in partially identified models can be reformulated in this way. Building on Shapiro (1991) and Fang and Santos (2019), we develop a pointwise valid inference procedure for the value of LPs. We modify this pointwise inference procedure to construct one-sided inference procedures that are uniformly valid over large classes of data-generating processes (DGPs). Our results provide alternative testing procedures for problems considered in Andrews, Roth, and Pakes (2023), Cox and Shi (2023), and Fang et al. (2023) (in the low-dimensional case), and remain valid when key components–such as the coefficient matrix–are unknown and must be estimated. Moreover, our framework also accommodates inference on the identified set of a subvector, in models defined by linear moment inequalities, and does so under weaker constraint qualifications than those in Gafarov (2025).

1 Introduction

Many economic models imply restrictions that can be formulated as linear equalities or inequalities involving estimable quantities. Such restrictions can be used for testing the model, or to provide identifying information about parameters of interest under weaker assumptions than are required to point identify them.

With either goal in mind, the inferential task often reduces to the problem of performing statistical inference on the value function ν of a linear program (LP). An LP takes the generic form:

$$\begin{aligned} \underset{\theta}{\text{minimize}} \quad & c^\top \theta \\ \text{subject to} \quad & A_I \theta \le b_I, \\ & A_E \theta = b_E, \end{aligned} \tag{1.1}$$

*This research was supported by the Social Sciences and Humanities Research Council of Canada. We thank Kirill Ponomarev for useful discussions and Alex Lu for excellent research assistance.

where the parameters $\mu = (A, b, c)$, with $A = \begin{pmatrix} A_E \\ A_I \end{pmatrix}$, may include unknown components that must be estimated from data. We show in Section 2 how special cases of (1.1) arise in settings such as testing moment inequalities, testing linear systems of equalities, and partial identification of linear functionals of model parameters (e.g. subvectors) with estimated constraints. In each setting, hypotheses of interest can be reformulated as statements about the value ν of the LPs.

In this paper, we develop a new approach to testing the optimal value of an LP when potentially all of the parameters μ are not know ex-ante and must be estimated from data. Given estimates $\hat{\mu} = (\hat{A}, \hat{b}, \hat{c})$ of the LP parameters, we let $\hat{\nu}$ denote the value of the sample analog LP. Building on results of Shapiro (1991) and Fang and Santos (2019), we begin by establishing a bootstrap approximation to the distribution of $\sqrt{n}(\hat{\nu} - \nu)$, given $\sqrt{n}$-consistent estimators of μ. The bootstrap approximation can be used to construct confidence sets for ν itself or for other linear functions of the parameter vector (e.g. its components). The result holds pointwise, under constraint qualification conditions on (A, b) that are weaker than those in recent work of Gafarov (2025).

We then modify our pointwise result to construct uniformly valid inference procedures tailored to four settings that are of particular interest in econometrics: (i) moment inequalities, (ii) moment inequalities with linear nuisance parameters, (iii) testing of linear systems with known coefficients, and (iv) testing linear systems with unknown coefficients. In doing so, our methods provide alternative testing procedures in settings considered by Andrews and Soares (2010), Andrews, Roth, and Pakes (2023), Cox and Shi (2023), and Fang et al. (2023). Within each of the four settings, we propose a bootstrap-based inference procedure that controls size uniformly over large classes of DGPs. Moreover the methods are not unnecessarily conservative, potentially obtaining asymptotically exact size in settings (i)-(iv).

Uniform size control is especially desirable in settings where confidence intervals are constructed by test inversion, as is usually done for moment inequalities. However, uniformity over an interestingly large set of DGPs is challenging in the general setting of (1.1). After illustrating the problem, we show that one can nevertheless move forward without overly strong assumptions by exploiting the additional structure that is present in the problems (i)-(iv). Computationally, our uniform inference procedures involve generating a test statistic by solving an LP in the original sample, and then computing a critical value though a bootstrap procedure in which an additional LP is solved within each bootstrap iteration.

In the case of (iv), our results extend recent work by Fang et al. (2023) (FSST) to settings in which the coefficient matrix A of a linear system of equalities (and/or inequalities) is also estimated. Apart from restrictions on the dimension of the problem, our assumptions are comparable to those in FSST, with the addition of a new assumption to control a "condition-like number" of the coefficient matrix (Assumption 4.27). Our results are established in a semi high-dimensional asymptotic regime, where the number of estimated entries in the coefficient matrix A and the right-hand side vector b remains bounded as the sample size increases, while the number of deterministic components in A and b is allowed to grow arbitrarily with the sample size (see Assumption 4.22). Our proposed procedure for this setting involves (in its baseline form) a quadratic program that must be solved once, in addition to the bootstrapped LPs as in the other three settings.

We demonstrate the empirical relevance of our method for an estimated A matrix (iv) to the problem of inference on the average treatment effect (ATE) in a setting where the ATE is partially identified using instrumental variables. Mogstad, Santos, and Torgovitsky (2018) show that instrumental variables generally yield systems of linear equalities that relate causal parameters to moments involving the outcome variable, which can be coupled with auxiliary assumptions to shrink the identified set for a desired treatment effect parameter. Existing inference procedures such as FSST cannot always be applied in such settings, as both the outcome moments and the coefficient matrix typically depend on the DGP. We design a Monte Carlo simulation around the empirical setting of Dupas (2014), in which valid instruments are provided by an experiment that randomized the price of an antimalarial bed net. Our procedure demonstrates desirable power properties, even with a relatively small sample size typical of field experiments.

The structure of the paper is as follows. In Section 2, we show how the settings (i)-(iv) above can be reformulated as special cases of problem (1.1). Section 3 establishes some basic properties of the problem and introduces our pointwise valid inference procedure. Section 4 then turns to uniform inference procedures tailored to problems (i)-(iv). Section 5 presents our application and simulation study. All proofs are deferred to the appendix.

1.1 Notation

Let x be a vector in Euclidean space and A a matrix. Unless stated otherwise, we use $\|x\|$ to denote the Euclidean norm of x, $\|x\|_p$ ($1 \leq p \leq \infty$) for its general l_p-norm, and $\|A\|$ to denote the spectral norm of A. We use $\mathbb{R}^p_+$ to denote the set of elements of $\mathbb{R}^p$ with non-negative entries. We use $\mathbb{1}$ to denote a vector with all entries equal to one. For a given vector x, we define $(x)_+$ and $(x)_-$ as the positive and negative parts of x, respectively, where the i^{th} entry of $(x)_+$ is given by $\max\{x_i, 0\}$ and the i^{th} entry of $(x)_-$ is given by $\max\{-x_i, 0\}$. Given a point x and a subset A of some Euclidean space, we denote the Euclidean distance from x to A by $d(x, A)$. For two sets A and B, we define $\vec{d}_H(A, B)$ and $d_H(A, B)$ as the directed (one-sided) Hausdorff distance from A to B and the Hausdorff distance between A and B, respectively. These are given by $\vec{d}_H(A, B) = \sup_{x \in A} d(x, B)$, $d_H(A, B) = \max\{\vec{d}_H(A, B), \vec{d}_H(B, A)\}$. Given a matrix $A \in \mathbb{R}^{m \times n}$ and a subset of row indices $I \subset [m]$, where for $m \in \mathbb{N}$ we have $[m] := \{1, \cdots, m\}$, we denote by A_I the submatrix obtained by retaining only the rows of A indexed by I and deleting all other rows. Given a vector $b \in \mathbb{R}^m$ and a subset $I \subset [m]$, we define b_I analogously. We use $\Rightarrow$, $\xrightarrow{p}$, and $\overset{p}{\Longrightarrow}$ to denote convergence in distribution, convergence in probability, and convergence in distribution in probability, respectively.

2 Motivating Examples

Below, we present several motivating examples of Eq. (1.1) that illustrate the practical importance and wide applicability of this inferential problem.

Example 2.1 (Moment equalities/inequalities). *Consider a moment inequality model where the true parameter β_0 is assumed to satisfy the moment conditions:*[1]

$$E_{F_0}[m_j(W, \beta_0)] \leq 0 \quad \textit{for } j = 1, \ldots, p, \tag{2.1}$$

where $m_j(\cdot,\cdot)$ for $j = 1, \ldots, p$ are known moment functions, and $W \sim F_0$.

Given a hypothetical value of the parameter $\beta \in \mathcal{B}$, the null hypothesis that β satisfies the moment conditions in (2.1) is equivalent to testing $H_0 : \nu = 0$ against the alternative $\nu > 0$,[2] *where ν denotes the value of the following linear program:*

$$\nu = \max\left\{b^\top \lambda \mid \lambda \in \mathbb{R}^p_+\right\}, \tag{2.2}$$

with $b = E_{F_0} m(W, \beta)$ and m is the vector-valued function with components $m_j(\cdot,\cdot)$ for $j = 1, \ldots, p$.

Below, we propose a new testing procedure for the null hypothesis that $H_0 : E_{F_0}[m(W, \beta_0)] \leq 0$, which exploits the linear programming characterization of the null hypothesis, and which is uniformly valid over the large classes of DGPs considered in Andrews and Soares (2010). Our test is not of the generalized moment selection (GMS) type, and as in Andrews and Soares (2010), our test can be inverted to construct a confidence set for β_0.

Example 2.2 (Moment inequalities with nuisance parameters entering linearly). *As above, consider a moment inequality model where the true parameter $\beta_0 = (\theta_0, \eta_0)$ is assumed to satisfy the moment inequality*[3]

$$E_{F_0}[m(W, \beta_0)] = E_{F_0}[h(W, \theta_0) - A\eta_0] \leq 0,$$

where the function h and the coefficient matrix A ($\in \mathbb{R}^{p\times d}$) are known. We are interested in inference on the component θ_0 of β_0, treating η_0 as a nuisance parameter.

A confidence set for θ_0 can be obtained by projecting a confidence set for β_0, using the approach in Andrews and Soares (2010). However, this approach is computationally expensive when the nuisance parameter η is high-dimensional, as it involves testing the null hypothesis at each point on a grid in $\mathbb{R}^{d_\beta}$, where d_β is the dimension of the parameter β_0. Moreover, as noted in Bugni, Canay, and Shi (2017), projection-based tests can be quite conservative.

Following Andrews, Roth, and Pakes (2023), and using Farkas' Lemma, testing whether there exists some η_0 such that $\beta_0 = (\theta_0, \eta_0)$ satisfies (2.4) is equivalent to testing whether $\nu = 0$ (against the alternative $\nu > 0$), where ν is the value of the following linear program:

$$\nu = \max\left\{b^\top \lambda \mid \lambda \in \mathbb{R}^p_+,\ A^\top \lambda = 0\right\}, \tag{2.3}$$

[1] Moment equalities can be equivalently expressed as pairs of moment inequalities, so this setting also accommodates moment equalities.

[2] Under the alternative, we actually have $\nu = +\infty$. In Section 4.1, we introduce additional constraints to our linear program in order to ensure that the value function is bounded and strictly positive under the alternative.

[3] Although a different hypothesis is initially posed, the one ultimately tested in Andrews, Roth, and Pakes (2023) and Cox and Shi (2023) coincides with ours, as their inference is conducted conditional on the realized values of the instruments.

with $b = E_{F_0}[h(W_i, \theta_0)]$.

For a hypothetical value of the parameter θ, *we propose below a testing procedure for the null hypothesis*

$$H_0 : \exists \eta \text{ such that } E_{F_0}[h(W_i, \theta) - A\eta] \leq 0, \tag{2.4}$$

which exploits the linear programming characterization of the null hypothesis and is uniformly valid over a large class of data-generating processes (DGPs). In particular, the class of DGPs we accommodate is less restrictive in some ways than those considered in Andrews, Roth, and Pakes (2023) and Cox and Shi (2023). Our results are established in a "semi high-dimensional" regime where the dimension of estimated component of b *is bounded while the dimension of the deterministic component of* b *can grow arbitrarily with the sample size. By contrast, the uniform validity of the methods of Andrews, Roth, and Pakes (2023) and Cox and Shi (2023) are established in a regime where the overall dimension of* b *is bounded. Another advantage of our procedure is that linear dependencies among the rows of* b *can be accommodated without the researcher specifying the form of that dependency.*

As in Andrews, Roth, and Pakes (2023) and Cox and Shi (2023), our test can be inverted to construct a confidence set for θ_0. *This involves conducting the test at a set of points on a grid in* $\mathbb{R}^{d_\theta}$, *where* d_θ *is the dimension of the parameter of interest* θ_0.

Example 2.3 (Inference on linear systems with known coefficients). *Fang et al. (2023) consider the problem of testing the following hypothesis:*

$$H_0 : \exists \eta \geq 0 \quad \text{such that} \quad b = A\eta, \tag{2.5}$$

where the coefficient matrix $A \in \mathbb{R}^{p \times d}$ *is known, and the vector* b *is to be estimated. Fang et al. (2023) provide several motivating examples for considering such a hypothesis.*

Using Farkas' lemma, testing whether (2.5) *holds is equivalent to testing whether* $\nu = 0$ *(against the alternative* $\nu > 0$*), where* ν *is the value of the following linear program:*

$$\nu = \max \left\{ b^\top \lambda \mid \lambda \in \mathbb{R}^p,\ A^\top \lambda \leq 0 \right\}. \tag{2.6}$$

Below, we propose a testing procedure for the hypothesis (2.5), *based on its linear programming formulation* (2.6). *Our method is valid over broad classes of DGPs, similar to those considered in Fang et al. (2023). Unlike Fang et al. (2023), who allow for the dimension* p *to grow at a controlled rate while the dimension* d *is unrestricted, our approach permits unrestricted growth only in the dimension d, but allows for* p *to grow with* N *as long as the dimension of the unknown (subject to estimation) component of* b *remains bounded. The procedures of Andrews, Roth, and Pakes (2023) and Cox and Shi (2023) can also be applied to test* (2.5), *but require both dimensions of* A *to remain uniformly bounded.*[4] *Our approach is also more general than that of Bai, Santos, and Shaikh (2022),*

[4]Note that the hypothesis in (2.5) is equivalent to testing $H_0 : \exists \eta$ such that $\tilde{b} - \tilde{A}\eta \leq 0$, where $\tilde{b} = (b', -b', 0'_{d_\eta})'$ and $\tilde{A} = (A', -A', -\mathbb{I}_{d_\eta})'$. Thus, hypotheses of the form (2.5) can be framed as special cases of the more general form (2.4), and

who consider testing (2.5) under fixed dimensions of A, while we allow the number of columns to grow freely. Furthermore, unlike their procedure, the test that we propose is scale invariant.

Example 2.4 (Inference on linear systems with unknown coefficients). *Consider the setting of Examples 2.2 and 2.3, where the coefficient matrix A is unknown and must be estimated from the data. In such settings, the linear programming characterizations of the null hypotheses in 2.3 and 2.6 remain valid. However, as the feasible regions of these LPs depend on A, the feasible region is now unknown and must be "estimated" from the data.*

One way in which such hypotheses may arise is when the researcher is interested in inferring $\tau_0 = C\theta_0$, for some known matrix C, where the identified set for θ_0 is defined by the linear system of inequalities:

$$A\theta \leq b$$

and both the matrix A and the vector b may be unknown. Following the suggestion of Fang et al. (2023), a confidence interval for the subvector τ_0 can be obtained by inverting a test of the following hypothesis for various candidate values of τ:

$$H_0 : \exists\, \theta \text{ such that } A\theta \leq b \quad \text{and} \quad C\theta = \tau. \tag{2.7}$$

It is easy to verify that the latter hypothesis can be written in the form of 2.4 or 2.5, where the coefficient matrix A is now potentially unknown. Below, we propose a uniformly valid inference procedure for the hypothesis in 2.4 or 2.5 when the matrix A must be estimated. The conditions required for uniform validity are similar to those in Andrews, Roth, and Pakes (2023), Cox and Shi (2023), and Fang et al. (2023)[5]*.*

When C is a vector—so that τ_0 is scalar—the identified set for τ_0 is given by the interval $[\nu_L, \nu_U]$, where ν_L and ν_U are the optimal values of the following linear programs:

$$\nu_L = \min\{C^\top\theta \mid A\theta \leq b\}, \quad \nu_U = \max\{C^\top\theta \mid A\theta \leq b\}. \tag{2.8}$$

Thus, the endpoints of the identified set correspond to the values of two LPs, and constructing a confidence interval for the identified set of τ_0 reduces to inferring the values ν_L and ν_U. We propose such an inference procedure below, which is uniformly valid over a smaller class of DGPs than those required to obtain uniform inference for the null in 2.7. Therefore, although the procedure that we propose for inferring ν_L and ν_U provides a computationally cheap way of conducting inference on the identified set of τ_0, it is uniformly valid under more stringent assumptions than are needed to construct uniformly valid confidence intervals for τ_0 through test inversion of the hypothesis in 2.7. For empirical applications that motivate such inference procedures, see Gafarov (2025) and Cho and Russell (2024).

vice versa. However, when the number of columns of A grows with sample size, so does the dimension of $\tilde{A}$, a case not covered by the procedures in Andrews, Roth, and Pakes (2023) or Cox and Shi (2023).

[5] The main additional assumption we impose is Assumption 4.27, which requires a uniform bound on a "condition number" of the matrices A.

3 Pointwise valid inference procedure

In this section, we develop a pointwise valid inference procedure for the value of a linear program (LP). Specifically, we consider the estimation of ν, the value of the LP

$$\min\{c^\top\theta \mid A_E\theta = b_E,\ A_I\theta \le b_I\}. \tag{3.1}$$

We focus on a setting where $A = (A_E', A_I')'$, b, and c are potentially unknown and must be estimated from data. To estimate ν, we consider the sample analogue of LP 3.1, given by

$$\hat{\nu} = \min\{\hat{c}^\top\theta \mid \hat{A}_E\theta = \hat{b}_E,\ \hat{A}_I\theta \le \hat{b}_I\}, \tag{3.2}$$

where $(\hat{A}, \hat{b}, \hat{c})$ is a consistent estimator of (A, b, c). In general, even if the population LP 3.1 is feasible and has a finite value, the sample LP 3.2 may be infeasible with probability bounded away from zero as the sample size N tends to infinity.

To address this issue, we impose Assumption 3.1, the Mangasarian-Fromovitz Constraint Qualification (MFCQ), which provides the minimal regularity condition required to ensure that the feasible region $\hat{\mathcal{P}}$ of LP 3.2 is nonempty with probability approaching one as $N \to \infty$ (see Lemmas 2 and 3 in Robinson (1977)).

Using a dual characterization of the MFCQ, we propose in Proposition 3.5 below a preprocessing step that transforms any LP that may not satisfy the MFCQ condition into an equivalent LP that does. As a result, Theorem 1 below, the main result of this section, applies broadly.

Assumption 3.1 (Mangasarian-Fromovitz Constraint Qualification)**.** The matrix A_E has full row rank, and there exists a point θ_0 such that $A_E\theta_0 = b_E$ and $A_I\theta_0 < b_I$.

We consider the following assumption, which guarantees that the values ν of our LP is finite.

Assumption 3.2. The feasible region

$$\mathcal{P} = \{\theta \in \mathbb{R}^d \mid A_E\theta = b_E \text{ and } A_I\theta \le b_I\}$$

is compact.

we show in the proof of Theorem 3.4 that Assumptions 3.1 and 3.2 imply that $\hat{\nu}$ is finite with probability approaching 1 as N tends to infinity.

The dual of the LP 3.1 is given by

$$\max\left\{\lambda_E^\top b_E + \lambda_I^\top b_I \mid \lambda_I \le 0,\ A_E^\top\lambda_E + A_I^\top\lambda_I = c\right\}. \tag{3.3}$$

Let $\hat{\mathcal{P}}$ be defined by

$$\hat{\mathcal{P}} := \{\theta \in \mathbb{R}^d \mid \hat{A}_E\theta = \hat{b}_E,\ \hat{A}_I\theta \le \hat{b}_I\}.$$

By convention, we set $\hat{\nu} = +\infty$ if $\hat{\mathcal{P}}$ is empty. When $\hat{\mathcal{P}}$ is nonempty, define $\hat{S}$ and $\hat{\Delta}$ as follows:

$$\hat{S} = \left\{\theta \in \mathbb{R}^d \mid \hat{A}_E\theta = \hat{b}_E,\ \hat{A}_I\theta \leq \hat{b}_I,\ \hat{c}^T\theta \leq \hat{\nu} + \kappa_N/\sqrt{N}\right\} \tag{3.4}$$

and

$$\hat{\Delta} = \left\{\lambda = (\lambda_E, \lambda_I) \in \mathbb{R}^p \mid \hat{A}_E^T\lambda_E + \hat{A}_I^T\lambda_I = \hat{c},\ \lambda_I \leq 0, \lambda_E^T\hat{b}_E + \lambda_I^T\hat{b}_I \geq \hat{\nu} - \kappa_N/\sqrt{N}\right\}, \tag{3.5}$$

for some sequence of constants κ_N such that $\kappa_N \to \infty$ and $\kappa_N/\sqrt{N} \to 0$. The sets $\hat{S}$ and $\hat{\Delta}$ can be viewed as estimators of the optimal solution sets of the primal and dual LPs, respectively.

Let $d = \text{vec}(A, b, c)$, and let $\hat{d}$ denote the sample counterpart of d. Let d_{Pr} denote the Prokhorov metric for weak convergence. The following assumption requires that $\sqrt{N}(\hat{d} - d)$ converges in distribution to a limiting law $\mathbb{G}$, and that there exists a bootstrap-type estimator ζ_N^* which consistently approximates this asymptotic distribution.

Assumption 3.3. Suppose that $d_{Pr}\left(\sqrt{N}(\hat{d} - d), \mathbb{G}\right) \to 0$, and $d_{Pr}\left(\mathcal{L}(\zeta_N^* \mid \mathbb{W}), \mathbb{G}\right) \xrightarrow{p} 0$, where ζ_N^* is a random variable with distribution that is completely determined from the data.

We now state the main result of this section, it builds on ideas from Shapiro (1991) and Fang and Santos (2019) to constuct a pointwise valid inference procedure for the distribution of $\sqrt{N}(\hat{\nu} - \nu)$. The bootstrap procedure we propose is computationally straightforward, as it requires solving only a single LP in each bootstrap iteration. In fact, leveraging LP duality, the min-max optimization problem can be reformulated as a linear program. The proof is provided in Section 7.0.1 of the appendix.

Theorem 3.4. *Suppose that Assumptions 3.1, 3.2 and 3.3 hold. Then*

$$d_{Pr}\left(\mathcal{L}\left(\min_{\theta\in\hat{S}}\max_{\lambda\in\hat{\Delta}}\langle\zeta_c^*, \theta\rangle + \lambda^T\left(\zeta_b^* - \zeta_A^*\theta\right)|\mathbb{W}\right),\ \min_{\theta\in S_0}\max_{\lambda\in\Delta_0}\langle\zeta_c, \theta\rangle + \lambda^T\left(\zeta_b - \zeta_A\theta\right)\right) = o_p(1)$$

where $(\zeta_c, \zeta_b, \zeta_A) \overset{d}{\sim} \mathbb{G}$.

When Assumption 3.1 is not satisfied, but it is known that the ball $\{\lambda \mid \|\lambda\|_\infty \leq M\}$ has a nonempty intersection with the solution set of the dual LP 3.3, the following proposition constructs an LP that has the same value as the original LP 3.1 and satisfies the MFCQ condition. This result follows from Lemma 7.11, which establishes that the MFCQ condition holds if and only if the set of optimal solutions to the dual LP is bounded (see parts (a) and (b) of Theorem 1 in Robinson (1977)). The proposition is derived by first adding the constraint $\|\lambda\|_\infty \leq M$ to the feasible region of the dual LP 3.3. The resulting modified dual LP has the same optimal value as the original LP. Taking the dual of this modified dual then yields a new primal LP that satisfies the MFCQ condition and retains the same optimal value.

Proposition 3.5. *Suppose that the LP 3.1 is feasible and has a finite optimal value. If there exists a constant $M > 0$ such that*

$$\|\lambda\|_\infty \leq M \quad \textit{for some } \lambda \in \arg\min\left\{\lambda_E^\top b_E + \lambda_I^\top b_I \mid \lambda_I \leq 0,\ A_E^\top\lambda_E + A_I^\top\lambda_I = c\right\},$$

then the optimal value ν *of LP 3.1 also satisfies*

$$\nu = \min \left\{ \theta^\top c + M\mathbb{1}^\top x + M\mathbb{1}^\top (y+z) \;\middle|\; \begin{array}{l} x \geq 0,\ y \geq 0,\ z \geq 0, \\ A_E\theta + y - z = b_E, \\ A_I\theta + x \geq b_I \end{array} \right\}, \tag{3.6}$$

and the above reformulated LP, which includes auxiliary variables x, y, *and* z, *satisfies the MFCQ condition.*

Remark 3.6. A little bit of algebra shows that the reformulated LP can be written as

$$\nu = \min_\theta c^\top\theta + M(b_I - A_I\theta)_+ + M\|b_E - A_E\theta\|_1.$$

When no inequality constraints are present, this yields the *penalty function estimators* considered in Voronin (2025). The penalty function estimator of the value of the LP studied in Voronin (2025) can therefore be interpreted as a method for transforming the original LP into one that satisfies the MFCQ condition. Moreover, the preceding proposition provides a general framework for constructing such *penalty function estimators*. Specifically, for any polytope $\mathcal{C} = \{\lambda \mid H\lambda \leq h\}$ that contains the origin in its interior, we can construct a corresponding penalty function estimator as follows: first, add the constraint $H\lambda \leq Mh$ to the feasible region of the dual LP 3.3, where $M > 0$ is a constant that serves as a tuning parameter. Note that if M is chosen large enough so that an optimal solution λ to the dual LP 3.3 satisfies $H\lambda \leq Mh$, then the value of the modified dual LP remains equal to the value of the original dual LP 3.3. Taking the dual of this modified dual LP yields a primal LP that satisfies the MFCQ condition and produces a penalty function estimator of the original LP value ν.

Remark 3.7. In Proposition 3.5, the bound M on a solution to the dual LP may not be known. Voronin (2025) derives and consistently estimates the asymptotic distribution of $\hat{\nu}$, obtained by solving the sample counterpart to LP 3.6, where M is allowed to grow with the sample size N at an appropriate rate.

4 Uniformly valid inference procedures

In this section, we adapt the result of Theorem 3.4 to derive uniformly valid inference procedures for each of the problems discussed in Examples 2.1 through 2.4. We then proceed to establish the validity of the resulting procedure—under appropriate assumptions—on a case-by-case basis in the subsections that follow.

The starting point of our argument is the observation that, in the context of uniform inference, a uniform version of Theorem 3.4 is unnecessarily restrictive. As shown by Andrews (2000), such a result can only hold over a necessarily "small" class of data-generating processes (DGPs).

To illustrate this point, consider the family $\mathcal{Q} = \{N(\mu, \mathbb{I}_2) \mid \mu \in \mathbb{R}^2\}$, consisting of bivariate normal distributions with unknown mean μ and identity covariance matrix. Suppose we are interested in testing the null hypothesis $H_0 : \mu \leq 0$ (interpreted component-wise) against the alternative that this inequality is violated. Assume we observe an i.i.d. sample $\{W_i\}_{i=1}^N$, where $W_i \sim N(\mu, \mathbb{I}_2)$.

The linear programming formulation of this testing problem yields the statistic

$$\hat{\nu} = \max\left\{\lambda^\top \bar{W} \mid \lambda \geq 0,\ \lambda^\top \mathbb{1} = 1\right\} = \bar{W}_1 \vee \bar{W}_2,$$

where $\bar{W}$ denotes the sample average of the W_i's. Without loss of generality, suppose $\mu_1 \geq \mu_2$, in which case we have

$$\sqrt{N}(\hat{\nu} - \nu) = \left(\sqrt{N}(\bar{W}_1 - \mu_1)\right) \vee \left(\sqrt{N}(\bar{W}_2 - \mu_2) - \sqrt{N}(\mu_1 - \mu_2)\right).$$

The distribution of $\sqrt{N}(\hat{\nu} - \nu)$ thus depends on the local parameter $\sqrt{N}(\mu_1 - \mu_2)$, for which there is no uniformly consistent estimator. This implies that a uniform version of Theorem 3.4 cannot hold even over the relatively small class $\mathcal{Q}$.

This observation motivates our focus on constructing uniformly valid one-sided inference procedures, as such procedures can be obtained under weaker conditions on the class of data-generating processes (DGPs), as will be shown below. Our one-sided procedures are based on a modification of the bootstrap procedure

$$\min_{\theta \in \hat{S}} \max_{\lambda \in \hat{\Delta}} \langle \zeta_c^*, \theta \rangle + \lambda^\top \left(\zeta_b^* - \zeta_A^* \theta\right) \tag{4.1}$$

that is used in Theorem 3.4 to estimate the distribution of $\sqrt{N}(\hat{\nu} - \nu)$. In the subsequent sections, we show that for each of the inference problems in Examples 2.1 through 2.4, a uniformly valid inference procedure can be obtained by either dropping the "min" operator in (4.1) and choosing suitable values of θ and $\hat{\Delta}$ to uniformly control the upper quantiles of $\sqrt{N}(\hat{\nu} - \nu)$, or by dropping the "max" operator and selecting appropriate values of λ and $\hat{S}$ to uniformly control the lower quantiles of $\sqrt{N}(\hat{\nu} - \nu)$. Moreover, in all cases, we show that the resulting procedures have asymptotically exact size and are therefore not unnecessarily conservative.

4.1 Moment Equalities/Inequalities

In this section, we derive a uniformly valid test for the hypothesis described in Example 2.1, which we restate below:

$$H_0 : Q \in \mathcal{Q}_0 \quad \text{versus} \quad H_1 : Q \in \mathcal{Q} \setminus \mathcal{Q}_0, \tag{4.2}$$

where

$$\mathcal{Q}_0 = \{Q \in \mathcal{Q} \mid b(Q) \leq 0\},$$

and $b(Q)$ is a vector-valued parameter of Q that must be estimated from data. The test we derive below is based on a characterization of the null hypothesis in terms of the value function of linear programs, and provides an alternative to several well-known tests available in the literature. We operate within a low-dimensional asymptotic framework, in which the dimension of b remains bounded as the sample size diverges.

Remark 4.1. To map the setting of Example 2.1 into 4.2, define $b(Q) := \mathbb{E}_F[m(W, \theta)]$, where m is a known, potentially vector-valued function of a parameter θ in a parameter space Θ, and W has an unknown distribution $F \in \mathcal{F}$. The class $\mathcal{F}$ is such that the expectation $\mathbb{E}_F[\|m(W,\theta)\|]$ is finite for each $\theta \in \Theta$. Then, we define

$$\mathcal{Q} = \{(\theta, F) \mid \theta \in \Theta,\ F \in \mathcal{F}\}, \quad \text{and} \quad \mathcal{Q}_0 = \{Q = (\theta, F) \in \mathcal{Q} \mid b(Q) = \mathbb{E}_F[m(W,\theta)] \leq 0\}.$$

The testing procedure that we develop below can be used to determine whether a given value $\theta_0 \in \Theta$ belongs to the identified set for θ, denoted $\Theta_I(F)$, where F denotes the true but unknown distribution of W, and defined by

$$\Theta_I(F) := \{\theta \in \Theta \mid (\theta, F) \in \mathcal{Q}_0\}.$$

We suppose that we have access to a uniformly consistent estimator $\hat{b}$ of b, which satisfies a uniform asymptotic normality property, as stated in the following assumption.

Assumption 4.2. For each $Q \in \mathcal{Q}$, there exists a symmetric matrix $\Sigma(Q)$, representing the asymptotic variance of $\hat{b}$, such that for all $\epsilon > 0$,

$$\overline{\lim}_{N\to\infty} \sup_{Q\in\mathcal{Q}} d_{pr}\left(\sqrt{N}(\hat{b} - b(Q)), \mathcal{N}(0, \Sigma(Q))\right) = 0.$$

Let $D(Q) := \operatorname{diag}(\Sigma(Q))^{1/2}$ denote the diagonal matrix formed by taking the square root of the diagonal elements of $\Sigma(Q)$. Since deterministic components of b can be discarded, we assume that the diagonal entries of $D(Q)$ are nonzero. The next assumption requires that these entries are uniformly bounded away from zero and infinity, and that a uniformly consistent estimator $\hat{D}$ of $D(Q)$ is available. The estimator $\hat{D}$ will be used to construct a scale-invariant test for the null hypothesis 4.2.

Assumption 4.3. i) There exist positive constants $\underline{\sigma}$ and $\bar{\sigma}$, with $\underline{\sigma} < \bar{\sigma}$, such that for all $Q \in \mathcal{Q}$, all diagonal elements of $D(Q)$ lie in the interval $[\underline{\sigma}, \bar{\sigma}]$.

ii) For all $\epsilon > 0$,

$$\overline{\lim}_{N\to\infty} \sup_{Q\in\mathcal{Q}} Q\left(\|\hat{D} - D(Q)\| > \epsilon\right) = 0.$$

Remark 4.4. In the context of Remark 4.1, for a given value of θ, let $\hat{b}$ ($=\hat{b}(\theta)$) and $\hat{D}$ ($=\hat{D}(\theta)$) be given by

$$\hat{b} = \bar{m}(\theta) = \frac{1}{N}\sum_{i=1}^{N} m(W_i, \theta), \quad \text{and} \quad \hat{D} = \operatorname{diag}(\hat{\Sigma})^{1/2},$$

where

$$\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} \left(m(W_i,\theta) - \bar{m}(\theta)\right)\left(m(W_i,\theta) - \bar{m}(\theta)\right)^\top.$$

Then Assumptions 4.2 and 4.3 (part (ii)) are satisfied if, for some $\delta > 0$, the following "$2+\delta$" condition holds[6]:

$$\sup_{Q\in\mathcal{Q}} \mathbb{E}_F\left[\|m(W,\theta)\|^{2+\delta}\right] < \infty.$$

[6]For details, see Section 11.4.2 of Lehman and Romano (2005)

As in Example 2.1, we frame the hypothesis testing problem in 4.2 in terms of inference on the value of a linear program. 4.2 equivalent to testing whether the value of the LP in 2.2 is equal to 0 versus the alternative that $\nu > 0$. For our test statistic, we recommend the sample counterpart of 2.1, where we add additional "normalizing" inequalities to make the feasible region of the LP bounded, and the test scale-invariant. Our test is

$$\hat{T}_N := \max\left\{\lambda^\top \hat{b} \mid \lambda \in \mathbb{R}^p_+,\ \|\hat{D}\lambda\|_1 \leq 1\right\} = \|(\hat{D}^{-1}\hat{b})_+\|_\infty. \tag{4.3}$$

Note that alternative choices can be made for our normalization, leading to different test statistics. For instance, one can instead consider the test statistic

$$\tilde{T}_N := \max\left\{\lambda^\top \hat{b} \mid \lambda \in \mathbb{R}^p_+,\ \|\hat{D}\lambda\|_\infty \leq 1\right\} = \|(\hat{D}^{-1}\hat{b})_+\|_1.$$

As noted by Andrews and Soares (2010), these two tests direct power along different directions in the space of alternatives; as such, the choice of normalization is not inconsequential. Our particular choice of the normalization in $\hat{T}_N$ is motivated by the fact that it leads to a simpler asymptotic analysis.

The next assumption requires the availability of a uniformly consistent "bootstrap" estimate of the asymptotic distributions $\{\mathcal{N}(0, \Sigma(Q)) \mid Q \in \mathcal{Q}\}$.

Assumption 4.5. We have available a random variable ζ^*_N whose distribution is entirely determined by the realized sample $\mathbb{W}$ and which satisfies: for all $\epsilon > 0$,

$$\overline{\lim}_{N\to\infty} \sup_{Q\in\mathcal{Q}} Q\left(d_{\Pr}\Big(\mathcal{L}(\zeta^*_N \mid \mathbb{W}), \mathcal{N}(0, \Sigma(Q))\Big) > \epsilon\right) = 0.$$

Remark 4.6. Continuing with Remark 4.4, the arguments in Section 11.4.2 of Lehman and Romano (2005) (see also Theorem 2.4 in Romano and Shaikh (2012)) can be adapted to show that the same $2+\delta$ condition implies that Assumption 4.5 holds for the bootstrap statistic

$$\zeta^*_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \big(m(W^*_i, \theta) - \bar{m}(\theta)\big), \quad \text{where} \quad W^*_i \mid \mathbb{W}_N \overset{\text{i.i.d.}}{\sim} \hat{F}_N,$$

and $\hat{F}_N$ denotes the empirical distribution of the data $\mathbb{W}_N$.

The critical values of our test statistic $\hat{T}_N$ are computed as follows: Given a significance level $\alpha \in (0, 1/2)$, we reject the null hypothesis in favor of the alternative if and only if $\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)$, where $\hat{c}_N(1-\alpha)$ is defined by

$$\hat{c}_N(1-\alpha) := \inf\{u \mid \hat{H}_N(u) \geq 1-\alpha\}, \tag{4.4}$$

and $\hat{H}_N(\cdot) := \mathcal{L}(\nu^* | \mathbb{W})$ is the conditional (on $\mathbb{W}$) distribution of ν^*, where ν^* is the value of the following LP:

$$\nu^* = \max_{\lambda\in\hat{\Delta}} \lambda^\top \zeta^* \quad \text{where} \quad \hat{\Delta} := \left\{\lambda \in \mathbb{R}^p_+ \;\middle|\; \|\hat{D}\lambda\|_1 \leq 1,\ \hat{b}^\top\lambda \geq \hat{T}_N - \kappa_N/\sqrt{N}\right\}.[7] \tag{4.5}$$

[7]Here, as in Theorem 3.4, κ_N is any positive sequence such that $\kappa_N \to \infty$ and $\kappa_N = o(\sqrt{N})$.

The following theorem establishes that our testing procedure has correct asymptotic size in a uniform sense. Moreover, it shows that, pointwise for each $Q \in \mathcal{Q}_0$, the bootstrap distribution $\hat{H}_N$ converges, with probability approaching 1, to the asymptotic distribution of $\hat{T}_N$.

Theorem 4.7. *Suppose that Assumptions 4.2, 4.3, and 4.5 hold, and let $\hat{T}_N$ and $\hat{c}_N(\cdot)$ be defined as in equations 4.3 and 4.4, respectively. Then, for all $\alpha \in (0, 1/2)$, we have*

$$\overline{\lim}_{N\to\infty} \sup_{Q\in\mathcal{Q}_0} Q\left(\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\right) \le \alpha. \tag{4.6}$$

Moreover, for each $Q \in \mathcal{Q}_0$ and any $\epsilon > 0$, we have

$$\lim_{N\to\infty} Q\left(\left\{\mathbb{W}_N \,\middle|\, d_{\Pr}(\mathcal{L}(\sqrt{N}\hat{T}_N), \hat{H}_N) > \epsilon\right\}\right) = 0, \tag{4.7}$$

where $\hat{H}_N(\cdot)$ is the conditional distribution (given $\mathbb{W}_N$) of ν^* as defined in equation (4.5).

Remark 4.8. Note that computing or approximating the distribution $\hat{H}$ only requires solving the linear program in equation 4.5 for each realization of $\zeta^* \mid \mathbb{W}$. This can be done over a large number B of bootstrap iterations. The distribution $\hat{H}$ can then be approximated by the empirical distribution of the values $\{\nu_b^*\}_{b=1}^B$, where ν_b^* corresponds to the value of the LP 4.6 for the b^{th} bootstrap iteration.

Remark 4.9. The constraint $\hat{b}^\top\lambda \ge \hat{T}_N - \kappa_N/\sqrt{N}$ in equation 4.5, performs a sort of moment selection. Indeed, we can think of the null hypothesis as testing whether the moments $\lambda^\top b$, with λ in the feasible region of the LP, are nonpositive. The constraints then only keeps those values of λ that correspond to moments that are close to being optimal ($\kappa_N/\sqrt{N} = o(1)$) in equation 4.3, and only those moments are considered in the computation of $\hat{H}$. However, this moment selection procedure is not of the GMS type, as it does not involve an explicit shift of the components of ζ^* according to the degree of slackness of the corresponding inequalities in $\hat{b}$.

Remark 4.10. The second part of Theorem 4.7 (equation 4.7) implies that the inequality in equation 4.6 can be replaced by an equality if, for some $Q \in \mathcal{Q}_0$, the asymptotic distribution of $\sqrt{N}\hat{T}_N$ is nondegenerate. In Lemma 7.3, we derive this asymptotic distribution, and it is nondegenerate if and only if at least one of the inequalities in the constraint $\mu(Q) \le 0$ binds. This shows, in particular, that our test is not unnecessarily conservative. We also note that the derivation of the asymptotic distribution of $\sqrt{N}\hat{T}_N$ in Lemma 7.3 does not follow from Shapiro (1991), as the constraint qualifications required there are not necessarily satisfied in our setting. In particular, we do not require $\sqrt{N}(\hat{D} - D)$ to converge in distribution, or for the set of optimal solution of the LP to be bounded. Our derivation relies crucially on the fact that the value of the linear program is zero under H_0 and remains nonnegative under small perturbations—that is, under H_0, the LP value functional is at a local minimum.

We now provide some heuristics for the results. The complete proof is provided in the appendix. For $\mu = (b, D)$, let $\phi(\mu)$ denote the value of the (unperturbed) LP

$$\phi(\mu) := \max\{\lambda^\top b \mid \lambda \ge 0,\ \|D\lambda\|_1 \le 1\},$$

viewed as a function of its inputs b and D. For a given sample size N, we consider perturbations $\xi = (\xi_b, \xi_D)$ of μ, where ζ_b is a vector of the same dimension as b and ζ_D is a diagonal matrix of the same dimension as D, and such that $\phi(\mu + \xi/\sqrt{N})$ remains finite. Denote the perturbed inputs by $b_N(\xi) = b + \xi_b/\sqrt{N}$, with a similar definition for $D_N(\xi)$, and set $\mu_N(\xi) = (b_N(\xi), D_N(\xi))$.

Note that $\hat{T}_N = \phi(\mu_N(\hat{\xi}))$, where $\hat{\xi} = \left(\sqrt{N}(\hat{b} - b), \sqrt{N}(\hat{D} - D)\right)$. Let $T_N(\xi) := \phi(\mu_N(\xi))$, and let $L_N(\lambda, x; \xi)$ denote the Lagrangian function for the LP corrresponding to the inputs $\mu_N(\xi)$ and given by

$$L_N(\lambda, x; \xi) = b_N(\xi)^\top \lambda + x\left(1 - \mathbb{1}^\top D_N(\xi)\lambda\right)$$

where x and λ are respectively a nonnegative scalar and a vector with nonnegative entries, and x represent a Lagrange multiplier. Let $L(\lambda, x)$ denote analogously the Lagrangian for the unperturbed LP:

$$L(\lambda, x) = b^\top \lambda + x\left(1 - \mathbb{1}^\top D\lambda\right).$$

Let Δ_0 and S_0 respectilvely denote the sets of optimal primal and dual solutions for the LP that corresponds to the unperturbed LP. Similarly, let $\Delta_{0,N}(\xi)$ and $S_{0,N}(\xi)$ denote the corresponding optimal solution sets for the perturbed LP with inputs $\mu_N(\xi)$.

The starting point in the derivation of the uniform validity of our procedure, is the following inequality that uses the saddle point property of the Lagrangian to upper bound our test statistic: Let $(\lambda_0, x_0) \in \Delta_0 \times S_0$ and $(\lambda_\xi, x_\xi) \in \Delta_{0,N}(\xi) \times S_{0,N}(\xi)$. Under H_0, $\phi(\mu) = 0$, and we have

$$\begin{aligned}
\sqrt{N}T_N(\xi) &= \sqrt{N}(\phi(\mu_N(\xi)) - \phi(\mu)) = \sqrt{N}\big(L_N(\lambda_\xi, x_\xi; \xi) - L(\lambda_0, x_0)\big) \\
&\leq \sqrt{N}\big(L_N(\lambda_\xi, x_0; \xi) - L(\lambda_\xi, x_0)\big) \\
&= \left(\sqrt{N}\big(b_N(\xi) - b\big)\right)^\top \lambda_\xi - x_0 \mathbb{1}^\top \Big(D_N(\xi) - D\Big)\lambda_0
\end{aligned}$$

As x_0 is any arbitrary dual solution, under H_0, it is easy to check that $S_0 = \{0\}$ [8]. Making the substitution $x_0 = 0$ in the last inequality yields

$$\sqrt{N}T_N(\xi) \leq \left(\sqrt{N}\big(b_N(\xi) - b\big)\right)^\top \lambda_\xi. \tag{4.8}$$

In the proof of Theorem 4.7, we show that, starting from equation 4.8, it is possible to establish

$$\sqrt{N}T_N(\tilde{\xi}) \leq \sup_{\lambda \in \hat{\Delta}} \lambda^\top \tilde{\zeta}_b + o_p(1), \tag{4.9}$$

where the $o_p(1)$ term is uniform over $Q \in \mathcal{Q}_0$, $\hat{\Delta}$ is defined as in equation 4.5, and $\tilde{\xi} = (\tilde{\xi}_b, \tilde{\xi}_D)$ is an independent copy of $\hat{\xi} = (\sqrt{N}(\hat{b} - b), \sqrt{N}(\hat{D} - D))$. A coupling argument is then used to replace $\tilde{\xi}_b$ on the right-hand side of the inequality with its bootstrap counterpart ζ_N^* from Assumption 4.5, leading to the

[8]This property of the independence of the Lagrange multiplier with respect to the null DGP is in part what motivates our choice of normalization. For alternative normalizations, for instance using the constraint $\|\hat{D}\lambda\|_1 = 1$, the Lagrange multiplier may depend on the null DGP, which complicates the theory.

derivation of inequality 4.6. The details are provided in Appendix A (7.0.1).

4.2 Moment Inequalities with Nuisance Parameters Entering Linearly

In this section, we develop a uniformly valid test for the hypothesis described in Example 2.2, which we restate as

$$H_0 : Q \in \mathcal{Q}_0 \quad \text{versus} \quad H_1 : Q \in \mathcal{Q} \setminus \mathcal{Q}_0, \tag{4.10}$$

where

$$\mathcal{Q}_0 = \left\{ Q \in \mathcal{Q} \mid \text{ there exists } \eta \in \mathbb{R}^d \text{ such that } b(Q) - A\eta \leq 0 \right\}.$$

We consider a setting in which the matrix A is known, while some components of $b(Q)$ are unknown and must be estimated from the data. The testing procedure we propose below is a slight modification of the method introduced in Section 4.1. Our inference procedure is uniformly valid for large families of data-generating processes that satisfy the assumptions outlined below. Throughout, we focus on a semi high-dimensional setting in which the dimension of the unknown components of b remains bounded, while the dimension of the deterministic component of b, as well as the dimension of the nuisance parameter are allowed to vary arbitrarily, as the sample size increases. The restrictions that we impose on the dimensions of the problem are summarized in Assumption 4.11 below.

Let $b(Q) = (b^u(Q), b^k(Q)) \in \mathbb{R}^p$, where $b^k(Q) \in \mathbb{R}^{p_k}$ represents the deterministic (known) component of b, and $b^u(Q) \in \mathbb{R}^{p_u}$ represents the unknown component of $b(Q)$.

Assumption 4.11. We assume that p_u remains bounded as N increases, with all other dimensions of the problem allowed to vary arbitrarily with N.[9]

To connect the setting of Example 2.2 to the hypothesis in equation (4.23), let $b(Q) = \mathbb{E}_F[h(W, \theta)]$, where, as in Remark 4.1, $\theta \in \Theta$ and $W \sim F$, with $F \in \mathcal{F}$. Let $\mathcal{A}$ denote a set of admissible coefficient matrices A. We then define the sets $\mathcal{Q}$ and $\mathcal{Q}_0$ as follows:

$$\mathcal{Q} = \{(\theta, A, F) \mid \theta \in \Theta,\ A \in \mathcal{A},\ F \in \mathcal{F}\}, \quad \text{and} \quad \mathcal{Q}_0 = \left\{ Q = (\theta, A, F) \in \mathcal{Q} \mid \ \exists \eta \in \mathbb{R}^d \text{ such that } b(Q) - A\eta \leq 0 \right\}.$$

As indicated in Example 2.2, this formulation covers some tests that arise from the conditional moment inequality framework considered by Andrews, Roth, and Pakes (2023) and Cox and Shi (2023), who study hypotheses of the form

$$H_0 : \ \exists \eta \in \mathbb{R}^d \text{ such that } \mathbb{E}[h(W, \theta_0) \mid Z] - A(Z, \theta)\eta \leq 0,$$

where Z denotes observed instrumental variables (with Z_i being a component of W_i). Here, the nuisance parameter η enters the moment inequality linearly with a coefficient matrix $A(Z, \theta)$ that may depend on both θ and Z, and is thus deterministic once we condition on the instruments. As the inequality defining

[9]That is, we allow p_k to vary arbitrarily with N, and, if $A \in \mathbb{R}^{p \times d}$, then d is also allowed to vary arbitrarily with N; however, it is without loss of generality to assume $d \leq p$, since the hypothesis 4.23 remains unchanged if A is replaced by a subset of its columns that spans its column space.

the null remains valid when it is multiplied through by a nonnegative function of the instrument, Andrews, Roth, and Pakes (2023) and Cox and Shi (2023) consider for their tests a set of inequalities that result when the function $h(\cdot)$ and matrix A are interacted with a nonnegative vector-valued function of the instruments, and then conduct inference conditional on the realized values of the instrument $\{Z_1, \ldots, Z_N\}$. By working conditional on the realized value of the instruments, the implication of the null that Andrews, Roth, and Pakes (2023) and Cox and Shi (2023) eventually test are hypotheses that have the unconditional moment inequalities form described in equation (4.23).

As in Section 4.1, suppose we have at our disposal estimators $\hat{b}$ and $\hat{D}$ of b and D (the matrix of asymptotic standard deviations), which satisfy Assumption 4.2 and Assumption 4.12 below. Assumption 4.12 modifies Assumption 4.3 to accommodate settings in which $b = (b^k, b^u)$ may include a deterministic component b^k, which associated diagonal entries in the asymptotic standard deviation matrix are equal to zero. The assumption in part requires that the diagonal entries of the asymptotic standard deviations matrix corresponding to the unknown component b^u of b are uniformly bounded away from zero and infinity.

Assumption 4.12. i) There exist positive constants $\underline{\sigma}$ and $\bar{\sigma}$, with $\underline{\sigma} < \bar{\sigma}$, such that for all $Q \in \mathcal{Q}$, all diagonal elements of $D(Q)$ that correspond to the estimated component of b lie in the interval $[\underline{\sigma}, \bar{\sigma}]$.

ii) For all $\epsilon > 0$,

$$\overline{\lim}_{N\to\infty} \sup_{Q\in\mathcal{Q}} Q\left(\|\hat{D} - D(Q)\| > \epsilon\right) = 0.$$

To compute the critical values of our test, we assume access to a bootstrap-like estimator ζ_N^* for the asymptotic distribution of $\sqrt{N}(\hat{b} - b)$, which satisfies Assumption 4.5.

Following Example 2.2, our hypothesis can be expressed in terms of inference on the value of a linear program. Indeed, 4.23 is equivalent to testing whether $\nu = 0$ vs. $\nu > 0$, where ν denotes the value of the linear program:

$$\nu = \max\{\lambda^\top b \mid \lambda \in \mathbb{R}^p_+,\ A^\top\lambda = 0,\ \|D\lambda\|_1 \leq 1\}. \tag{4.11}$$

This naturally leads to the test statistic:

$$\hat{T}_N = \max\{\lambda^\top \hat{b} \mid \lambda \in \mathbb{R}^p_+,\ A^\top\lambda = 0,\ \|\hat{D}\lambda\|_1 \leq 1\} = \min_{\eta\in\mathbb{R}^d} \|[\hat{D}^{-1}(\hat{b} - A\eta)]_+\|_\infty \tag{4.12}$$

where the last equality only holds when $\hat{D}$ is invertible (thus all components of b are estimated). As in Section 4.1, normalization serves two purposes: it ensures that the set of values of the component of λ that corresponds to the estimated part of b, for λ in the feasible region of the LP, is bounded. Moreover, it renders the test statistic scale-invariant by accounting for heteroskedasticity through $\hat{D}$. Note however, that when some components of b are deterministic, in contrast to the analysis of Section 4.1, it is possible here for the value of the LP in 4.12, and thus for our test statistic $\hat{T}_N$, to be infinite.[10] However, when N is sufficiently large that $\hat{D}$ and D have the same nonzero diagonal elements, Farkas' Lemma implies that it can occur if and only if there exists $\lambda \in \mathbb{R}^p_+$ such that $A^\top\lambda = 0$, the component of λ corresponding to the estimated

[10] This possibility is not considered in Andrews, Roth, and Pakes (2023).

component of b in zero, and $\lambda^\top b > 0$. In other words, when this occurs, it is possible to reject the null hypothesis solely based on its implications for the deterministic components of b.

Our test differs from that proposed in Andrews, Roth, and Pakes (2023) only in that we normalize using a weak inequality constraint ($\|\hat{D}\lambda\|_1 \leq 1$), whereas they use a strict equality ($\|\hat{D}\lambda\|_1 = 1$). We prefer the weak inequality formulation for two reasons. First, it simplifies the analysis: under the null, the value function of the LP is insensitive to estimation of the D matrix (see Remark 4.10). Second, it guarantees that the feasible region is always non-empty (as it always contains the origin), which may not hold when the strict normalization $\|\hat{D}\lambda\|_1 = 1$ is imposed.[11]

The computation of the critical values proceeds as in Section 4.1. Given a significance level $\alpha \in (0, 1/2)$, we reject the null hypothesis in favor of the alternative if and only if $\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)$, where $\hat{c}_N(1-\alpha)$ is defined by

$$\hat{c}_N(1-\alpha) := \inf\{u \mid \hat{H}_N(u) \geq 1-\alpha\}, \tag{4.13}$$

and $\hat{H}_N(\cdot) := \mathcal{L}(\nu^*|\mathbb{W})$ is the conditional (on $\mathbb{W}$) distribution of ν^*, where ν^* is the value of the following LP:

$$\nu^* = \max_{\lambda \in \hat{\Delta}} \lambda^\top \zeta^* \quad \text{where} \quad \hat{\Delta} := \left\{\lambda \in \mathbb{R}^p_+ \;\middle|\; A^\top\lambda = 0,\ \|\hat{D}\lambda\|_1 \leq 1,\ \hat{b}^\top\lambda \geq \hat{T}_N - \kappa_N/\sqrt{N}\right\}, \tag{4.14}$$

and where $\kappa_N = o(\sqrt{N})$, $\kappa_N \to \infty$.

To establish the uniform validity of our procedure, we consider an additional assumption, which puts restrictions on the relationship between the coefficient matrix A and the asymptotic variance matrix $\Sigma(Q)$ for DGPs in the null family $\mathcal{Q}_0$. Assumption 4.13 below ensures that the asymptotic distribution of our test statistic does not "concentrate too much" along arbitrary sequences of DGPs $Q_N \in \mathcal{Q}_0$, which could otherwise undermine the uniform validity of our procedure. It is possible however to show that whithout this assumption, our method remains valid if we modify the rejection rule and instead reject iff $\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha) + \tau$, for some fixed arbitrary $\tau > 0$.

Before stating the assumption, we introduce some additional notation. Let $extr(\mathcal{D}(Q))$ denote the set of extreme points of the feasible region of the LP 4.11, defined by

$$\mathcal{D}(Q) := \left\{\lambda \in \mathbb{R}^p_+ \,\middle|\, A^\top\lambda = 0,\ \|D\lambda\|_1 \leq 1\right\}.$$

Let $\Delta_0(Q)$ denote the subset of $\mathcal{D}(Q)$, given by the optimal solutions to the linear program in equation (4.11).[12]

Assumption 4.13. There exists a constant $\rho > 0$ such that for all $Q \in \mathcal{Q}_0$:

i) If $\Delta_0(Q) \neq \{0\}$, then there exists $\lambda \in \Delta_0(Q)$ such that $\lambda^\top\Sigma(Q)\lambda \geq \rho$.

ii) If $\Delta_0(Q) = \{0\}$ and $extr(\mathcal{D}(Q)) \setminus \{0\} \neq \emptyset$, then for all $\lambda \in extr(\mathcal{D}(Q)) \setminus \{0\}$, we have $\lambda^\top\Sigma(Q)\lambda \geq \rho$.

Assumption 4.13 is trivially satisfied if the variance matrices $\Sigma(Q)$ are non-singular, with their eigenvalues

[11] However, as noted in Andrews, Roth, and Pakes (2023), the feasible region is empty under the normalization iff the matrix A has full range, in which case the null always holds (for any b).

[12] Note that $extr(\mathcal{D}(Q))$ is non-empty as it always contains zero.

uniformly bounded away from zero and infinity.[13] While Assumption 4.13 is analogous to Assumption 4 in Andrews, Roth, and Pakes (2023), it is weaker.[14] In particular, our assumption allows for the possibility that asymptotically, with positive probability, two distinct vertices arise as optimal solutions of the sample LP 4.12, a posibility that is ruled out by Assumption 4 in Andrews, Roth, and Pakes (2023), and which can arise when $\Sigma(Q)$ is singular. The approach in Cox and Shi (2023) does not impose this type of assumption. Instead, as in Andrews, Roth, and Pakes (2023), it assumes that in the case of degeneracy (when $\Sigma(Q)$ is singular), the structure of the degeneracy is *known*. Specifically, if $b(Q) = \mathbb{E}[m(W, \theta)]$ and $\Sigma(Q) = \text{Var}(m(W, \theta))$ is singular, then there exists a *known* matrix B and a function g such that $b(Q) = B\mathbb{E}[g(W, \theta)]$, where $\text{Var}(g(W, \theta))$ is non-singular, with eigenvalues uniformly bounded away from zero and infinity. By contrast, in the degenerate case, our framework does not require the structure of the linear dependence among the components of b to be known.

We now state the main result of this section. The proof is in Section 7.0.1 of the appendix.

Theorem 4.14. *Suppose that Assumptions 4.11, 4.2, 4.5, 4.12, and 4.13 hold, and let $\hat{T}_N$ and $\hat{c}_N(\cdot)$ be defined as in equations 4.12 and 4.13, respectively. Then, for all $\alpha \in (0, 1/2)$, we have*

$$\overline{\lim}_{N\to\infty} \sup_{Q \in \mathcal{Q}_0} Q\left(\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\right) \leq \alpha. \tag{4.15}$$

Moreover, for each $Q \in \mathcal{Q}_0$ and any $\epsilon > 0$, we have

$$\lim_{N\to\infty} Q\left(\left\{\mathbb{W}_N \;\middle|\; d_{\Pr}(\mathcal{L}(\sqrt{N}\hat{T}_N), \hat{H}_N) > \epsilon\right\}\right) = 0, \tag{4.16}$$

where $\hat{H}_N(\cdot)$ is the conditional distribution (given $\mathbb{W}_N$) of ν^* as defined in equation (4.14).

Remark 4.15. As in Remark 4.10, equation 4.16 implies that the inequality in equation 4.15 can be replaced by an equality if, for some $Q \in \mathcal{Q}_0$, the asymptotic distribution of $\sqrt{N}\hat{T}_N$ is nondegenerate. The asymptotic distribution is derived in Lemma 7.3, and it is nondegenerate if and only if the LP 4.11 has at least one optimal solution $\lambda \in \Delta_0(Q)$ such that its component λ^u that corresponds to the estimated part of b is not zero. Thus our approach yields pointwise asymptotic exact size, for all DGPs on the "boundary" of the null region $\mathcal{Q}_0$. By contrast, the LF (least favorable) and hybrid tests of Andrews, Roth, and Pakes (2023) have exact pointwise asymptotic size only for DGPs where all feasible λ are optimal (i.e., $\Delta_0(Q) = \mathcal{D}(Q)$)[15], which (By Farkas' Lemma) is not possible in settings where for each $Q \in \mathcal{Q}_0$, the (reverse) inequality $b(Q) - A\eta \geq 0$ does not have a solution η.[16] We refer the reader to Andrews, Roth, and Pakes (2023) for details.

Remark 4.16. Under Assumption 4.11, since we allow p to vary arbitrarily with N as long as p_u remains bounded, Assumption 4.2 is mild—it only requires uniform asymptotic normality for the estimated component $b^u(Q)$ of b, which has bounded dimension. In this respect, our result is stronger than those of Andrews,

[13]In that case, the nonzero extreme points are uniformly bounded away from zero as they satisfy $\|D\lambda\|_1 = 1$, and thus $\|\lambda\|_1 \geq 1/\underline{\sigma}$.

[14]As in Andrews, Roth, and Pakes (2023), in part ii) of the assumption, it suffices to only consider vertices that occur with positive probability as optimal solutions of the sample LP, when $N \to \infty$.

[15]Equivalently, this corresponds to DGPs for which there exists a value of the nuisance parameter η such that $b(Q) - A\eta = 0$.

[16]In the moment inequality setting (where $A = 0$) this is for instance the case if all moment inequalities cannot bind simulataneously, i.e., the inequality $b(Q) \leq 0$, never holds as an equality for all $Q \in \mathcal{Q}_0$.

Roth, and Pakes (2023) and Cox and Shi (2023), who establish uniform validity of their procedures in settings where all dimensions of the problem are bounded with respect to N.

Remark 4.17. The proofs of Theorems 4.7 and 4.14 are very similar. Note that when all components of b are unknown and must be estimated, the test statistic in (4.12) coincides with the test statistic from Section 4.1 when $A = 0$. The main additional complication in the proof of Theorem 4.14 arises from the fact that when some components of b are known, the matrices $\hat{D}$ are no longer guaranteed to be non-singular, and the feasible region $\hat{\mathcal{D}}$ of the linear program in (4.14) may be unbounded.

4.3 Inference on Linear Systems with Known Coefficients

In this section, we develop a uniformly valid test for the hypothesis presented in Example 2.3, which we restate as

$$H_0 : Q \in \mathcal{Q}_0 \quad \text{versus} \quad H_1 : Q \in \mathcal{Q} \setminus \mathcal{Q}_0, \tag{4.17}$$

where

$$\mathcal{Q}_0 = \left\{ Q \in \mathcal{Q} \mid \text{ there exists } \eta \in \mathbb{R}^d_+ \text{ such that } b(Q) = A\eta \right\}.$$

We consider a setting in which the matrix A ($\in \mathbb{R}^{p \times d}$) is known, while some components of $b(Q)$ ($\in \mathbb{R}^p$) are unknown and must be estimated from the data. The testing procedure we propose is a slight modification of the method introduced in Section 4.2. In fact, through suitable transformations and the use of auxiliary variables, one can show that the hypotheses are equivalent. Specifically, the null hypothesis H_0 in equation 4.18 holds with $\eta \geq 0$ if and only if η satisfies

$$\tilde{b}(Q) - \tilde{A}\eta \leq 0,$$

where

$$\tilde{b}(Q) = \begin{pmatrix} 0 \\ b(Q) \\ -b(Q) \end{pmatrix}, \quad \tilde{A} = \begin{pmatrix} \mathbb{I}_d \\ A \\ -A \end{pmatrix}.$$

On the other hand, the hypothesis H_0 in 4.23 holds if and only if there exists $x \geq 0$ such that

$$\tilde{A}x = b,$$

with

$$\tilde{A} = \begin{pmatrix} A & -A & \mathbb{I}_p \end{pmatrix}.$$

The inference procedure we describe below could be derived by applying the approach from Section 4.2 to the equivalent hypothesis

$$H_0 : \ \exists\, \eta \geq 0 \quad \text{such that} \quad \tilde{b}(Q) - \tilde{A}\eta \leq 0,$$

and the uniform validity of the resulting test would hold, given that the assumptions from that section hold for the transformed parameters $(\tilde{b}(Q), \tilde{A}(Q))$. However, to facilitate a direct comparison with the results in Fang et al. (2023), we choose instead to state our assumptions in terms of the original parameters $(b(Q), A(Q))$. For illustrative applications that motivate consideration of hypothesis 4.18, we refer the reader to Fang et al. (2023).

As in Section 4.2, we focus on a semi high-dimensional setting in which the dimension of the unknown components of b remains bounded, while the dimension of the deterministic component of b and the number of columns of the matrix A are allowed to vary arbitrarily, as the sample size increases. By contrast, the method of Fang et al. (2023) allows the dimension of the estimated component of b to grow with N, although they do not formally consider, as we do, the possibility that the dimension of the deterministic component of b varies arbitrarily with N. The restrictions that we impose on the dimensions of the problem are summarized in Assumption 4.18 below.

Let $b(Q) = (b^u(Q), b^k(Q)) \in \mathbb{R}^p$ and p_u be as in the statement of Assumption 4.11.

Assumption 4.18. We assume that p_u remains bounded as N increases, with all other dimensions of the problem allowed to vary arbitrarily with N.[17]

As in Section 4.2, we assume the availability of estimators $\hat{b}$ and $\hat{D}$ for b and D, respectively, which satisfy Assumptions 4.2 and 4.12. In addition, we assume access to a bootstrap-like estimator ζ_N^* for the asymptotic distribution of $\hat{b}$, which satisfies Assumption 4.5.

Following the linear programming characterization developed in Example 2.3, testing the hypothesis H_0 (vs H_1) in 4.18, is equivalent to testing whether $\nu = 0$ (vs $\nu > 0$), where ν represent the value of the linear program

$$\nu = \max\{b^\top \lambda \mid A^\top \lambda \leq 0,\ \|D\lambda\|_1 \leq 1\}.$$

Our test statistic is given by the value of the sample analogue of the foregoing LP:

$$\hat{T}_N := \max\{\hat{b}^\top \lambda \mid A^\top \lambda \leq 0,\ \|\hat{D}\lambda\|_1 \leq 1\} = \min_{x \geq 0} \|\hat{D}^{-1}(\hat{b} - Ax)\|_\infty \tag{4.18}$$

where the last equality only holds when $\hat{D}$ is invertible, and follows from duality. We reject the null hypothesis in favor of the alternative whenever $\sqrt{N}\hat{T}_N$ is large. Specifically, given a significance level $\alpha \in (0, 1/2)$, we reject the null hypothesis in favor of the alternative if and only if $\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)$, where $\hat{c}_N(1-\alpha)$ is defined by

$$\hat{c}_N(1-\alpha) := \inf\{u \mid \hat{H}_N(u) \geq 1 - \alpha\}, \tag{4.19}$$

and $\hat{H}_N(\cdot) := \mathcal{L}(\nu^* | \mathbb{W})$ is the conditional (on $\mathbb{W}$) distribution of ν^*, where ν^* is the value of the following LP:

$$\nu^* = \max_{\lambda \in \hat{\Delta}} \lambda^\top \zeta_N^* \quad \text{where} \quad \hat{\Delta} := \left\{\lambda \in \mathbb{R} \;\middle|\; A^\top \lambda \leq 0,\ \|\hat{D}\lambda\|_1 \leq 1,\ \hat{b}^\top \lambda \geq \hat{T}_N - \kappa_N/\sqrt{N}\right\}. \tag{4.20}$$

Here, ζ_N^* is the bootstrap-like estimator of the asymptotic distribution of $\hat{b}$ given in Assumption 4.5, and κ_N is a tuning parameter that satisfies $\kappa_N = o(\sqrt{N})$ and $\kappa_N \to \infty$.

[17] Note that unlike the null hypothesis 4.23, where we can always assume w.l.o.g. that $d \leq p$, it is possible here for "simplest" representation of the null to have d much larger than p. This is because the null 4.18 requires $b(Q)$ to be in the set cone(A), the conic hull of the columns of A, which can be a set with a large number of facets.

Here, as in Section 4.2, the use of the normalization $\|\hat{D}\lambda\|_1 \leq 1$ serves two purposes: It ensures that the component of λ which corresponds to the estimated part of b remains bounded, for values of λ in the feasible region of our LPs. More importantly, the use of the asymptotic standard deviation matrix in our normalization makes our tests scale invariant. Note however, that when some components of b are known, and the corresponding diagonal components of D (and $\hat{D}$) are equal to zero, the feasible regions $\hat{\mathcal{D}} := \{\lambda \in \mathbb{R}^p \mid A^\top \lambda \leq 0,\ \|\hat{D}\lambda\|_1 \leq 1\}$ may be unbounded, and it is possible to have $\hat{T}_N = \infty$. However, when the latter occurs, it is neccesarily the case (by Farkas' Lemma) that there exists $\lambda \in \mathbb{R}^p$ such that $\lambda^u = 0$ (λ^u being the component of λ that corresponds to b^u), $A^\top \lambda \leq 0$ and $\hat{b}^\top \lambda = \langle b^k, \lambda^k \rangle > 0$. That is, when $\hat{T}_N = \infty$, it is possible to reject the null hypothesis solely by focusing on the implication of the null for the deterministic component b^k of b without recourse to the data.[18] Finally, we would like to point out that given the dual characterization of our test statistic in equation 4.18, our test can be viewed as the scale invariant counterpart to the test proposed in Bai, Santos, and Shaikh (2022).

To establish the uniform validity of our procedure, we will impose Assumption 4.19 below, which is analogous to Assumption 4.13. The modification of Assumption 4.13 is necessary in the present context because, unlike the setting of Section 4.2—where the feasible regions always contain at least one extreme point (the origin)—in the current setting, the presence of a deterministic component in the vector b can lead to feasible regions $\mathcal{D}(Q) := \{\lambda \in \mathbb{R}^p \mid A^\top \lambda \leq 0,\ \|D\lambda\|_1 \leq 1\}$ that have no extreme points.[19] However, as in Section 4.2, our approach remains uniformly valid without Assumption 4.19 below, if we modify the rejection rule and reject iff $\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha) + \tau$, for some fixed arbitrary $\tau > 0$.

Before stating Assumption 4.19, we introduce some additional notation. Let $[p] = K \cup U$ be a partition of $[p]$, the set of indices corresponding to the coordinates of b, where K denotes the indices of the deterministic components and U denotes the indices of the components to be estimated. Define $\Pi^u : \mathbb{R}^p \to \mathbb{R}^p$ as the projection onto the subspace spanned by the cannonical basis vectors $\{e_i \mid i \in U\}$, and given by

$$[\Pi^u(\lambda)]_i = \lambda_i \ \forall i \in U \qquad \text{and} \qquad [\Pi^u(\lambda)]_i = 0 \ \forall i \in K.$$

We let $\mathcal{L}(Q)$ denote the *lineality space* of $\mathcal{D}(Q)$, for $Q \in \mathcal{Q}$, given by

$$\mathcal{L}(Q) := \{\lambda \in \mathbb{R}^p \mid \Pi^u(\lambda) = 0,\ A^\top \lambda = 0\}.$$

Let $\mathcal{L}(Q)^\perp$ denote the orthogonal complement of the subspace $\mathcal{L}(Q)$. The key facts regarding the lineality space are as follows: if $\mathcal{D}(Q)$ contains at least one extreme point, then $\mathcal{L}(Q) = \{0\}$ and $\mathcal{L}(Q)^\perp = \mathbb{R}^p$. Furthermore, the sets $\mathcal{D}(Q) \cap \mathcal{L}(Q)^\perp$ always contain at least one extreme point.[20] Our assumption will be formulated in terms of the extreme points of the sets $\mathcal{D}(Q) \cap \mathcal{L}(Q)^\perp$.

[18] This is akin to a situation in the moment inequality setting in which we aim to test the hypothesis $b(Q) \leq 0$, but where the deterministic component b^k of b contains a positive entry. In such a case, we can reject the null hypothesis immediately, without requiring further analysis.

[19] For example, let $b = \begin{pmatrix} 0 \\ b_2 \end{pmatrix} \in \mathbb{R}^2$ (i.e., first coordinate is known and equal to zero) and $A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. If the asymptotic variance of $\hat{b}_2$ is 1, then the feasible region is $\mathcal{D}(Q) = \{\lambda \in \mathbb{R}^2 \mid -1 \leq \lambda_2 \leq 0\}$, which is a horizontal strip, and has no extreme points.

[20] See Chapter 8 of Schrijver (1999) for the definition and properties of the lineality space.

Assumption 4.19. There exists a constant $\rho > 0$ such that for all $Q \in \mathcal{Q}_0$:

i) If $\Delta_0(Q) \neq \{0\}$, then there exists $\lambda \in \Delta_0(Q)$ such that

$$\lambda^\top \Sigma(Q)\lambda \geq \rho.$$

ii) If $\Delta_0(Q) = \{0\}$ and $\operatorname{extr}(\mathcal{D}(Q) \cap \mathcal{L}(Q)^\perp) \setminus \{0\} \neq \emptyset$, then

$$\lambda^\top \Sigma(Q)\lambda \geq \rho \quad \text{for all} \quad \lambda \in \operatorname{extr}(\mathcal{D}(Q) \cap \mathcal{L}(Q)^\perp) \setminus \{0\}.$$

Heuristically, Assumptions 4.13 and 4.19 are needed to guarantee that on some events $E_N(Q)$ that we construct in our proofs and which satisfy

$$\underline{\lim}_{N\to\infty} \inf_{Q \in \mathcal{Q}_0} Q(E_N(Q)) = 1,$$

and for all sufficiently large N, whenever $\hat{\vartheta} > 0$ occurs (which is a necessary condition to reject the null hypothesis), it must be the case that $\hat{\Delta}$ contains an element λ such that $\lambda^\top \Sigma(Q)\lambda > \rho/2$. The existence of such a λ is what is needed in our proofs to guarantee that the distributions $\hat{H}$ do not concentrate (see Proposition 7.2), and any assumption that can guarantee the existence of such λ whenever rejection occurs, would equally imply the uniform validity of our procedure. In our proofs, we show that on the events $E_N(Q)$, the sets $\Delta_0(Q)$ are at an asymptotically negligeable distance from the set $\hat{\Delta}$, and thus part i) of the assumption guarantees the existence of a $\lambda \in \hat{\Delta}$ that satisfies the desired condition. Moreover, when $\Delta_0(Q) = \{0\}$, and part i) cannot hold, we use the fact that when a rejection occurs, the value $\hat{\vartheta} > 0$ is achieved at a point in $\operatorname{extr}(\hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^\perp) \setminus \{0\}$. In Lemma 7.6, we show that the sets $\operatorname{extr}(\hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^\perp)$ and $\operatorname{extr}(\mathcal{D}(Q) \cap \mathcal{L}(Q)^\perp)$ are asymptotically arbitrarily close in the Hausdorff sense, and thus part ii) of the assumption guarantees that the optimal λ satisfies the condition $\lambda^\top \Sigma(Q)\lambda > \rho/2$ for all sufficiently large N. We refer the reader to the proofs for details.

Assumption 4.19 is not directly comparable to the assumptions in Fang et al. (2023). However, a more direct comparison becomes possible if we impose restrictions on the support of $\hat{b}$, as done in Assumption 4.3 of Fang et al. (2023). By adopting similar support restrictions (as we do below), we can replace Assumption 4.19 with Assumption 4.20, which aligns more closely with the framework in Fang et al. (2023). The support condition in Assumption 4.20 should be satisfied whenever the parameter b represents a mean vector.

Assumption 4.20. We assume that the following support restriction holds

$$\underline{\lim}_{N\to\infty} \inf_{Q \in \mathcal{Q}} Q\Big\{\big(\hat{b} - b(Q)\big\} \in \operatorname{Range}(\Sigma(Q))\Big) = 1.$$

We assume that there exists $\rho > 0$ such that for all $Q \in \mathcal{Q}_0$ and λ such that $\lambda \in \operatorname{extr}(\mathcal{D}(Q) \cap \mathcal{L}(Q)^\perp)$, then

$$\text{Either} \quad \lambda^\top \Sigma(Q)\lambda = 0 \qquad \text{or} \qquad \lambda^\top \Sigma(Q)\lambda \geq \rho.$$

Assumption 4.20 is similar to the assumptions made in Theorem 4.2 of Fang et al. (2023). Note, however, that the assumptions in Fang et al. (2023) are stated under a deterministic normalization (i.e., $\hat{D}$ is replaced by a deterministic matrix). An analogous assumption for the case in which the normalizing matrix is estimated was considered in an earlier version of their paper, but involved conditions based on the extreme points of certain random polytopes. In contrast, our framework relies solely on conditions involving the extreme points of population polytopes. Lemma 7.6, which can be viewed as a generalization of Lemma 4 in Andrews, Roth, and Pakes (2023), enables us to show that the extreme points of the sample polytopes remain close to those of their population counterparts.

To build intuition, the support restrictions ensure that when rejection occurs (and thus $\hat{\nu} > 0$), the corresponding optimal λ cannot be an extreme point of $\hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^{\perp}$ satisfying $\lambda^{\top}\Sigma(Q)\lambda = 0$. Specifically, the support restriction guarantees that (with high probability) $\lambda^{\top}\hat{b} = \lambda^{\top}b$ for such extreme points. Under H_0, we have $\hat{b}^{\top}\lambda \leq 0$ for all $\lambda \in \mathcal{D}$, and therefore also for $\lambda \in \hat{\mathcal{D}}$ (by homogeneity). Hence, the support restriction in Assumption 4.20 implies that, for sufficiently large N, whenever rejection occurs, there must exist an element of $\operatorname{extr}(\hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^{\perp}) \setminus \{0\}$ that is optimal (and thus lies in the set $\hat{\Delta}$) and satisfies $\lambda^{\top}\Sigma(Q)\lambda > \rho/2$. This condition is sufficient to ensure the required anti-concentration.

We now state the main theorem of this section. A proof is provided in Section 7.0.1 of the appendix.

Theorem 4.21. *Suppose that Assumptions 4.18, 4.2, 4.5, 4.12, and 4.19 (or 4.20) hold, and let $\hat{T}_N$ and $\hat{c}_N(\cdot)$ be defined as in equations 4.12 and 4.13, respectively. Then, for all $\alpha \in (0, 1/2)$, we have*

$$\overline{\lim}_{N\to\infty} \sup_{Q\in\mathcal{Q}_0} Q\left(\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\right) \leq \alpha. \tag{4.21}$$

Moreover, for each $Q \in \mathcal{Q}_0$ and any $\epsilon > 0$, we have

$$\lim_{N\to\infty} Q\left(\left\{\mathbb{W}_N \;\middle|\; d_{\Pr}(\mathcal{L}(\sqrt{N}\hat{T}_N), \hat{H}_N) > \epsilon\right\}\right) = 0, \tag{4.22}$$

where $\hat{H}_N(\cdot)$ is the conditional distribution (given $\mathbb{W}_N$) of ν^* as defined in equation (4.20).

4.4 Inference on Linear Systems with Unknown Coefficients

In this section, we develop a uniformly valid test for the hypothesis described in Example 2.4, which we restate as

$$H_0 : Q \in \mathcal{Q}_0 \quad \text{versus} \quad H_1 : Q \in \mathcal{Q} \setminus \mathcal{Q}_0,^{21} \tag{4.23}$$

where

$$\mathcal{Q}_0 = \left\{Q \in \mathcal{Q} \;|\; \text{ there exists } \eta \in \mathbb{R}^d \text{ such that } b(Q) - A(Q)\eta \leq 0\right\}.$$

We consider now a setting where some entries of the matrix $A(Q)$ ($\in \mathbb{R}^{p\times d}$), as well as some components of $b(Q)$ are unknown and must be estimated from the data. Our inference procedure is uniformly valid for large families of data-generating processes that satisfy the assumptions outlined below. Throughout, we focus on a semi high-dimensional setting in which the dimension of the unknown components of b and A

[21] As shown in Section 4.3, a hypothesis of the form $H_0 :\ \exists x \geq 0$ s.t. $Ax = b$, with A and b unknown, can be reformulated into a equivalent hypothesis of the current form.

remains bounded, while the dimensions of their deterministic components are allowed to vary arbitrarily, as the sample size increases. The restrictions that we impose on the dimensions of the problem are summarized in Assumption 4.22 below.

To simplify the analysis, we assume that the number of columns of A is fixed, while the number of rows may grow with the sample size. Let $K \cup U = [p]$ denote a partition of $[p]$, such that $i \in K$ iff A_{ij}, for $j \in [d]$, and b_i are known and not subject to estimation. Thus U denotes the set of indices i such that either b_i is unknown or A_{ij} is unknown for some $j \in [d]$. Let $p^u := |U|$ and $p^k = |K|$. We impose the following assumption.[22] We let $\Pi^u(\lambda)(\cdot)$ denote the projection on on the U-coordinates defined by

$$[\Pi^u(\lambda)]_i = \lambda_i \quad \forall i \in U \qquad \text{and} \qquad [\Pi^u(\lambda)]_i = 0 \quad \forall i \in K.$$

Assumption 4.22. We assume that d and p^u are uniformly bounded in N, while p^k may grow arbitrarily with N.

Let $d(Q) = \text{vec}\big([A(Q) \;\; b(Q)]\big)$. Analogously to Assumption 4.2, we assume access to uniformly consistent estimators of the matrix A and vector b, denoted by $\hat{d}$, which satisfy a uniform asymptotic normality property, as formalized in the following assumption.

Assumption 4.23. For each $Q \in \mathcal{Q}$, there exists a symmetric matrix $\Sigma(Q)$, representing the asymptotic variance of $\hat{d}$, such that for all $\epsilon > 0$,

$$\overline{\lim}_{N\to\infty} \sup_{Q\in\mathcal{Q}} d_{pr}\left(\sqrt{N}(\hat{d} - d(Q)), \mathcal{N}(0, \Sigma(Q))\right) = 0.$$

We note that although the dimension of d may vary with N, Assumption 4.23 is not restrictive in light of Assumption 4.22, which implies that $\sqrt{N}(\hat{d} - d)$ is nonzero for only a uniformly (in N) bounded number of components.

Let $D(Q) = \text{diag}(\Sigma(Q))^{1/2}$ denote the asymptotic standard deviation matrix of $\hat{d}$, and let $\hat{D}$ be an estimator of D. Analogously to Assumption 4.12, the following assumption requires that $\hat{D}$ is a uniformly consistent estimator of D. We also assume that the asymptotic standard deviations of the estimated components of d are uniformly bounded away from zero and infinity over all $Q \in \mathcal{Q}$.

Assumption 4.24. i) There exist positive constants $\underline{\sigma}$ and $\bar{\sigma}$, with $\underline{\sigma} < \bar{\sigma}$, such that for all $Q \in \mathcal{Q}$, all diagonal elements of $D(Q)$ that correspond to the estimated component of d lie in the interval $[\underline{\sigma}, \bar{\sigma}]$.

ii) For all $\epsilon > 0$,

$$\overline{\lim}_{N\to\infty} \sup_{Q\in\mathcal{Q}} Q\left(\|\hat{D} - D(Q)\| > \epsilon\right) = 0.$$

The next assumption requires the availability of a uniformly consistent bootstrap-like estimate of the asymptotic distributions $\{\mathcal{N}(0, \Sigma(Q)) \mid Q \in \mathcal{Q}\}$.

[22] It is possible to allow the number of columns of A to grow with the sample size; what matters is that the number of estimated components of A remains uniformly bounded.

Assumption 4.25. There exists a random variable ζ_N^*, whose distribution is entirely determined by the realized sample $\mathbb{W}$, such that for all $\epsilon > 0$,

$$\overline{\lim}_{N\to\infty} \sup_{Q\in\mathcal{Q}} Q\left(d_{\Pr}\left(\mathcal{L}(\zeta_N^* \mid \mathbb{W}), \mathcal{N}(0, \Sigma(Q))\right) > \epsilon\right) = 0.$$

To describe the structural conditions imposed on the matrix A, we introduce the following definition, which serves as an analogue of a condition number for systems of inequalities. The quantity $\Lambda(A)$ considers subsets of rows of A that form a basis for its row space and measures how close such subsets are from being linearly dependent. It thus serves as a measure of multicollinearity, becoming large when some basis of the row space are nearly collinear.

Definition 4.26. Given a symmetric matrix B, let $\lambda_{\min}(B)$ denote its smallest eigenvalue. For an arbitrary matrix $A \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) = r$, define the (strictly positive) quantity $\Lambda(A)$ by

$$\Lambda(A)^{-1} := \min\left\{\lambda_{\min}(A_F A_F^\top) \mid F \subseteq [m],\ |F| = r,\ \operatorname{rank}(A_F) = r\right\}, \tag{4.24}$$

where A_F denotes the submatrix formed by the rows indexed by F.

As demonstrated in Lemma 7.8 in the appendix, $\Lambda(A)$ provides an upper bound on the Hoffman constant, which characterizes the local behavior of polyhedrons $\{x \mid Ax \le b\}$ when A and b are subject to perturbations. While better bounds for the Hoffman constant exist in the literature—such as the Robinson constant (see Robinson (1973))—we employ $\Lambda(\cdot)$ due to its interpretability.

We consider the following assumption, which mainly restricts the coefficient matrices that are considered in our uniformity result.

Assumption 4.27. The coefficient matrices $A(Q)$ and vectors $b(Q)$ satisfy

i) $\sup_{Q\in\mathcal{Q}_0} \Lambda(A(Q)) < \infty$

ii) For $Q \in \mathcal{Q}_0$, let $\eta_Q := \arg\min\{\|\eta\| \mid b(Q) - A(Q)\eta \le 0\}$. We assume that $\sup_{Q\in\mathcal{Q}_0} \|\eta_Q\| < \infty$.

Part (i) of Assumption 4.27 is mild, as it imposes only an upper bound on $\Lambda(A)$, a quantity that is finite for any fixed matrix A. However, the function $\Lambda(\cdot)$ is not continuous, and the condition $\sup_{A\in\mathcal{A}} \Lambda(A) < \infty$ excludes any set $\mathcal{A}$ that *contains a neighborhood* of a matrix with linearly dependent rows.[23] Part (ii) of the assumption implies the existence of a constant $M > 0$ such that all feasible regions $\{\eta \mid b(Q) - A(Q)\eta \le 0\}$, for $Q \in \mathcal{Q}_0$, intersect the Euclidean ball $B(0, M)$. This condition is used in the proof to control the sensitivity of the statistic $\hat{T}_N$ with respect to the estimation of the matrix A.

We now describe our test statistic. Given the asymptotic standard deviation matrix D, let $\sigma_{A_{ij}}$ and σ_{b_i} denote the asymptotic standard deviations of the estimators for the (i, j)-th entry of the matrix A and the

[23] If a matrix $A \in \mathbb{R}^{m\times n}$ has a subset $F \subseteq [m]$ of rows, with $|F| \le n$, that are linearly dependent, then arbitrarily small perturbations $\tilde{A}$ of A, can result in $\tilde{A}_F$ with linearly independent rows that are nearly collinear. This causes $\Lambda(\tilde{A}_F)$—and hence $\Lambda(\tilde{A})$—to become arbitrarily large.

i-th entry of the vector b, respectively. These correspond to diagonal entries of D. Let $\hat{\sigma}_{A_{ij}}$ and $\hat{\sigma}_{b_i}$ be the analogous entries from the estimator $\hat{D}$. Define the diagonal matrices Ω and $\hat{\Omega}$ (both in $\mathbb{R}^{p\times p}$) by:

$$\Omega_{ii} = \sigma_{b_i} \vee \max_{j\in[d]} \sigma_{A_{ij}}, \qquad \hat{\Omega}_{ii} = \hat{\sigma}_{b_i} \vee \max_{j\in[d]} \hat{\sigma}_{A_{ij}}.$$

Note that the diagonal entry Ω_{ii} is set to zero for $i \in K$, and is strictly positive for $i \in U$. These matrices serve the same normalizing role as $\hat{D}$ in Sections 4.1–4.3, and are used below to ensure that the test statistic is scale invariant. Following the linear programming characterization developed in Example 2.4, testing the hypothesis H_0 (vs H_1) in 4.18, is equivalent to testing whether $\nu = 0$ (vs $\nu > 0$), where ν represent the value of the linear program

$$\nu = \max\left\{b^\top\lambda \mid \lambda \geq 0,\ A^\top\lambda = 0,\ \|\Omega\lambda\|_1 \leq 1\right\}, \tag{4.25}$$

Our test statistic is given by the value of the sample analogue of the preceding LP:

$$\hat{T}_N := \max\left\{\hat{b}^\top\lambda \mid \lambda \geq 0,\ \hat{A}^\top\lambda = 0,\ \|\hat{\Omega}\lambda\|_1 \leq 1\right\}. \tag{4.26}$$

We reject the null hypothesis in favor of the alternative whenever $\sqrt{N}\hat{T}_N$ is large. Specifically, given a significance level $\alpha \in (0, 1/2)$, we reject the null hypothesis in favor of the alternative if and only if

$$\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha),$$

where $\hat{c}_N(1-\alpha)$ is defined by

$$\hat{c}_N(1-\alpha) := \inf\{u \mid \hat{H}_N(u) \geq 1-\alpha\}, \tag{4.27}$$

and $\hat{H}_N(\cdot) := \mathcal{L}(\nu^*|\mathbb{W})$ is the conditional (on $\mathbb{W}$) distribution of ν^*, where ν^* is the value of the following LP:

$$\nu^* = \max_{\lambda\in\hat{\Delta}} \langle \zeta_b^* - \zeta_A^*\hat{\eta}, \lambda\rangle \tag{4.28}$$

where

$$\hat{\Delta} := \left\{\lambda \in \mathbb{R}_+^p \;\middle|\; -\frac{\kappa_{1N}}{\sqrt{N}}\mathbb{1} \leq \hat{A}^\top\lambda \leq \frac{\kappa_{1N}}{\sqrt{N}}\mathbb{1},\ \|\hat{\Omega}\lambda\|_1 \leq 1,\ \hat{b}^\top\lambda \geq \hat{T}_N - \kappa_{2N}/\sqrt{N}\right\}, \tag{4.29}$$

and $\hat{\eta}$ is given by the (unique) solution of the quadratic program[24]

$$\hat{\eta} = \arg\min\left\{\|\eta\|^2 \mid \hat{b} - \hat{A}\eta \leq (\hat{T}_N + \frac{\kappa_{3N}}{\sqrt{N}})\hat{\Omega}\mathbb{1}\right\}. \tag{4.30}$$

Here, ζ_b^* and ζ_A^* are the b and A components of ζ_N^*, the bootstrap-like estimator of the asymptotic distribution of $\hat{b}$ given in Assumption 4.25. The tuning parameters κ_{1N}, κ_{1N}, and κ_{3N} are all required to satisfy $\kappa_N = o(\sqrt{N})$ and $\kappa_N \to \infty$. In the proof of Theorem 4.29, we show that $\hat{\eta}$ is a uniformly (in $Q \in \mathcal{Q}_0$) consistent estimator of η_Q, defined in Assumption 4.27. The inclusion of $\hat{T}_N$ in equation 4.30 guarantees that the feasible region of the quadratic program is always non-empty.[25]Moreover, the term $\frac{\kappa_{3N}}{\sqrt{N}}$, which

[24]We could replace the quadratic program with a linear program by minimizing the ℓ_1 norm instead. However, in that case, Assumption 4.28 would need to hold for all covariance matrices $V(Q)$, where η_Q is now any element of the set $\arg\min\{\|\eta\|_1 \mid b - A\eta \leq 0\}$, which may now contain multiple elements.

[25]This can be seen by considering the dual of the LP 4.26.

asymptotically dominates $\hat{T}_N = O(1/\sqrt{N})$, ensures that η_Q lies within the feasible region of the quadratic program with probability approaching one, uniformly over $Q \in \mathcal{Q}_0$, as $N \to \infty$.

As in the preceding sections, we note that the feasible regions of the LP (4.26) are not necessarily bounded when $K \neq \emptyset$ (i.e., there exists $i \in [p]$ such that b_i and the i-th row of A are deterministic). In such cases, it is possible for $\hat{T}_N$ to be infinite. However, as previously argued, when $\hat{\Omega}$ and Ω have the same nonzero diagonal elements, Farkas' Lemma implies that $\hat{T}_N$ cannot be infinite under the null.

The following assumptions is needed to guarantee on the event where our test rejects, the distributions of our test statistic (for $Q \in \mathcal{Q}_0$) do not asymptotically concentrate. We can do without this assumption, if we are willing to modify the rejection rule to: reject iff $\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha) + \tau$, where $\tau > 0$ is a fixed small constant.

Let the random elements (ζ_A, ζ_b), with $\zeta_A \in \mathbb{R}^{p\times d}$ and $\zeta_b \in \mathbb{R}^p$, be such that $\mathcal{L}\Big(vec([\zeta_A \quad \zeta_b])\big|Q\Big) \sim \mathcal{N}(0, \Sigma(Q))$. For $Q \in \mathcal{Q}_0$, let $V(Q) \in \mathbb{R}^{p\times p}$ be defined by

$$\zeta_b - \zeta_A \eta_Q \sim \mathcal{N}(0, V(Q))$$

where η_Q is as in Assumption 4.27.

Assumption 4.28. There exists a constant $\rho > 0$ such that for all $Q \in \mathcal{Q}_0$:

i) If $\Delta_0(Q) \neq \{0\}$, then there exists $\lambda \in \Delta_0(Q)$ such that $\lambda^\top V(Q)\lambda \geq \rho$.

ii) If $\Delta_0(Q) = \{0\}$ and $\mathrm{extr}(\mathcal{D}(Q)) \setminus \{0\} \neq \emptyset$, then for all $\lambda \in \mathrm{extr}(\mathcal{D}(Q)) \setminus \{0\}$, we have $\lambda^\top V(Q)\lambda \geq \rho$.

We now state the main result of this section. Its proof is provided in Section 7.0.1 of the appendix.

Theorem 4.29. *Suppose that Assumptions 4.22 through 4.28 hold, and let $\hat{T}_N$ be as in equation (4.26). Then, for all $\alpha \in (0, 1/2)$,*

$$\overline{\lim}_{N\to\infty} \sup_{Q\in\mathcal{Q}_0} Q\left(\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\right) \leq \alpha, \tag{4.31}$$

where $\hat{c}_N(1-\alpha)$ is defined as in equation (4.27).

Moreover, for any $Q \in \mathcal{Q}_0$ such that the system $b(Q) - A(Q)\eta \leq 0$ has η_Q as its unique solution, we have

$$\lim_{N\to\infty} Q\left(\left\{\mathbb{W}_N \,\middle|\, d_{\Pr}(\mathcal{L}(\sqrt{N}\hat{T}_N), \hat{H}_N) > \epsilon\right\}\right) = 0, \tag{4.32}$$

for any $\epsilon > 0$. Here $\hat{H}_N(\cdot)$ denotes the conditional distribution (given $\mathbb{W}_N$) of v^* as defined in equation (4.28).

Remark 4.30. Equation 4.32 implies that the inequality in Equation 4.31 can be replaced by an equality if, for some $Q \in \mathcal{Q}_0$, the inequality system $b(Q) - A(Q)\eta \leq 0$ has η_Q as its unique solution, and the asymptotic distribution of $\sqrt{N}\hat{T}_N$ is nondegenerate. In Lemma 7.4, we derive the asymptotic distribution of $\sqrt{N}\hat{T}_N$ under such DGPs and show that it is nondegenerate if and only if the optimal solution set $\Delta_0(Q)$ of the linear program in Equation 4.25 strictly contains the set $\{0\}$. As a consequence, our test is not unduly conservative when the family $\mathcal{Q}_0$ includes such DGPs. For DGPs $Q \in \mathcal{Q}_0$ where the inequality system $b(Q) - A(Q)\eta \leq 0$

has multiple solutions and $\Delta_0(Q)$ strictly contains $\{0\}$, it can be shown that the pointwise (i.e., for fixed Q) asymptotic limit of $\hat{H}_N(\cdot)$ stochastically dominates the pointwise asymptotic distribution of $\sqrt{N}\hat{T}_N$ in the first-order stochastic dominance sense, with both distributions being nondegenerate. In such cases, our testing procedure is expected to be conservative, with the pointwise asymptotic size of the test potentially much smaller than the nominal level α.

5 Simulation exercise: inference on the average treatment effect

In this section we study the finite-sample performance of our inference methods, focusing on the procedure of Section 4.4 for testing linear systems with unknown coefficients. To do so, we apply the method to a setting in which the parameter of interest is the average treatment effect of a binary treatment, which is partially identified using instrumental variables (IV). The problem belongs to a class studied by Mogstad, Santos, and Torgovitsky (2018) (MST) for identification in IV models with heterogeneous treatment effects.

We build our simulation around an empirical application studied in the working paper version of MST (Mogstad, Santos, and Torgovitsky (2017)). This application considers a field experiment from Dupas (2014) that offered randomized prices Z for an antimalarial bed net in Kenya. The treatment $D \in \{0, 1\}$ indicates whether a given household purchased the bed net at the price they were offered, and the outcome $Y \in \{0, 1\}$ whether the household reports actually using the bed net two months after the experiment. Given potential outcomes $Y(0), Y(1)$, we consider the average treatment effect $\text{ATE} = \mathbb{E}[Y(1) - Y(0)]$ as the parameter of interest, which is partially identified in this setting given standard IV assumptions.[26] A simplification afforded by the setting of Dupas (2014) is that $Y(0) = 0$ with probability one, because this particular bed net was not available at the time outside of the field experiment. Hence the ATE reduces to $\text{ATE} = \mathbb{E}[Y(1)]$, the utilization rate that would occur if everyone purchased the bed net.

A strength of the MST approach to identification with instrumental variables is that additional assumptions about how selection is related to potential outcomes can be used to reduce the identified set for partially identified parameters like the ATE. Following MST, assume that purchases of the bed net are governed by $D = D(Z)$, where

$$D(z) = \mathbb{1}(U \leq p(z)) \tag{5.1}$$

where $U \sim \text{Unif}[0, 1]$, and $p(z) = \mathbb{E}[D \mid Z = z]$ is the propensity score function. Instrument validity implies that $Z \perp (U, Y_i(1))$. Vytlacil (2002) establishes an equivalence between this model and the LATE model of Imbens and Angrist (1994), which, instead of Eq. (5.1), imposes the absence of defiers.

Given the variable U that governs selection into treatment (purchase of the net), MST propose leveraging assumptions on marginal treatment response (MTR) curves m_0 and m_1, where $m_d(u) := \mathbb{E}[Y_i(d)|U = u]$ for $d \in \{0, 1\}$ and $u \in [0, 1]$. In our setting $m_0(u) = \mathbb{E}[Y(0)|U = u] = 0$ for all u so we focus on the treated MTR curve $m_1(u)$. A natural assumption is that $m_1(u)$ is weakly decreasing in u, indicating that individuals who would buy the bed net at a higher price are also more likely to use the net if they purchase it.

[26] The ATE is not point identified under the standard LATE model assumptions of Imbens and Angrist (1994) due to the presence of never-takers (who do not purchase the bed net even at the lowest price offered) and always-takers (who purchase the bed net even at the highest price offered). Since the outcome $Y \in \{0, 1\}$ is bounded, the ATE remains partially identified.

It is known that, even with a binary instrument $Z \in \{0,1\}$, the ATE is point identified if one imposes the assumption that the MTR curves $m_d(u)$ are linear in u (Brinch, Mogstad, and Wiswall (2017)). To allow for partial identification, we instead impose that $m_1(u)$ is quadratic: $m_1(u) = \alpha + \beta \cdot u + \gamma \cdot u^2$ for some parameters (α, β, γ). Given this assumption, the average treatment effect can be written as

$$ATE = \mathbb{E}[Y_i(1)] = \int_0^1 \mathbb{E}[Y(1)|U=u] \cdot du = \int \{\alpha + \beta \cdot u + \gamma \cdot u^2\} \cdot du = \alpha + \frac{\beta}{2} + \frac{\gamma}{3} \tag{5.2}$$

For simplicity, we focus on a case with a binary instrument $Z \in \{0,1\}$ that reflects a lower versus higher price for the bed net, e.g. $Z = 1$ corresponds to a price of 50 Kenyan Shillings and $Z = 0$ to 150 Kenyan shillings. These prices are a subset of those randomized in the experiment of Dupas (2014), and correspond to the prices at which $p(0) \approx 1/3$ and $p(1) \approx 2/3$ of households buy the bed net. For either value of $z \in \{0,1\}$:

$$\mathbb{E}[Y \cdot D|Z=z] = \mathbb{E}[Y_i(1)\mathbb{1}(U \le p(z))] = \int_0^{p(z)} \mathbb{E}[Y(1)|U=u] \cdot du = p(z) \cdot \alpha + \frac{p(z)^2}{2} \cdot \beta + \frac{p(z)^3}{3} \cdot \gamma \tag{5.3}$$

Eq. (5.3) constitutes a system of equations that are linear in the parameters $\lambda = (\alpha, \beta, \gamma)$, which provide information about the parameter of interest ATE which is also linear in those parameters by Eq. (5.2).

The researcher uses the estimands $\mathbb{E}[Y \cdot D|Z=0]$ and $\mathbb{E}[Y \cdot D|Z=1]$ to provide identifying information from the outcome data. They assume that $\mathbb{E}[Y(1)|U=u] = \alpha + \beta \cdot u + \gamma \cdot u^2$, that $\mathbb{E}[Y(1)|U=u]$ is weakly decreasing in u, that $\mathbb{E}[Y(1)|U=u] \in [0,1]$ for all $u \in [0,1]$ (as implied by Y being binary), and that $\mathbb{E}[Y(0)|U=u] = 0$. Monotonicity implies that $\frac{d}{du}\mathbb{E}[Y(1)|U=u] = \beta + 2\gamma u$ is nonpositive on the unit interval, i.e. $\beta \le 0$ and $\beta + 2\gamma \le 0$. Boundedness implies that $0 \le \alpha \le 1$ and $0 \le \alpha + \beta + \gamma \le 1$.[27]

If $ATE = \theta$, we then obtain the system of linear equalities $A(Q)\lambda \ge b(Q,\theta)$, where:

$$A(Q) = \left[\begin{array}{ccc} p(0) & p(0)^2/2 & p(0)^3/3 \\ -p(0) & -p(0)^2/2 & -p(0)^3/3 \\ p(1) & p(1)^2/2 & p(1)^3/3 \\ -p(1) & -p(1)^2/2 & -p(1)^3/3 \\ \hline 1 & 1/2 & 1/3 \\ -1 & -1/2 & -1/3 \\ \hline -1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & -1 & -2 \end{array}\right], \quad b(Q,\theta) = \left(\begin{array}{c} \mathbb{E}[Y \cdot D|Z=0] \\ -\mathbb{E}[Y \cdot D|Z=0] \\ \mathbb{E}[Y \cdot D|Z=1] \\ -\mathbb{E}[Y \cdot D|Z=1] \\ \hline \theta \\ -\theta \\ \hline -1 \\ 0 \\ 0 \\ 0 \end{array}\right), \quad \begin{array}{c} \text{(data)} \\ \text{(data)} \\ \text{(data)} \\ \text{(data)} \\ \text{(hypothesis } ATE \ge \theta) \\ \text{(hypothesis } ATE \le \theta) \\ m_1(0) \le 1 \\ m_1(1) \ge 0 \\ m_1(u) \text{ decreasing at } u = 0 \\ m_1(u) \text{ decreasing at } u = 1 \end{array}$$

where note that the first four rows of both A and b involve estimated quantities. We test the null that $A(Q)\lambda \ge b(Q,\theta)$ for a grid of θ values ranging from 0 to 1, using the procedure described in Section 4.4. The data is drawn from a DGP in which $P(Z=1) = 1/2$, D is generated via Eq. (5.1) with $p(0) = 1/3$, $p(z) = 2/3$, and $U|Z \sim Unif[0,1]$, and $Y(1) = \mathbb{1}(W \le \alpha + \beta U + \gamma U^2)$, where $W|Z,U \sim Unif[0,1]$.

[27]Note that $m_1(0) \ge 0$ is implied by $m_1(1) \ge 0$ and m_1 being non-increasing, and similarly $m_1(1) \le 1$ is implied by $m_1(0) \le 1$ and m_1 being non-increasing. Thus in the we have dropped two redundant inequality restrictions in A and b.

Figure 1 plots rejection probabilities for 1000 draws for a test with nominal size of 5%, across two versions of this DGP. We begin by setting $n = 250$.[28] In the left panel a), we set $\alpha = 1, \beta = -1, \gamma = 1/2$. The true value of the ATE in this 0.67 (depicted by a vertical solid red line), and the identified set is $[0.58, 0.67]$ (depicted by the hatched region). The horizontal dotted line depicts 5%. In this version of the DGP, the slope of $m_1(u)$ is zero at $u = 1$, and the true value is on the boundary of the identified set. The right panel b) of Figure 1 considers a second DGP in which $\alpha = 0.9$, $\beta = -1$, and $\gamma = 1/3$. In this setting the true value of the ATE is 0.51, which lies in the interior of the identified set $[0.47, 0.53]$. In both panels, two choices of the tuning parameter κ_n are compared, $\kappa_n = \sqrt{\log n}$ and $\kappa_n = \log n$. The results depicted in Figure 1 appear not to be sensitive to this choice, given the DGP and sample size.

In both DGPs depicted in Figure 1, the test controls size well throughout the identified set for the ATE, and gains power relatively quickly outside of it. We note that the power gain is asymmetric. For example, in panel a), the rejection probability remains below 5% until $\theta \approx 0.75$, well outside of the identified set. This is a small sample phenomenon only however: in panel a) we also show the power curve with a sample size of $n = 5000$, which traces out the identified set exactly (up to our grid spacing for θ: multiples of 0.05).

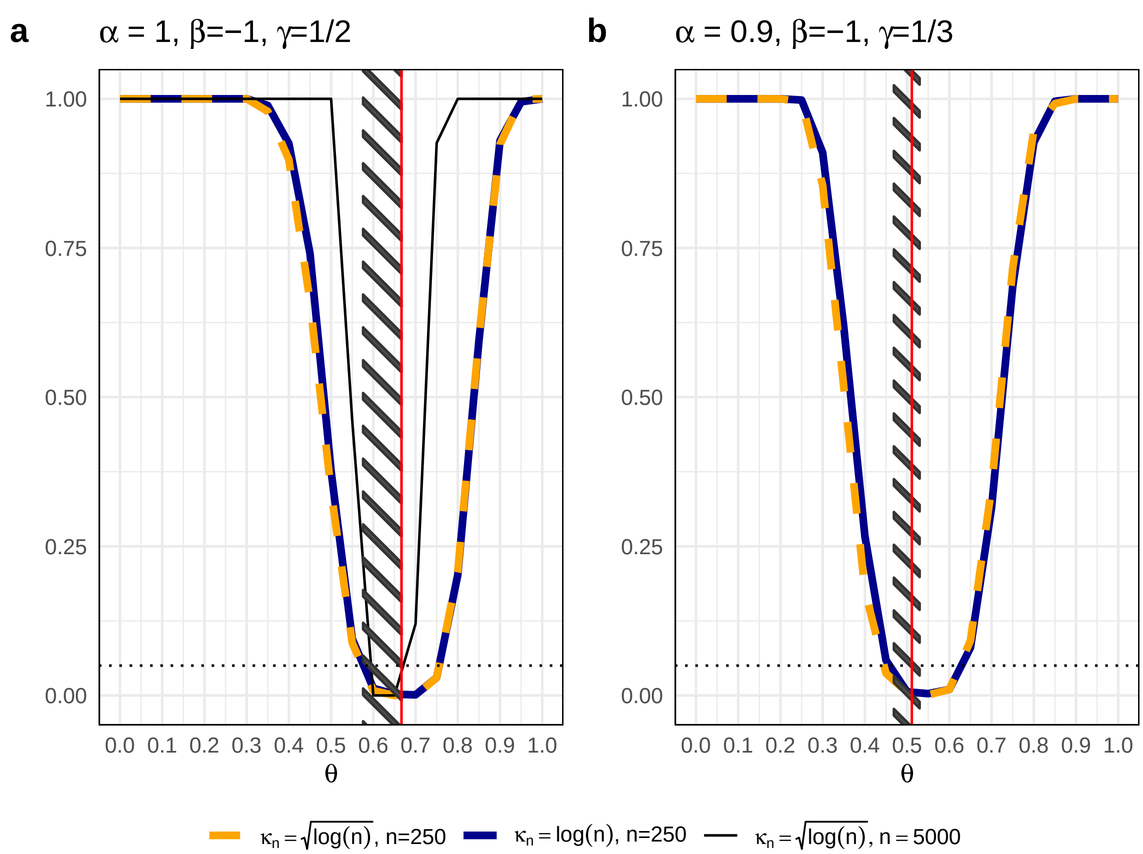

Figure 1: Rejection probabilities for testing hypothetical values θ of the ATE.

[28] This corresponds approximately to the sample size in Dupas (2014) for our two instrument values: $Z = 1$(a price of 50 Kenyan Shillings), and $Z = 0$ (a price of 150 Kenyan shillings).

6 Conclusion

This paper proposes a new approach to inference on the value ν of a linear program when the parameters of the program are estimated from data. Building on a bootstrap approximation to the distribution of $\sqrt{n}(\hat{\nu} - \nu)$ that is pointwise valid, we develop modified bootstrap procedures that provide uniformly valid inference in settings of central interest to econometrics. Our procedure demonstrates good performance in a simulation study that involves testing linear systems with unknown coefficient matrices. In future work, we aim to extend our results to high-dimensional settings, propose data-driven methods for selecting the tuning parameter κ_n, and conduct detailed comparisons between the performance of our procedure and that of existing methods in the settings that we consider.

7 Appendix

7.0.1 Proof of Theorem 3.4

Before proceeding to the proof of Theorem 3.4, we first establish the following lemma, which shows that the sets $\hat{S}$ and $\hat{\Delta}$ are Hausdorff consistent to the solution sets of the primal and dual problems, respectively.

Lemma 7.1. *Suppose that 3.1, 3.2 and 3.3 hold. Then* $d_H(\hat{S}, S_0) = o_p(1)$ *and* $d_H(\hat{\Delta}, \Delta_0) = o_p(1)$.

Proof of Lemma 7.1. We proceed in two steps. First, we establish the upper hemicontinuity of the solution sets of the primal and dual problems using Theorem 1 of Robinson (1977). Second, we establish the lower hemicontinuity of these solution sets using Theorem 1 of Robinson (1975). Both references study the stability of the feasible regions and solution sets of linear programs under perturbations.

Step 1 (Upper hemicontinuity)

Note that by Lemma 7.11, the MFCQ ensures that the set of dual optimal solutions, Δ_0, is bounded. Indeed, there is a fundamental duality between the regularity of the primal problem and the boundedness of the solution sets of the dual problem(see also parts (a) and (b) of Theorem 1 in Robinson (1977)). Since the feasible set of the primal problem, $\mathcal{P}$, is compact, the set of optimal solutions for the primal, S_0, is also bounded. Consequently, by part (c) of Theorem 1 in Robinson (1977), there exist positive constants c_1 and c_2 that depend only on the input d $(= \text{vec}(A, b, c))$, such that

$$\|\hat{d} - d\| \leq c_1 \tag{7.1}$$

implies that the primal and dual problems corresponding to the inputs $\hat{d}$ are feasible, and satisfy

$$\sup_{(\theta,\lambda)\in\hat{S}_0\times\hat{\Delta}_0} d\left((\theta,\lambda), S_0 \times \Delta_0\right) < c_2\|\hat{d} - d\|, \tag{7.2}$$

where the sets $\hat{S}_0$ and $\hat{\Delta}_0$, which should not be confused with $\hat{S}$ and $\hat{\Delta}$, represent the sets of optimal solutions for the primal and dual problems, respectively, corresponding to the input $\hat{d}$. Inequality 7.2 establishes that when the estimation error is sufficiently small, the solution sets to the perturbed primal and dual problems remain close to those of the unperturbed primal and dual problems, with the distance being bounded by a fixed constant times the approximation error. Let $\hat{\theta}_0$ and θ_0 are the solutions to the perturbed and unperturbed primal problems, respectively, then the values $\hat{\nu}$ and ν of the perturbed and unperturbed primal programs satisfy $\hat{\nu} = \hat{c}^T\hat{\theta}_0$ and $\nu = c^T\theta_0$. Thus, whenever inequality 7.1 holds, inequality 7.2 and the first part of Assumption 3.3 imply that

$$\hat{\nu} - \nu = O_p(1/\sqrt{N}). \tag{7.3}$$

Note that since $\hat{S}$ and $\hat{\Delta}$ are, respectively, enlargements of $\hat{S}_0$ and $\hat{\Delta}_0$, inequalities 7.1 and 7.2 imply that $\hat{S}$ and $\hat{\Delta}$ are nonempty whenever inequality 7.1 holds. Lemma 7.9 (see also Lemma 2 of Robinson (1977)) applied to the systems

$$S_0 = \{\theta \in \mathbb{R}^d \mid A_E\theta = b_E, A_I\theta \leq b_I, c^T\theta \leq \nu\}$$

and

$$\Delta_0 = \{\lambda = (\lambda_E, \lambda_I) \in \mathbb{R}^p \mid A_E^T\lambda_E + A_I^T\lambda_I = c,\ \lambda_I \leq 0,\ \lambda_E^T b_E + \lambda_I^T b_I \geq \nu\},$$

implies that there exists a positive constant c_3 and c_4 that depend only on the input d such that whenever

$$\hat{\delta} := \max\{\|\hat{d} - d\|, |\nu - \hat{\nu} - \kappa_N/\sqrt{N}|, |\nu - \hat{\nu} + \kappa_N/\sqrt{N}|\} \leq c_3 \tag{7.4}$$

we get

$$\sup_{\theta\in\hat{S}} d(\theta, S_0) \leq c_4\hat{\delta} \quad \sup_{\lambda\in\hat{\Delta}} d(\lambda, \Delta_0) \leq c_4\hat{\delta}. \tag{7.5}$$

Assumption 3.3 and inequality 7.3 then imply that $\hat{\delta} = O_p(\kappa_N/\sqrt{N})$, from which we conclude that

$$\sup_{\theta\in\hat{S}} d(\theta, S_0) = o_p(1) \quad \text{and} \quad \sup_{\lambda\in\hat{\Delta}} d(\lambda, \Delta_0) = o_p(1). \tag{7.6}$$

Step 2 (Lower hemicontinuity)

By Lemma 7.11, the compactness of the primal feasible region implies that the dual problem satisfies the MFCQ. By Lemma 7.10 (see also Theorem 1 in Robinson (1975)), since both the primal and dual problems satisfy the MFCQ, For any $R > 0$, there exist constants c_1 and c_2 such that

$$\|\hat{d} - d\| \leq c_1 \tag{7.7}$$

ensures that the feasible regions of the perturbed primal $(\hat{\mathcal{P}})$ and perturbed dual $(\hat{\mathcal{D}})$ problems, corresponding to the input $\hat{d}$, are nonempty and satisfy[29]

$$\max\Big\{\sup_{\theta\in\mathcal{P}} d(\theta, \hat{\mathcal{P}}), \sup_{\lambda\in\mathcal{D}, \|\lambda\|\leq R} d(\lambda, \hat{\mathcal{D}})\Big\} \leq c_2\|\hat{d} - d\|. \tag{7.8}$$

Given a closed and convex subset $A \in \mathbb{R}^m$ for some $m \geq 1$, let $\Pi_A(x)$ denote the closest element to x in A.

[29] Note that the feasible region of the dual is not necessarily bounded.

For any $\theta \in S_0$, we have

$$|\hat{\nu} - \hat{c}^T \Pi_{\hat{\mathcal{P}}}(\theta)| \leq |\hat{\nu} - \nu| + |\hat{c}^T(\theta - \Pi_{\hat{\mathcal{P}}}(\theta))| + |(\hat{c} - c)^T \theta|.$$

By Assumptions 3.2 and 3.3, along with inequalities 7.3 and 7.8, it follows that

$$\sup_{\theta \in S_0} |\hat{\nu} - \hat{c}^T \Pi_{\hat{\mathcal{P}}}(\theta)| = O_p(1/\sqrt{N}).$$

Thus, we conclude that the following event occurs with probability approaching 1:

$$\Pi_{\hat{\mathcal{P}}}(S_0) \subset \hat{S} = \left\{\theta \in \hat{\mathcal{P}} \mid \hat{c}^T \theta \leq \hat{\nu} + \kappa_N/\sqrt{N}\right\}. \tag{7.9}$$

The latter, together with inequality 7.8, implies that

$$\sup_{\theta \in S_0} d(\theta, \hat{S}) = O_p(1/\sqrt{N}).$$

To derive the analogous bound for Δ_0, first note that whenever inequality 7.1 holds, the boundedness of Δ_0 and inequality 7.2 imply uniform (over all realizations of the sample satisfying 7.1) boundedness of Δ_0 and $\hat{\Delta}_0$. Moreover, the constant R in inequality 7.8 can be chosen sufficiently large to serve as such a uniform bound. Then arguing as above, we obtain

$$\sup_{\lambda \in \Delta_0} d(\lambda, \hat{\Delta}) = O_p(1/\sqrt{N}).$$

Consequently, we obtain

$$\sup_{\theta \in S_0} d(\theta, \hat{S}) = o_p(1) \quad \text{and} \quad \sup_{\lambda \in \Delta_0} d(\lambda, \hat{\Delta}) = o_p(1). \tag{7.10}$$

This completes the proof. $\square$

Proof of Theorem 3.4. The second part of Assumption 3.3 and Lemma 7.1 imply that there exists a positive sequence of constants ϵ_N converging to zero sufficiently slowly such that the event (here d_w denotes the dimension of the sample space for one observation)

$$E_N := \left\{\mathbb{W} \in \mathbb{R}^{d_w \times N} \;\middle|\; \max\left\{d_H(\hat{S}, S_0), d_H(\hat{\Delta}, \Delta_0), d_{Pr}\left(\mathcal{L}\left(\zeta_N^* \mid \mathbb{W}\right), \mathbb{G}\right)\right\} \leq \epsilon_N\right\}$$

has probability tending to one asymptotically. Now we show that for any sequence $\{\mathbb{W}_N\}_{N \geq 1}$ such that $\mathbb{W}_N \in E_N$, we have

$$\mathcal{L}\left(\min_{\theta \in \hat{S}} \max_{\lambda \in \hat{\Delta}} \langle \zeta_c^*, \theta \rangle + \lambda^T \left(\zeta_b^* - \zeta_A^* \theta\right) \mid \mathbb{W}_N\right) \Rightarrow \min_{\theta \in S_0} \max_{\lambda \in \Delta_0} \langle \zeta_c, \theta \rangle + \lambda^T \left(\zeta_b - \zeta_A \theta\right) \tag{7.11}$$

with $\zeta = (\zeta_A, \zeta_b, \zeta_c) \overset{d}{\sim} \mathbb{G}$, from which the claim of the theorem follows. Indeed, given such a sequence of data realizations $\{\mathbb{W}_N\}$, it follows from the definition of E_N that

$$\mathcal{L}\left(\zeta_N^* \mid \mathbb{W}_N\right) \Rightarrow \mathbb{G}.$$

Also, as $\hat{S}$ and $\hat{\Delta}$ are completely determined from the data, along the data realizations $\{\mathbb{W}_N\}$, the sets $\{\hat{S}\}$, and $\{\hat{\Delta}\}$ are nonrandom and converge in the Haussdorf sense to S_0 and Δ_0, respectively. By the Skorokhod representation theorem, there exists a sufficiently rich probability space where the latter convergence holds almost surely. Thus, given the realizations $\{\mathbb{W}_N\}$, we can assume that $\zeta_N^* \xrightarrow{a.s.} \zeta$ with $\zeta \overset{d}{\sim} \mathbb{G}$. Given a value of ζ, let $F(\theta, \lambda; \zeta)$ be defined by

$$F(\theta, \lambda; \zeta) = \langle \zeta_c, \theta \rangle + \lambda^T \left(\zeta_b - \zeta_A \theta\right).$$

Consider an outcome in our rich probability space where $\zeta_N^* \to \zeta$ (by assumption, the latter occurs with probability one) and set

$$F_N(\theta, \lambda) := F(\theta, \lambda; \zeta_N^*)$$

and

$$F(\theta, \lambda) := F(\theta, \lambda; \zeta).$$

Note that by the definition of E_N, the sets $\hat{S}$ and $\hat{\Delta}$ that correspond to realizations of the data $\mathbb{W}_N$ in E_N are uniformly bounded over N, and we can assume that there exist compact sets S and Δ such that for realizations of $\mathbb{W}_N \in E_N$, we have $\hat{S} \times \hat{\Delta} \subset S \times \Delta$. It is then not difficult to see (using the definitions of F_N and F) that F_N converges uniformly to F on $S \times \Delta$. As $\max\{d_H(\hat{S}, S_0), d_H(\hat{\Delta}, \Delta_0)\} = o(1)$ along the sequence $\{\mathbb{W}_N\}$, an application of Lemma 7.7 then yields that

$$\min_{\theta \in \hat{S}} \max_{\lambda \in \hat{\Delta}} F_N(\theta, \lambda) \to \min_{\theta \in S_0} \max_{\lambda \in \Delta_0} F(\theta, \lambda).$$

As the latter holds almost surely on our rich probability space, we conclude that

$$\min_{\theta \in \hat{S}} \max_{\lambda \in \hat{\Delta}} F(\theta, \lambda; \zeta_N^*) \xrightarrow{a.s.} \min_{\theta \in S_0} \max_{\lambda \in \Delta_0} F(\theta, \lambda; \zeta),$$

which in turn implies 7.11.

□

Proof of Theorem 4.7. The starting point of the proof is equation 4.8 in Section 4.1, which we reproduce here:

$$\sqrt{N} T_N(\xi) \leq \left(\sqrt{N}\left(b_N(\xi) - b\right)\right)^\top \lambda_\xi. \tag{7.12}$$

For all $\xi = (\xi_b, \xi_D)$ such that $\phi(\mu_N(\xi))$ is finite. The reader is refered to Section 4.1 for the definition of the various terms involved in equation 7.12, and for some of the notation that is used below. Note that Assumption 4.2 and 4.3-i) implies that the family $\{\mathcal{N}(0, \Sigma(Q)) \mid Q \in \mathcal{Q}\}$ is tight, and Assumption 4.2 then implies that the family $\{\sqrt{N}(\hat{b} - b(Q)) \mid Q \in \mathcal{Q}\}$ is asymptotically tight. Thus, Assumptions 4.2 through 4.5

imply that there exists $\alpha_N \downarrow 0$, such that for any $M_N \uparrow \infty$, we have

$$\sup_{Q \in \mathcal{Q}} d_{Pr}\left(\sqrt{N}(\hat{b} - b), \mathcal{N}(0, \Sigma(Q))\right) \le \alpha_N$$

and for all $Q \in \mathcal{Q}$, the event $E_N(Q)$ defined by

$$E_N(Q) := \left\{ \mathbb{W} \mid \|\hat{D} - D\| \le \alpha_N,\ d_{Pr}\big(\mathcal{L}(\zeta^* \mid \mathbb{W}), \mathcal{N}(0, \Sigma(Q))\big) \le \alpha_N, \sqrt{N}\|\hat{b} - b(Q)\| \le M_N \right\}$$

satisfies

$$\underline{\lim}_{N \to \infty} \inf_{Q \in \mathcal{Q}} Q(E_N(Q)) = 1.$$

Choose $M_N \uparrow \infty$ such that $M_N = o(\kappa_N)$ and $M_N \alpha_N = o(1)$. Consider the set of pertubations $\mathcal{X}_N$ defined by

$$\mathcal{X}_N := \left\{ \xi = (\xi_b, \xi_D) \,\middle|\, \|\xi_b\| \le M_N,\ \|\xi_D\| \le \sqrt{N}\alpha_N \right\},$$

where ξ_D are diagonal matrices of the same dimension as D. Below, we use C to denote a generic constant that does not depend on $Q \in \mathcal{Q}$ or N for all sufficiently large N. We allow the value of the constant C to change from one line to the next. In the first three steps below, we establish inequality 4.6. The fourth step gives the proof of inequality 4.7.

Step 1 In this step, we show there exists $\beta_N = o(1)$ such that for all sufficiently large N's and for all $Q \in \mathcal{Q}_0$, we have on the event $E_N(Q)$:

$$\sup_{\xi \in \mathcal{X}_N} \sup_{\lambda \in \Delta_{0,N}(\xi)} \left\{ \xi_b^\top \lambda - \sup_{\lambda \in \hat{\Delta}} \xi_b^\top \lambda \right\} \le \beta_N, \tag{7.13}$$

where $\hat{\Delta}$ is as in equation 4.5. Let $\mathcal{D}$ and $\mathcal{D}_N(\xi)$ denote respectively the feasible regions of the perturbed and unperturbed LP, and defined by

$$\mathcal{D} = \{\lambda \in \mathbb{R}_+^p \mid \|D\lambda\|_1 \le 1\} \quad \text{and} \quad \mathcal{D}_N(\xi) = \{\lambda \in \mathbb{R}_+^p \mid \|D_N(\xi)\lambda\|_1 \le 1\}.$$

We first show that

$$\sup_{\xi \in \mathcal{X}_N} d_H(\mathcal{D}, \mathcal{D}_N(\xi)) \le C\alpha_N. \tag{7.14}$$

Indeed, for $\xi \in \mathcal{X}_N$ and for all N sufficiently large that $\alpha_N < \underline{\sigma}/2$, the diagonal entries of the matrix $D_N(\xi)$ are strictly positive and bounded below by $\underline{\sigma}/2$, where $\underline{\sigma}$ is as in Assumption 4.3. We have

$$\sigma_{min}(D_N(\xi)D^{-1})\|D\lambda\|_1 \le \|D_N(\xi)\lambda\|_1 \le \sigma_{max}(D_N(\xi)D^{-1})\|D\lambda\|_1$$

where $\sigma_{min}(D)$ and $\sigma_{max}(D)$ denote respectively the smallest and largest eigenvalue of the diagonal matrix D. As the sets $\mathcal{D}$ and $\mathcal{D}_N(\xi)$ are star-shaped, [30] the latter inequality implies that

$$\left[\lambda \in \mathcal{D}_N(\xi)\right] \Rightarrow \left[\sigma_{min}(D_N(\xi)D^{-1})\lambda \in \mathcal{D}\right] \tag{7.15}$$

[30] We say that a set $\mathcal{C}$ is star-shaped (with respect to the origin) if $x \in \mathcal{C}$ implies that $tx \in \mathcal{C}$ for all $t \in [0, 1]$.

and

$$\left[\lambda \in \mathcal{D}\right] \Rightarrow \left[\sigma_{min}(D_N(\xi)^{-1}D)\lambda \in \mathcal{D}_N(\xi)\right]. \tag{7.16}$$

As $\|D_N(\xi) - D\| \leq \alpha_N \sqrt{N}$ on $\mathcal{X}_N$, we have

$$\sigma_{min}(D_N(\xi)D^{-1}) \wedge \sigma_{min}(D_N(\xi)^{-1}D) \geq 1 - C\alpha_N \tag{7.17}$$

for all sufficiently large N's. Note that for all sufficiently large N's, the sets $\mathcal{D}$ and $\mathcal{D}_N(\xi)$ are bounded in the l_1-norm by $2/\underline{\sigma}$. Therefore, combining equations 7.15, 7.16 and 7.17, and the fact that $\mathcal{D}$ and $\mathcal{D}_N(\xi)$ are star-shaped imply that inequality 7.14 holds. Note that on the event $E_N(Q)$, we have $\hat{\xi} = (\sqrt{N}(\hat{b} - b(Q)), \sqrt{N}(\hat{D} - D(Q))) \in \mathcal{X}_N$. Hence, the preceding argument also yields that on the event $E_N(Q)$ we have

$$d_H(\mathcal{D}, \hat{\mathcal{D}}) \leq C\alpha_N. \tag{7.18}$$

where $\hat{\mathcal{D}} := \{\lambda \in \mathbb{R}^p_+ \mid \|\hat{D}\lambda\|_1 \leq 1\}$. Moreover, inequality 7.17 combined with its analogue for $\hat{D}$ and D, imply that every $\lambda \in \mathcal{D}$ has a positive scalar multiple that is in $\hat{\mathcal{D}}_N(\xi)$ and that is at a distance of at most $C\alpha_N$ from λ, and vice versa.

Given $\lambda \in \mathbb{R}^p$, let $\hat{\Pi}(\lambda)$ denote the element on the ray $\{t\lambda \mid t \geq 0\}$ that is closest to λ and that is in $\hat{\mathcal{D}}$. That is $\hat{\Pi}(\lambda)$ is the projection of λ on the segment $\{t\lambda \mid t \geq 0\} \cap \hat{\mathcal{D}}$. We now show that for all $\xi \in \mathcal{X}_N$, $Q \in \mathcal{Q}_\prime$, and all sufficiently large N's, on the event $E_N(Q)$ we have

$$\hat{\Pi}(\Delta_{0,N}(\xi)) \subset \hat{\Delta}. \tag{7.19}$$

Indeed, we have

$$\hat{T}_N = \sup_{\lambda \in \hat{\mathcal{D}}} \hat{b}^\top \lambda = \sup_{\lambda \in \hat{\mathcal{D}}} (\hat{b} - b)^\top \lambda + b^\top \lambda \leq \sup_{\lambda \in \hat{\mathcal{D}}} (\hat{b} - b)^\top \lambda \leq CM_N/\sqrt{N},$$

where we have used the fact that $Q \in \mathcal{Q}_0$ implies that $b^\top \lambda \leq 0$ for all $\lambda \in \mathcal{D}$ (as well as those in $\hat{\mathcal{D}}$, by homogeneity), and on $E_N(Q)$ the sets $\hat{\mathcal{D}}$ are uniformly bounded for all large N (such that $\alpha_N \leq \underline{\sigma}/2$) and $\|\hat{b} - b\| \leq M_N/\sqrt{N}$. And for all $\lambda \in \Delta_{0,N}(\xi)$, we have

$$\hat{b}^\top \hat{\Pi}(\lambda) = \left(\hat{b} - b_N(\xi)\right)^\top \hat{\Pi}(\lambda) + b_N(\xi)^\top \hat{\Pi}(\lambda) \geq \left(\hat{b} - b_N(\xi)\right)^\top \hat{\Pi}(\lambda) \geq -CM_N/\sqrt{N},$$

since $b_N(\xi)^\top \lambda \geq 0$ for all $\lambda \in \Delta_{0,N}(\xi)$ (note that we always have $0 \in \mathcal{D}_N(\xi)$), and $\|\hat{b} - b_N(\xi)\| \leq 2M_N/\sqrt{N}$ on the event $E_N(Q)$ and for $\xi \in \mathcal{X}_N$. As $M_N = o(\kappa_N)$, the latter two inequalities imply that for all $\xi \in \mathcal{X}_N$, $Q \in \mathcal{Q}_\prime$, and all sufficiently large N's, on the event $E_N(Q)$, we have

$$\lambda \in \Delta_{0,N}(\xi) \quad \Rightarrow \quad \hat{b}^\top \hat{\Pi}(\lambda) \geq \hat{T}_N - \kappa_N/\sqrt{N}$$

which yields 7.19.

We now establish equation 7.13. For all N sufficiently large, $\xi \in \mathcal{X}_N$, $\lambda \in \Delta_{0,N}(\xi)$, $Q \in \mathcal{Q}_0$, and on the event $E_N(Q)$, we have

$$\xi_b^\top \lambda = \xi_b^\top(\lambda - \hat{\Pi}(\lambda)) + \xi_b^\top \hat{\Pi}(\lambda) \le CM_N\alpha_N + \sup_{\lambda\in\hat{\Lambda}} \xi_b^\top \lambda$$

where we have used the fact that every $\lambda \in \mathcal{D}_N(\xi)$ has a positive scalar multiple that is in $\hat{\mathcal{D}}$ and that is at a distance of at most $C\alpha_N$ from λ, equation 7.19, and the fact that $\|\xi_b\| \le M_N$ for $\xi \in \mathcal{X}_N$. Equation 7.13 then follows by taking $\beta_N = C\alpha_N M_N$, and noting that M_N is chosen such that $M_N\alpha_N = o(1)$.

Step 2 We now use equation (7.13) to couple our test statistic $\hat{T}_N$ with a dominating statistic (in the sense of first-order stochastic dominance) whose distribution can be estimated uniformly with respect to $Q \in \mathcal{Q}_0$. This step culminates in equation (7.26), which establishes inequality (4.6) in Theorem 4.7. Let $\hat{\xi}^* = (\sqrt{N}(\hat{b}^* - b(Q)), \sqrt{N}(\hat{D}^* - D(Q)))$ be an independent and identically distributed version of the root $\hat{\xi}(Q) = (\sqrt{N}(\hat{b} - b(Q)), \sqrt{N}(\hat{D} - D(Q)))$, computed from a sample $\mathbb{W}_N^*$, such that $\mathbb{W}_N^*$ and $\mathbb{W}_N$ are independent and identically distributed. Using the notation of Section 4.1 that is introduced prior to the derivation of equation 4.8, let $\hat{T}_N^* = \phi(\mu_N(\hat{\xi}^*))$ be an independent version of $\hat{T}_N$ constructed from $\hat{\xi}^*$. For $Q \in \mathcal{Q}$, let the event $F_N^* = F_N^*(Q)$ be defined by

$$F_N^*(Q) := \left\{ \mathbb{W}^* \;\middle|\; \|\hat{\xi}_b^*\| > M_N \text{ or } \|\hat{\xi}_D^*\| > \sqrt{N}\alpha_N \right\}.$$

By our choice of M_N and α_N, we have

$$\delta_N := \sup_{Q\in\mathcal{Q}} Q\big(F_N^*(Q)\big) \quad \text{satisfies} \quad \lim_{N\to\infty} \delta_N = 0$$

Equations 7.12 and 7.13 then imply that on the event $E_N(Q)$, $Q \in \mathcal{Q}_0$, we have

$$\sqrt{N}\hat{T}_N^* \le \sup_{\lambda\in\hat{\Lambda}} \langle \hat{\xi}_b^*, \lambda\rangle + \beta_N + \infty \cdot \mathbb{1}_{\{F_N^*\}}. \tag{7.20}$$

Let $J_N(\cdot; Q)$ denote the CDF of $\sqrt{N}\hat{T}_N$. Define

$$\hat{G}_N(\cdot; Q) := \mathcal{L}\left(\sup_{\lambda\in\hat{\Lambda}} \langle \hat{\xi}_b^*, \lambda\rangle + \beta_N + \infty \cdot \mathbb{1}_{\{F_N^*\}} \;\middle|\; \mathbb{W}_N \right)$$

as the conditional (on $\mathbb{W}_N$) CDF of the right-hand side of inequality (7.20), viewed as a random element taking values in $\mathbb{R} \cup \{\infty\}$. Similarly, define

$$\hat{G}_\infty(\cdot; Q) := \mathcal{L}\left(\sup_{\lambda\in\hat{\Lambda}} \langle \zeta, \lambda\rangle \;\middle|\; \mathbb{W}_N \right),$$

where $\zeta \sim N(0, \Sigma(Q))$. And let $\hat{H}_N(;Q)$ be defined as in Theorem 4.7

$$\hat{H}_N(\cdot; Q) := \mathcal{L}\left(\sup_{\lambda\in\hat{\Lambda}} \langle \zeta_N^*, \lambda\rangle \;\middle|\; \mathbb{W}_N \right),$$

where ζ^* is the bootstrap estimate of of ζ that is given in Assumption 4.5. By Strassen's theorem (coupling characterization of the Prokhorov distance), Assumption 4.2, and our choice of α_N, we have on the event $E_N(Q)$

$$
\begin{aligned}
d_{Pr}\left(\hat{G}_N, \hat{G}_\infty\right) &\le d_{Pr}\left(\hat{G}_N, \mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\hat{\xi}^*_b, \lambda\rangle \,\middle|\, \mathbb{W}_N\right)\right) \\
&\quad + d_{Pr}\left(\hat{G}_\infty, \mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\hat{\xi}^*_b, \lambda\rangle \,\middle|\, \mathbb{W}_N\right)\right) \\
&\le \beta_N + Q(F^*_N) + C\alpha_N
\end{aligned}
\tag{7.21}
$$

where we have used in part the fact that $d_{Pr}(\hat{\xi}^*_b, \zeta) \le \alpha_N$ and Strassen's theorem imply that[31]

$$
d_{Pr}\left(\hat{G}_\infty, \mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\hat{\xi}^*_b, \lambda\rangle \,\middle|\, \mathbb{W}_N\right)\right) = d_{Pr}\left(\mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\zeta, \lambda\rangle \,\middle|\, \mathbb{W}_N\right), \mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\hat{\xi}^*_b, \lambda\rangle \,\middle|\, \mathbb{W}_N\right)\right) \le C\alpha_N,
$$

since for all sufficiently large N, the sets $\hat{\Delta}$ are uniformly bounded on the event E_N. Similarly, since we have the inequality

$$
d_{Pr}\left(\mathcal{L}(\zeta^* \mid \mathbb{W}), \mathcal{N}(0, \Sigma(Q))\right) \le \alpha_N
$$

on $E_N(Q)$ (by the definition of $E_N(Q)$), it follows from Strassen's theorem that

$$
d_{Pr}\left(\hat{G}_\infty, \hat{H}_N\right) \le C\alpha_N
$$

on $E_N(Q)$, for all sufficiently large N.

In conclusion, there exists $\gamma_N = o(1)$ such that for all sufficiently large N and all $Q \in \mathcal{Q}$, on the event $E_N(Q)$ we have

$$
d_{Pr}\left(\hat{G}_N, \hat{G}_\infty\right) \vee d_{Pr}\left(\hat{H}_N, \hat{G}_\infty\right) \le \gamma_N. \tag{7.22}
$$

Let the significance level α and the critical value $\hat{c}_N(1-\alpha)$ be as in Theorem 4.7. For now, assume that there exists a constant C such that for all sufficiently large N, for all $Q \in \mathcal{Q}_0$, and for all sample realizations $\mathbb{W}_N$ on the event $E_N(Q) \cap \{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}$, we have the anti-concentration condition

$$
\hat{G}_\infty\left(\hat{c}_N(1-\alpha) + \gamma_N\right) - \hat{G}_\infty\left(\hat{c}_N(1-\alpha) - \gamma_N\right) \le C\gamma_N. \tag{7.23}
$$

We prove in the next step that inequality (7.23) indeed holds. For $\mathbb{W}_N \in E_N(Q) \cap \{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}$, and $Q \in \mathcal{Q}_0$, we then have[32]

$$
\begin{aligned}
J_N\left(\hat{c}_N(1-\alpha)\right) &\ge \hat{G}_N\left(\hat{c}_N(1-\alpha)\right) \\
&\ge \hat{G}_\infty\left(\hat{c}_N(1-\alpha) - \gamma_N\right) - \gamma_N \\
&\ge \hat{G}_\infty\left(\hat{c}_N(1-\alpha) + \gamma_N\right) - C\gamma_N \\
&\ge \hat{H}_N\left(\hat{c}_N(1-\alpha)\right) - C\gamma_N \\
&\ge 1 - \alpha - C\gamma_N
\end{aligned}
\tag{7.24}
$$

[31]Strassen's theorem implies that if $d_{Pr}(X, Y) \le \epsilon$ and f is a Lipschitz function that satisfies $\|f\|_{Lip} \le C$, then $d_{Pr}(f(X), f(Y)) \le (C \vee 1)\epsilon$. The Lipschitz boundedness in our case follows since $\hat{\Delta}$ are uniformly (w.r.t sufficiently large N and $Q \in \mathcal{Q}_0$, and on the event $E_N(Q)$) bounded.

[32]This argument is similar to the proof of Theorem 2.1 in Romano and Shaikh (2012)

where we have used (7.23) to derive the third inequality, and where the other inequalities follow from (7.22) and the fact that $\hat{H}_N(\hat{c}_N(1-\alpha)) \geq 1-\alpha$ by the definition of $\hat{c}_N(1-\alpha)$. Let $q_N(\cdot)$ denote the quantiles of $J_N(\cdot)$. Inequality 7.24 then implies that on the event $E_N(Q) \cap \{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}$, with $Q \in \mathcal{Q}_0$, we have

$$q_N(1-\alpha-C\gamma_N) \leq \hat{c}_N(1-\alpha). \tag{7.25}$$

We are now ready to establish inequality 4.6 from the theorem. We have

$$\begin{aligned}
\sup_{Q\in\mathcal{Q}_0} Q\Big(\{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}\Big) &\leq \sup_{Q\in\mathcal{Q}_0} Q\Big(\{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}\cap E_N(Q)\Big) + \delta_N \\
&\leq \sup_{Q\in\mathcal{Q}_0} Q\Big(\{\sqrt{N}\hat{T}_N > q_N(1-\alpha-C\gamma_N)\}\Big) + \delta_N \\
&\leq \alpha + C\gamma_N + \delta_N.
\end{aligned} \tag{7.26}$$

which yields equation 4.6.

<u>Step 3</u> We now establish the anti-concentration condition in equation 7.23. First, note that on the event $\{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}$—that is, when the test rejects the null—the optimum of the linear program defining $\hat{T}_N$ is achieved at a vertex of the feasible set other than the origin, as rejection implies that $\hat{T}_N > 0$, and $\hat{T}_N$ is given by

$$\hat{T}_N = \max\left\{\hat{b}^\top\lambda \;\middle|\; \lambda\in\mathbb{R}^p_+,\ \|\hat{D}\lambda\|_1 \leq 1\right\}.$$

It is straightforward to verify that the set of extreme points of the feasible region $\hat{\mathcal{D}}$, denoted $\mathrm{extr}(\hat{\mathcal{D}})$, is given by

$$\mathrm{extr}(\hat{\mathcal{D}}) = \left\{0, \frac{e_1}{\hat{\sigma}_1}, \ldots, \frac{e_p}{\hat{\sigma}_p}\right\},$$

where e_j is the jth canonical basis vector in $\mathbb{R}^p$ (i.e., the vector with a 1 in the jth coordinate and zeros elsewhere), and $\hat{\sigma}_i$ denotes the ith diagonal element of $\hat{D}$. For $i = 1, \ldots, p$, define $\hat{\lambda}_i := e_i/\hat{\sigma}_i$.

From this, we conclude that on the event of a rejection, it must be that (with $\hat{\Delta}$ as in equation 4.5)

$$\hat{\Delta} \cap \left\{\hat{\lambda}_i \mid i = 1, \ldots, p\right\} \neq \emptyset. \tag{7.27}$$

Moreover, for all N sufficiently large such that $\alpha_N \leq \underline{\sigma}/2$, and for all $Q \in \mathcal{Q}$, on the event $E_N(Q)$, for any $i \in [p]$, we have

$$\mathcal{L}\left(\langle\zeta, \hat{\lambda}_i\rangle \mid \mathbb{W}_N\right) \sim \mathcal{N}\left(0, \frac{\sigma_i(Q)^2}{\hat{\sigma}_i^2}\right) \quad \text{and} \quad \frac{\sigma_i(Q)^2}{\hat{\sigma}_i^2} \geq 4/9 \tag{7.28}$$

where $\sigma_i(Q)^2$ denotes the i^{th} diagonal element of the asymptotic covariance matrix $\Sigma(Q)$. Indeed, on the event $E_N(Q)$, we have $|\hat{\sigma}_i - \sigma_i| \leq \alpha_N$, and the condition $\alpha_N \leq \underline{\sigma}/2$ implies that $\hat{\sigma}_i \leq \frac{3}{2}\sigma_i$ for any $i \in \{1, \ldots, p\}$, and thus $\sigma_i/\hat{\sigma}_i \geq \frac{2}{3}$.

Recalling that the sets $\hat{\mathcal{D}}$ (and thus $\hat{\Delta}$) are uniformly bounded for all sufficiently large N and for all $Q \in \mathcal{Q}$, equations 7.27 and 7.28 can be used to verify the conditions of Proposition 7.2, which yields the desired conclusion, namely equation 7.23.

<u>Step 4</u> Fix $Q \in \mathcal{Q}_0$. Under Assumption 4.2, Lemma 7.3 implies that the asymptotic distribution of $\sqrt{N}\hat{T}_N$

is given by

$$\max\left\{\zeta^\top\lambda \;\middle|\; \lambda \in \Delta_0(Q)\right\}, \tag{7.29}$$

where $\Delta_0(Q) := \arg\max\{b(Q)^\top\lambda \mid \lambda \in \mathbb{R}^p_+, \|D(Q)\lambda\|_1 \le 1\}$ and $\zeta \sim \mathcal{N}(0, \Sigma(Q))$. Let $G_\infty(\cdot)$ denote the CDF of (7.29). Then $d_{\Pr}(\mathcal{L}(\sqrt{N}\hat{T}_N), G_\infty) = o(1)$. To prove (4.7), it suffices to show

$$d_{\Pr}(\hat{H}_N, G_\infty) = o_p(1), \tag{7.30}$$

where $\hat{H}_N$ is the conditional (given $\mathbb{W}_N$) CDF in (4.5). We show below that $d_H(\Delta_0(Q), \hat{\Delta}) = o_p(1)$, from which (7.30) follows via an almost sure representation argument as in Theorem 3.4.[33]

From (7.18) and (7.19) (with $\xi = 0$), on $E_N(Q)$ we have

$$d_H(\mathcal{D}, \hat{\mathcal{D}}) \le C\alpha_N, \quad \text{and} \quad \vec{d}_H(\Delta_0, \hat{\Delta}) \le C\alpha_N. \tag{7.31}$$

It remains to show $\vec{d}_H(\hat{\Delta}, \Delta_0) \to 0$ on $E_N(Q)$. Fix $\epsilon > 0$. On $E_N(Q)$, (7.31) implies[34]

$$\hat{\Delta} \subset \hat{\mathcal{D}} \subseteq [\mathcal{D}]^{C\alpha_N},$$

and for any $\lambda \in \hat{\mathcal{D}}$, there exists $\Gamma(\lambda) \in \mathcal{D}$ with $\|\lambda - \Gamma(\lambda)\| \le C\alpha_N$. If $\lambda \in \hat{\mathcal{D}} \backslash [\Delta_0]^\epsilon$, then $d(\Gamma(\lambda), \Delta_0) > \epsilon - C\alpha_N > \epsilon/2$ for all large N. By the definition of Δ_0, there exists $\delta > 0$ such that $\langle b, \lambda\rangle < -\delta$ for all $\lambda \in \mathcal{D}\backslash[\Delta_0]^\epsilon$. Then on $E_N(Q)$, for $\lambda \in \hat{\mathcal{D}}\backslash[\Delta_0]^\epsilon$ and all large N,

$$\begin{aligned} \langle \hat{b}, \lambda\rangle &= \langle \hat{b}, \lambda - \Gamma(\lambda)\rangle + \langle \hat{b} - b, \Gamma(\lambda)\rangle + \langle b, \Gamma(\lambda)\rangle \\ &\le C(M_N/\sqrt{N} + \alpha_N) - \delta \le -\delta/2, \end{aligned} \tag{7.32}$$

where we have used the fact that for $Q \in \mathcal{Q}_0$, $\langle b, \lambda\rangle \le 0$ for any $\lambda \in \mathcal{D}$. Also, from the proof of (7.19), for $Q \in \mathcal{Q}_0$, we have $\hat{T}_N \le CM_N/\sqrt{N}$ on $E_N(Q)$, so

$$\hat{T}_N - \kappa_N/\sqrt{N} = o(1)$$

since $\kappa_N = o(\sqrt{N})$. Combining this with (7.32), we get $\hat{\Delta} \subset [\Delta_0]^\epsilon$ on $E_N(Q)$ for all large N. As $\epsilon > 0$ is arbitrary and $Q(E_N) \to 1$, this yields $\vec{d}_H(\hat{\Delta}, \Delta_0) = o_p(1)$, as claimed. □

Proof of Theorem 4.14. As in the proof of Theorem 4.7, let $L(\lambda, x, y)$ denote the Lagrangian associated with the LP in (4.11), defined as

$$L(\lambda, x, y) = b^\top\lambda + x(1 - \mathbb{1}^\top D\lambda) - y^\top A^\top\lambda,$$

where x is a nonnegative scalar and $y \in \mathbb{R}^p$ are the Lagrange multipliers. Let Δ_0 and S_0 denote, respectively, the sets of optimal solutions to the primal and dual problems associated with the LP in (4.11).

To analyze the value function under perturbations, let the input triple be denoted by $\mu = (b, A, D)$, and let

[33] If $d_H(\Delta_n, \Delta) \to 0$ and $\zeta_n \xrightarrow{a.s.} \zeta$, then $\max_{\lambda\in\Delta_n} \zeta_n^\top\lambda \to \max_{\lambda\in\Delta} \zeta^\top\lambda$ a.s.

[34] We use $[A]^\epsilon$ to denote the ϵ neighborhood of the set A.

$\phi(\mu)$ denote the value of the LP in (4.11) given these inputs. For perturbations $\xi = (\xi_b, \xi_A, \xi_D)$ and sample size N, define the perturbed inputs as $\mu_N(\xi) = (b_N(\xi), A_N(\xi), D_N(\xi)) = \left(b + \frac{\xi_b}{\sqrt{N}}, A + \frac{\xi_A}{\sqrt{N}}, D + \frac{\xi_D}{\sqrt{N}}\right)$.

Since the matrix A is known, we will only consider perturbations with $\xi_A = 0$, and by abuse of notation, write $\xi = (\xi_b, \xi_D)$ and $\mu_N(\xi) = (b_N(\xi), D_N(\xi))$. Let $L_N(\lambda, x, y; \xi)$ denote the Lagrangian associated with the LP in (4.11) under the perturbed inputs $\mu_N(\xi)$, defined by

$$L_N(\lambda, x, y; \xi) = b_N(\xi)^\top \lambda + x(1 - \mathbb{1}^\top D_N(\xi)\lambda) - y^\top A^\top \lambda.$$

Let $\Delta_{0,N}(\xi)$ and $S_{0,N}(\xi)$ denote, respectively, the sets of optimal solutions to the perturbed primal and dual problems. Note that the functions L and L_N, as well as the sets Δ_0, $\Delta_{0,N}$, and others, all depend on μ (or on Q); however, this dependence is omitted for notational simplicity.

The first step in our proof, as in that of Theorem 4.7, is to establish the analogue of equation (4.8). To that end, let $(\lambda_0, x_0, y_0) \in \Delta_0 \times S_0$ and $(\lambda_\xi, x_\xi, y_\xi) \in \Delta_{0,N}(\xi) \times S_{0,N}(\xi)$. By the saddle point property of the Lagrangian, for sufficiently small perturbations such that the LP remains feasible, we have

$$\begin{aligned}
\sqrt{N}\big(\phi(\mu_N(\xi)) - \phi(\mu)\big) &= \sqrt{N}\big(L_N(\lambda_\xi, x_\xi, y_\xi; \xi) - L(\lambda_0, x_0, y_0)\big) \\
&\leq \sqrt{N}\big(L_N(\lambda_\xi, x_0, y_0; \xi) - L(\lambda_\xi, x_0, y_0)\big) \\
&= \sqrt{N}(b_N(\xi) - b)^\top \lambda_\xi - x_0 \mathbb{1}^\top \sqrt{N}(D_N(\xi) - D)\lambda_\xi.
\end{aligned}$$

Now, for $Q \in \mathcal{Q}_0$ we have $\phi(\mu) = 0$, and $(x_0, y_0) \in S_0$ implies that $x_0 = 0$ and $b - Ay_0 \leq 0$. Therefore, the inequality simplifies to

$$\sqrt{N}\,\phi(\mu_N(\xi)) \leq \xi_b^\top \lambda_\xi, \tag{7.33}$$

which corresponds exactly to inequality (4.8).

Now, note that Assumptions 4.2 and 4.12-i) imply that the family $\{\mathcal{N}(0, \Sigma(Q)) \mid Q \in \mathcal{Q}\}$ is tight, and Assumption 4.2 then implies that the family $\{\sqrt{N}(\hat{b} - b(Q)) \mid Q \in \mathcal{Q}\}$ is asymptotically tight. Thus, Assumptions 4.2, 4.5 and 4.12 imply that there exists $\alpha_N \downarrow 0$, such that for any $M_N \uparrow \infty$, we have

$$\sup_{Q \in \mathcal{Q}} d_{\Pr}\Big(\sqrt{N}(\hat{b} - b), \mathcal{N}(0, \Sigma(Q))\Big) \leq \alpha_N$$

and for all $Q \in \mathcal{Q}$, the event $E_N(Q)$ defined by

$$E_N(Q) := \Big\{\mathbb{W} \mid \|\hat{D} - D\| \leq \alpha_N,\ d_{\Pr}\big(\mathcal{L}(\zeta^* \mid \mathbb{W}), \mathcal{N}(0, \Sigma(Q))\big) \leq \alpha_N, \sqrt{N}\|\hat{b} - b(Q)\| \leq M_N\Big\}$$

satisfies

$$\underline{\lim}_{N \to \infty} \inf_{Q \in \mathcal{Q}} Q(E_N(Q)) = 1.$$

Choose $M_N \uparrow \infty$ such that $M_N = o(\kappa_N)$ and $M_N \alpha_N = o(1)$. When N is sufficiently large that $\alpha_N < \underline{\sigma}$, the indices of the diagonal entries of the matrices $\hat{D}$ and D that are nonzero coincide, and correspond to the indices of the estimated components of b. In what follows, we let Π^u denote the projection on the unknown components of b that must be estimated. [35] Consider the set of pertubations $\mathcal{X}_N$ defined by

$$\mathcal{X}_N := \left\{ \xi = (\xi_b, \xi_D) \;\middle|\; \|\xi_b\| \leq M_N,\ \|\xi_D\| \leq \sqrt{N}\alpha_N \right\},$$

where, ξ_D denotes diagonal matrices of the same dimensions as D, with diagonal entries corresponding to the deterministic components of b equal to zero, and ξ_b is a vector of the same dimension as b with entries corresponding to the deterministic component of b equal to zero. It is natural to consider such restrictive perturbations ξ_D, as they are similar to perturbations that arise from estimation ($\hat{D} - D$ has diagonal entries corresponding to known components of b equal to zero). Throughout the argument below, we use C to represent a generic constant that does not depend on $Q \in \mathcal{Q}$ or on N, for all sufficiently large N. The value of C may vary from line to line.

In the first three steps below, we establish inequality (4.15). The fourth step provides the proof of inequality (4.16). These steps closely mirror those used in the proof of Theorem 4.7, and we occasionally refer to that proof when the arguments are essentially the same.

Step 1 In this step, we show there exists $\beta_N = o(1)$ such that for all sufficiently large N's and for all $Q \in \mathcal{Q}_0$, we have on the event $E_N(Q)$:

$$\sup_{\xi \in \mathcal{X}_N} \sup_{\lambda \in \Delta_{0,N}(\xi)} \left\{ \xi_b^\top \lambda - \sup_{\lambda \in \hat{\Delta}} \xi_b^\top \lambda \right\} \leq \beta_N, \tag{7.34}$$

where $\hat{\Delta}$ is as in equation 4.14. Let $\mathcal{D}$ and $\mathcal{D}_N(\xi)$ denote respectively the feasible regions of the unperturbed and perturbed LP. Note that since some diagonal entries of the matrices D and $D_N(\xi)$ can be zero, it is possible for the sets $\mathcal{D}$ and $\mathcal{D}_N(\xi)$ to be unbounded. we first show that for all sufficiently large N, for all $Q \in \mathcal{Q}_0$, and $\xi \in \mathcal{X}_N$, we have that for each $\lambda \in \mathcal{D}_N(\xi)$ there exists $\Gamma(\lambda) \in \hat{\mathcal{D}}$ such that $\Gamma(\lambda)$ is a positive scalar multiple of λ and

$$\|\Pi^u(\lambda) - \Pi^u(\Gamma(\lambda))\| \leq C\alpha_N \tag{7.35}$$

where Π^u is as in Footnote 35. Note that for all large N such that $\alpha_N < \underline{\sigma}/2$ (with $\underline{\sigma}$ given in Assumption 4.12) and for $\xi \in \mathcal{X}_N$, we have

$$\|D_N(\xi)\lambda\|_1 = \|\Pi^u(D_N(\xi)\lambda)\|_1 \geq \underline{\sigma}/2\|\Pi^u(\lambda)\| \qquad \forall \lambda \in \mathbb{R}^p,$$

which yields[36]

$$\sup_{\lambda \in \mathcal{D}_N(\xi)} \|\Pi^u(\lambda)\|_1 \leq 2/\underline{\sigma}. \tag{7.36}$$

Focusing on the component of λ that corresponds to the estimated part of b, and arguing as in the proof of Theorem 4.7, we have $\forall \lambda \in \mathbb{R}^p$

$$(1 - \alpha_N/\underline{\sigma})\|\Pi^u(D\lambda)\|_1 \leq \|\Pi^u(D_N(\xi)\lambda)\|_1 \leq (1 + \alpha_N/\underline{\sigma})\|\Pi^u(D\lambda)\|_1. \tag{7.37}$$

[35] For example, if $b = (b_1, b_2, b_3, b_4)$, and b_1 and b_3 are known while b_2 and b_4 are unknown and must be estimated, then $\Pi^u(b)$ denotes the vector $(0, b_2, 0, b_4)$. Similarly, $\Pi^u(\lambda) = (0, \lambda_2, 0, \lambda_4)$.

[36] The same is true on the event $E_N(Q)$, when $\mathcal{D}_N(\xi)$ is replaced by $\hat{\mathcal{D}}$.

As a consequence, since the sets $\mathcal{D}$ and $\mathcal{D}_N(\xi)$ are star-shaped w.r.t. the origin, for each $\lambda \in \mathcal{D}_N(\xi)$, we have $(1-\alpha_N/\underline{\sigma})\lambda \in \mathcal{D}$. Similarly, since $\|\hat{D}-D\| \leq \alpha_N$ on $E_N(Q)$, for all $\lambda \in \mathbb{R}^p$ and N such that $\alpha_N < \underline{\sigma}/2$, we have

$$(1-2\alpha_N/\underline{\sigma})\|\Pi^u(\hat{D}\lambda)\|_1 \leq \|\Pi^u(D\lambda)\|_1 \leq (1+2\alpha_N/\underline{\sigma})\|\Pi^u(\hat{D}\lambda)\|_1. \tag{7.38}$$

Hence, for each $\lambda \in \mathcal{D}$, we have $(1-2\alpha_N/\underline{\sigma})\lambda \in \hat{\mathcal{D}}$. Combining the latter two observations implies that for all $\lambda \in \mathcal{D}_N(\xi)$, we can set $\Gamma(\lambda) = (1-\alpha_N/\underline{\sigma})(1-2\alpha_N/\underline{\sigma})\lambda \in \hat{\mathcal{D}}$, and we get

$$\|\Pi^u(\Gamma(\lambda)) - \Pi^u(\lambda)\| \leq C\alpha_N\|\Pi^u(\lambda)\| \leq C\alpha_N,$$

where we have used equation 7.36. This proves equation 7.35.

Returning to the proof of equation 7.34, for $\xi \in \mathcal{X}_N$ and on the event $E_N(Q)$, we have

$$\begin{aligned}\sup_{\lambda\in\Delta_{0,N}(\xi)} \xi_b^\top\lambda &\leq \sup_{\lambda\in\Delta_{0,N}(\xi)} \xi_b^\top\left(\Pi^u(\lambda-\Gamma(\lambda))\right) + \sup_{\lambda\in\Delta_{0,N}(\xi)} \xi_b^\top\Gamma(\lambda)\\ &\leq CM_N\alpha_N + \sup_{\lambda\in\Delta_{0,N}(\xi)} \xi_b^\top\Gamma(\lambda),\end{aligned}$$

where we have used equation 7.35 and the fact that the part of ξ_b that corresponds to the deterministic component of b is equal to zero. Inequality 7.34 then follows, with $\beta_N = CM_N\alpha_N$, if we show that for all $Q \in \mathcal{Q}_0$, $\xi \in \mathcal{X}_N$, and sufficiently large N, we have

$$\{\Gamma(\lambda) \mid \lambda \in \Delta_{0,N}(\xi)\} \subseteq \hat{\Delta} \tag{7.39}$$

on the event $E_N(Q)$. We now proceed to prove 7.39. For $\lambda \in \hat{\mathcal{D}}$, for $Q \in \mathcal{Q}_0$ and on the event $E_N(Q)$, we have

$$\langle\hat{b},\lambda\rangle = \langle\hat{b}-b,\Pi^u(\lambda)\rangle + \langle b,\lambda\rangle \leq CM_N/\sqrt{N},$$

where we have used 7.36 and the fact that for $Q \in \mathcal{Q}_0$, $b^\top\lambda \leq 0$ for all $\lambda \in \mathcal{D}$ (and thus in $\hat{\mathcal{D}}$). This yields

$$0 \leq \hat{\vartheta} \leq CM_N/\sqrt{N}. \tag{7.40}$$

Also, for $\xi \in \mathcal{X}_N$, $\lambda \in \Delta_{0,N}(\xi)$, and on the event $E_N(Q)$, we have

$$\begin{aligned}\langle\hat{b},\Gamma(\lambda)\rangle &= \langle\hat{b}-b,\Pi^u(\Gamma(\lambda))\rangle + \langle b-b_N(\xi),\Pi^u(\Gamma(\lambda)) + \langle b_N(\xi),\Gamma(\lambda)\rangle\rangle\\ &\geq -CM_N/\sqrt{N},\end{aligned}$$

where we have used the fact that $b_N(\xi)^\top\Gamma(\lambda) \geq 0$ for all $\lambda \in \Delta_{0,N}(\xi)$, which holds since $b_N(\xi)^\top\lambda \geq 0$ for all $\lambda \in \Delta_{0,N}(\xi)$ (the origin is always feasible), and $\Gamma(\lambda)$ is a positive scalar multiple of λ. By our choice of M_N, we have $M_N = o(\kappa_N)$, and this combined with the preceding inequalities yields that for all sufficiently large N, for $\xi \in \mathcal{X}_N$ and $Q \in \mathcal{Q}_0$, we have

$$\langle\hat{b},\Gamma(\lambda)\rangle \geq \hat{\vartheta} - \kappa_N/\sqrt{N}$$

for all $\lambda \in \Delta_{0,N}(\xi)$, which yields equation 7.39, and concludes the proof of equation 7.34.

Step 2 In this step, we use equation (7.34) and a coupling argument to upper bound the test statistic by another statistic whose distribution can be uniformly estimated for all $Q \in \mathcal{Q}_0$. The proof closely parallels Step 2 in the proof of Theorem 4.7. Let $\hat{\xi}^* = (\sqrt{N}(\hat{b}^* - b(Q)), \sqrt{N}(\hat{D}^* - D(Q)))$ be an independent and identically distributed version of the root $\hat{\xi}(Q) = (\sqrt{N}(\hat{b} - b(Q)), \sqrt{N}(\hat{D} - D(Q)))$, computed from a sample $\mathbb{W}_N^*$, such that $\mathbb{W}_N^*$ and $\mathbb{W}_N$ are independent and identically distributed. For $Q \in \mathcal{Q}$, let the event $F_N^* = F_N^*(Q)$ be defined by

$$F_N^*(Q) := \left\{ \mathbb{W}^* \;\middle|\; \|\hat{\xi}_b^*\| > M_N \text{ or } \|\hat{\xi}_D^*\| > \sqrt{N}\alpha_N \right\}.$$

By our choice of M_N and α_N, we have

$$\delta_N := \sup_{Q \in \mathcal{Q}} Q\big(F_N^*(Q)\big) \quad \text{satisfies} \quad \lim_{N \to \infty} \delta_N = 0$$

Equations 7.33 and 7.34 then imply that on the event $E_N(Q)$, $Q \in \mathcal{Q}_0$, we have

$$\sqrt{N}\phi(\mu_N(\hat{\xi}^*)) \leq \sup_{\lambda \in \hat{\Lambda}} \langle \hat{\xi}_b^*, \lambda \rangle + \beta_N + \infty \cdot \mathbb{1}_{\{F_N^*\}}. \tag{7.41}$$

Let $J_N(\cdot; Q)$ denote the CDF of $\sqrt{N}\hat{T}_N$. Define

$$\hat{G}_N(\cdot; Q) := \mathcal{L}\left(\sup_{\lambda \in \hat{\Lambda}} \langle \hat{\xi}_b^*, \lambda \rangle + \beta_N + \infty \cdot \mathbb{1}_{\{F_N^*\}} \;\middle|\; \mathbb{W}_N \right)$$

as the conditional (on $\mathbb{W}_N$) CDF of the right-hand side of inequality (7.41), viewed as a random element taking values in $\mathbb{R} \cup \{\infty\}$. Similarly, define

$$\hat{G}_\infty(\cdot; Q) := \mathcal{L}\left(\sup_{\lambda \in \hat{\Lambda}} \langle \zeta, \lambda \rangle \;\middle|\; \mathbb{W}_N \right),$$

where $\zeta \sim N(0, \Sigma(Q))$. And let $\hat{H}_N(;Q)$ be defined as in Theorem 4.14

$$\hat{H}_N(\cdot; Q) := \mathcal{L}\left(\sup_{\lambda \in \hat{\Lambda}} \langle \zeta_N^*, \lambda \rangle \;\middle|\; \mathbb{W}_N \right),$$

where ζ^* is the bootstrap estimate of of ζ that is given in Assumption 4.5.

By Strassen's theorem and our choice of α_N, we have on the event $E_N(Q)$

$$\begin{aligned} d_{\mathrm{Pr}}\left(\hat{G}_N, \hat{G}_\infty\right) &\leq d_{\mathrm{Pr}}\left(\hat{G}_N, \mathcal{L}\left(\sup_{\lambda \in \hat{\Lambda}} \langle \hat{\xi}_b^*, \lambda \rangle \;\middle|\; \mathbb{W}_N \right)\right) \\ &\quad + d_{\mathrm{Pr}}\left(\hat{G}_\infty, \mathcal{L}\left(\sup_{\lambda \in \hat{\Lambda}} \langle \hat{\xi}_b^*, \lambda \rangle \;\middle|\; \mathbb{W}_N \right)\right) \\ &\leq \beta_N + Q(F_N^*) + C\alpha_N. \end{aligned} \tag{7.42}$$

Here, we have in part used the fact that, equation 7.36, the inequality $d_{\mathrm{Pr}}(\hat{\xi}_b^*, \zeta) \leq \alpha_N$ on $E_N(Q)$, and

Strassen's theorem, imply that

$$d_{\mathrm{Pr}}\left(\hat{G}_{\infty}, \mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\hat{\xi}_b^*, \lambda\rangle \,\middle|\, \mathbb{W}_N\right)\right) = d_{\mathrm{Pr}}\left(\mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\zeta, \Pi^u(\lambda)\rangle \,\middle|\, \mathbb{W}_N\right), \mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\hat{\xi}_b^*, \Pi^u(\lambda)\rangle \,\middle|\, \mathbb{W}_N\right)\right) \le C\alpha_N.$$

Similarly, as

$$d_{\mathrm{Pr}}\big(\mathcal{L}(\zeta^* \mid \mathbb{W}), \mathcal{N}(0, \Sigma(Q))\big) \le \alpha_N$$

holds on the event $E_N(Q)$ (by the definition of $E_N(Q)$), it follows from Strassen's theorem that

$$d_{\mathrm{Pr}}\left(\hat{G}_{\infty}, \hat{H}_N\right) \le C\alpha_N$$

on $E_N(Q)$, for all sufficiently large N.

In conclusion, there exists $\gamma_N = o(1)$ such that for all sufficiently large N and all $Q \in \mathcal{Q}_0$, on the event $E_N(Q)$ we have

$$d_{\mathrm{Pr}}\big(\hat{G}_N, \hat{G}_{\infty}\big) \vee d_{\mathrm{Pr}}\big(\hat{H}_N, \hat{G}_{\infty}\big) \le \gamma_N. \tag{7.43}$$

Let the significance level α and the critical value $\hat{c}_N(1-\alpha)$ be as in Theorem 4.14. For now, assume that there exists a constant C such that for all sufficiently large N, for all $Q \in \mathcal{Q}_0$, and for all sample realizations $\mathbb{W}_N$ on the event $E_N(Q) \cap \{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}$, we have the anti-concentration condition

$$\hat{G}_{\infty}\big(\hat{c}_N(1-\alpha) + \gamma_N\big) - \hat{G}_{\infty}\big(\hat{c}_N(1-\alpha) - \gamma_N\big) \le C\gamma_N. \tag{7.44}$$

We prove in the next step that inequality (7.44) indeed holds. For $\mathbb{W}_N \in E_N(Q) \cap \{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}$, and $Q \in \mathcal{Q}_0$, arguing as in 7.24, we have

$$J_N\big(\hat{c}_N(1-\alpha)\big) \ge 1 - \alpha - C\gamma_N. \tag{7.45}$$

Let $q_N(\cdot)$ denote the quantiles of $J_N(\cdot)$. Inequality 7.45 then implies that on the event $E_N(Q) \cap \{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}$, with $Q \in \mathcal{Q}_0$, we have

$$q_N(1 - \alpha - C\gamma_N) \le \hat{c}_N(1-\alpha). \tag{7.46}$$

Arguing as in equation 7.26, we have

$$\sup_{Q\in\mathcal{Q}_0} Q\Big(\{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}\Big) \le \alpha + C\gamma_N + \delta_N. \tag{7.47}$$

which yields equation 4.15 from Theorem 4.14, as $\gamma_N = o(1)$ and $\delta_N = o(1)$.

Step 3 In this step, we establish the anti-concentration condition 7.44. We first recall that as $0 \in \hat{\mathcal{D}}$ and $\lambda \ge 0$ for all $\lambda \in \hat{\mathcal{D}}$, 0 is an extreme point of $\hat{\mathcal{D}}$, and $\mathrm{extr}(\hat{\mathcal{D}}) \ne \emptyset$. It then follows that any face of $\hat{\mathcal{D}}$ contains at least one extreme point of $\hat{\mathcal{D}}$ (see Section 8.5 of Schrijver (1999)). On the event of a rejection, we necessarily have $\hat{\nu} > 0$, in which case $\hat{\Delta}$ contains an element of $\mathrm{extr}(\hat{\mathcal{D}})\backslash\{0\}$. Indeed, as the set

$\hat{\Delta}_0 = \arg\max\{\hat{b}^\top \lambda \mid \lambda \in \hat{\mathcal{D}}\} = \{\lambda \in \hat{\mathcal{D}} \mid \hat{b}^\top \lambda = \hat{v}\}$ is a face of $\hat{\mathcal{D}}$, it must contain an element λ of $\mathrm{extr}(\hat{\mathcal{D}})$, and it must be the case that $\lambda \neq 0$, as $\hat{v} > 0$. Since $\hat{\Delta}_0 \subseteq \hat{\Delta}$, we conclude that $\hat{\Delta} \cap [\mathrm{extr}(\hat{\mathcal{D}}) \backslash \{0\}] \neq \emptyset$.

We first consider the case where $Q \in \mathcal{Q}_0$ is such that $\Delta_0(Q) = \{0\}$. Suppose that the following event occurs

$$E_N(Q) \cap \{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}$$

and let $\hat{\lambda} \in \hat{\Delta} \cap [\mathrm{extr}(\hat{\mathcal{D}}) \backslash \{0\}]$. Then Lemma 7.6 implies that there exists $\lambda \in \mathrm{extr}(\mathcal{D}(Q)) \backslash \{0\}$ such that

$$\|\Pi^u(\hat{\lambda}) - \Pi^u(\lambda)\| \leq C\|\Pi^u(\hat{\lambda})\|\|\hat{D} - D\| \leq C\alpha_N \tag{7.48}$$

where we have used equation 7.36 and the fact that $\|\hat{D} - D\| \leq \alpha_N$ on $E_N(Q)$. The latter inequality implies that

$$\begin{aligned}
\hat{\lambda}^\top \Sigma(Q)\hat{\lambda} &= \Pi^u(\hat{\lambda})^\top \Sigma(Q)\Pi^u(\hat{\lambda}) = \|\Sigma(Q)^{1/2}\Pi^u(\hat{\lambda})\|^2 \\
&\geq (1/2)\|\Sigma(Q)^{1/2}\Pi^u(\lambda)\|^2 - \|\Sigma(Q)^{1/2}(\Pi^u(\hat{\lambda}) - \Pi^u(\lambda))\|^2 \\
&= (1/2)\lambda^\top \Sigma(Q)\lambda - \|\Sigma(Q)^{1/2}(\Pi^u(\hat{\lambda}) - \Pi^u(\lambda))\|^2 \\
&\geq (1/2)\lambda^\top \Sigma(Q)\lambda - C\alpha_N^2
\end{aligned} \tag{7.49}$$

where we have used equation 7.48 and part i) of Assumption 4.12. Hence, using part ii) of Assumption 4.13, for all sufficiently large N and for $Q \in \mathcal{Q}_0$ such that $\Delta_0(Q) = \{0\}$, there exists $\hat{\lambda} \in \hat{\Delta}$ such that

$$\hat{\lambda}^\top \Sigma(Q)\hat{\lambda} \geq \rho/4 \tag{7.50}$$

on the event $E_N(Q) \cap \{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}$.

We now establish the analogue of equation 7.50 for $Q \in \mathcal{Q}_0$ such that $\Delta_0(Q) \neq \{0\}$. By part i) of Assumption 4.13, there exists $\lambda_Q \in \Delta_0(Q)$ such that

$$\lambda_Q^\top \Sigma(Q)\lambda_Q \geq \rho. \tag{7.51}$$

Equation 7.35 (with $\xi = 0$) implies that for all sufficiently large N, and on the event $E_N(Q)$, there exist $\hat{\lambda} \in \hat{\mathcal{D}}(Q)$ such that

$$\|\Pi^u(\hat{\lambda}) - \Pi^u(\lambda_Q)\| \leq C\|\Pi^u(\lambda_Q)\|\|\hat{D} - D\| \leq C\alpha_N. \tag{7.52}$$

Moreover, since equation 7.35 gives $\hat{\lambda}$ as a positive scalar multiple of λ_Q and $\lambda_Q \in \Delta_0(Q)$, for sufficiently large N and on the event $E_N(Q)$, we have

$$|\hat{b}^\top \hat{\lambda}| = |(\hat{b} - b)^\top \hat{\lambda} + b^\top \hat{\lambda}| = |(\hat{b} - b)^\top \Pi^u(\hat{\lambda})| \leq CM_N/\sqrt{N}$$

where we have used homogeneity and the fact that $\lambda_Q \in \Delta_0(Q)$ implies that $b^\top \lambda_Q = 0$, as well as equation 7.36. The latter inequality, equation 7.40, and the fact that $M_N = o(\kappa_N)$, then imply that for sufficiently large N and on the event $E_N(Q)$, we have

$$\left[\hat{b}^\top \hat{\lambda} \geq \hat{v} - \kappa_N/\sqrt{N}\right] \qquad \text{which implies that} \qquad \left[\hat{\lambda} \in \hat{\Delta}\right]. \tag{7.53}$$

Reasoning as in equation 7.49, equation 7.52 and the part i) of Assumption 4.12, then yield

$$\hat{\lambda}^\top \Sigma(Q)\hat{\lambda} \geq (1/2)\lambda_Q^\top \Sigma(Q)\lambda_Q - C\alpha_N^2.$$

For all sufficiently large N, the latter inequality and equation 7.51 then give

$$\hat{\lambda}^\top \Sigma(Q)\hat{\lambda} \geq \rho/4.$$

In conclusion, for all sufficiently large N and for $Q \in \mathcal{Q}_0$ such that $\Delta_0(Q) \neq \{0\}$, on the event $E_N(Q)$, there exists a $\hat{\lambda} \in \hat{\Delta}$ such that

$$\hat{\lambda}^\top \Sigma(Q)\hat{\lambda} \geq \rho/4. \tag{7.54}$$

Combining equations 7.50 and 7.54, we conclude that for all sufficiently large N and for all $Q \in \mathcal{Q}_0$, on the event $E_N(Q) \cap \{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}$, there exists a $\lambda \in \hat{\Delta}$ such that

$$\Pi^u(\hat{\lambda})^\top \Sigma(Q)\Pi^u(\hat{\lambda}) = \hat{\lambda}^\top \Sigma(Q)\hat{\lambda} \geq \rho/4,$$

and the anti-concentration condition 7.44 then follows from Proposition 7.2, using the fact that the sets $\Pi^u(\hat{\Delta})$ are eventually uniformly bounded on the events $E_N(Q)$, and

$$\sup_{\lambda \in \hat{\Delta}} \langle \zeta, \lambda \rangle = \sup_{\lambda \in \hat{\Delta}} \langle \Pi^u(\zeta), \Pi^u(\lambda) \rangle.$$

Step 4 We now prove inequality 4.16. Fix $Q \in \mathcal{Q}_0$. Under Assumptions 4.2, 4.11 and 4.12, Lemma 7.3 implies that the asymptotic distribution of $\sqrt{N}\hat{T}_N$ is given by

$$\max\left\{\zeta^\top \lambda \;\middle|\; \lambda \in \Delta_0(Q)\right\} = \max\left\{\zeta^\top \Pi^u(\lambda) \;\middle|\; \lambda \in \Delta_0(Q)\right\}, \tag{7.55}$$

where $\Delta_0(Q) := \arg\max\{b(Q)^\top \lambda \mid \lambda \in \mathbb{R}_+^p, A^\top \lambda = 0,\ \|D(Q)\lambda\|_1 \leq 1\}$ and $\zeta \sim \mathcal{N}(0, \Sigma(Q))$. Let $G_\infty(\cdot)$ denote the CDF of the value of the program in equation 7.55. Then $d_{Pr}(\mathcal{L}(\sqrt{N}\hat{T}_N), G_\infty) = o(1)$. To prove (4.16), it suffices to show

$$d_{Pr}(\hat{H}_N, G_\infty) = o_p(1), \tag{7.56}$$

where $\hat{H}_N$ is the conditional (given $\mathbb{W}_N$) CDF in (4.14). We show below that

$$d_H(\Pi^u(\Delta_0(Q)), \Pi^u(\hat{\Delta})) = o_p(1), \tag{7.57}$$

from which (7.56) follows from Assumption 4.5, via an almost sure representation argument as in Theorem 3.4.

To establish equation 7.57, we first show that

$$\vec{d}_H\big(\Pi^u(\Delta_0), \Pi^u(\hat{\Delta})\big) = o_p(1), \tag{7.58}$$

and then establish the converse in equation 7.61 below. Arguing as in the proof of equation 7.38, for all N sufficiently large N that $\alpha_N < \underline{\sigma}/2$, and on the event $E_N(Q)$, we have that

$$\forall \lambda \in \mathcal{D} \quad \hat{\Gamma}(\lambda) := (1 - (2/\underline{\sigma})\alpha_N)\lambda \in \hat{\mathcal{D}} \qquad \text{and} \qquad \sup_{\lambda \in \mathcal{D}} \|\Pi^u(\lambda) - \Pi^u(\hat{\Gamma}(\lambda))\| \le C\alpha_N. \tag{7.59}$$

We now show that, on $E_N(Q)$, $\hat{\Gamma}(\Delta_0)$ is included in $\hat{\Delta}$ for all sufficiently large N. Equation 7.59 would then imply that for all sufficiently large N and on the event $E_N(Q)$, we have

$$\vec{d}_H\big(\Pi^u(\Delta_0), \Pi^u(\hat{\Delta})\big) \le \vec{d}_H\big(\Pi^u(\Delta_0), \Pi^u(\hat{\Gamma}(\Delta_0)))\big) + \vec{d}_H\big(\hat{\Gamma}(\Delta_0), \hat{\Delta})\big) = \vec{d}_H\big(\Pi^u(\Delta_0), \Pi^u(\hat{\Gamma}(\Delta_0)))\big) \le C\alpha_N,$$

which would yield equation 7.58. Indeed, on $E_N(Q)$ and for all sufficiently large N, we have

$$\begin{aligned} \inf_{\lambda \in \Delta_0} \langle \hat{b}, \hat{\Gamma}(\lambda) \rangle &= \langle \hat{b} - b, \Pi^u(\hat{\Gamma}(\lambda)) \rangle + \langle b, \hat{\Gamma}(\lambda) \rangle \\ &\ge -\frac{M_N}{\sqrt{N}} \sup_{\lambda \in \hat{\mathcal{D}}} \|\Pi^u(\lambda)\| \\ &\ge -C\frac{M_N}{\sqrt{N}}, \end{aligned} \tag{7.60}$$

where we have used the fact that $\langle b, \hat{\Gamma}(\lambda) \rangle = 0$ since $Q \in \mathcal{Q}_0$, $\lambda \in \Delta_0$, and $\hat{\Gamma}(\lambda)$ is a positive scalar multiple of λ; we have also used the analogue of equation 7.36, for $\hat{\mathcal{D}}$ on the event $E_N(Q)$. Since by equation 7.40 we have $0 \le \hat{\nu} \le CM_N/\sqrt{N}$, equation 7.60 and the fact that $M_N = o(\kappa_N)$ imply that on the event $E_N(Q)$, and for all sufficiently large N, we have

$$\hat{\Gamma}(\Delta_0) \subseteq \{\lambda \in \hat{\mathcal{D}} \mid \langle \hat{b}, \lambda \rangle \ge \hat{\nu} - \kappa_N/\sqrt{N}\}.$$

Hence, $\hat{\Gamma}(\Delta_0)$ is eventually a subset of $\hat{\Delta}$ on the event $E_N(Q)$, and this completes the proof of equation 7.58.

We now establish the converse of equation 7.58, and show that

$$\vec{d}_H\big(\Pi^u(\hat{\Delta}), \Pi^u(\Delta_0)\big) = o_p(1). \tag{7.61}$$

As in equation 7.59, it can be shown that on the event $E_N(Q)$ and for all sufficiently large N such that $\alpha_N < \underline{\sigma}$, we have

$$\forall \lambda \in \hat{\mathcal{D}} \quad \Gamma(\lambda) := (1 - (1/\underline{\sigma})\alpha_N)\lambda \in \mathcal{D} \qquad \text{and} \qquad \sup_{\lambda \in \hat{\mathcal{D}}} \|\Pi^u(\lambda) - \Pi^u(\Gamma(\lambda))\| \le C\alpha_N. \tag{7.62}$$

Moreover, on $E_N(Q)$ and for all sufficiently large N, we have

$$
\begin{aligned}
0 \geq \sup_{\lambda \in \hat{\Delta}} \langle b, \Gamma(\lambda) \rangle &\geq \inf_{\lambda \in \hat{\Delta}} \langle b, \Gamma(\lambda) \rangle \\
&= \inf_{\lambda \in \hat{\Delta}} \langle b - \hat{b}, \Pi^u(\Gamma(\lambda)) \rangle + \langle \hat{b}, \Gamma(\lambda) \rangle \\
&\geq \inf_{\lambda \in \hat{\Delta}} \langle b - \hat{b}, \Pi^u(\Gamma(\lambda)) \rangle - \kappa_N/\sqrt{N} \\
&\geq \inf_{\lambda \in \mathcal{D}} \langle b - \hat{b}, \Pi^u(\lambda) \rangle - \kappa_N/\sqrt{N} \\
&\geq -\frac{M_N}{\sqrt{N}} \sup_{\lambda \in \mathcal{D}} \|\Pi^u(\lambda)\| - \kappa_N/\sqrt{N} \\
&\geq -C\kappa_N/\sqrt{N}
\end{aligned}
\tag{7.63}
$$

where the first inequality follows since $\nu = 0$ (for $Q \in \mathcal{Q}_0$); the third inequality follows from the fact that $\hat{\nu} \geq 0$ (we always have $0 \in \hat{\mathcal{D}}$), $\langle \hat{b}, \lambda \rangle \geq \hat{\nu} - \kappa_N/\sqrt{N} \geq -\kappa_N/\sqrt{N}$ for all $\lambda \in \hat{\Delta}$, and that $\Gamma(\lambda)$ is a scalar multiple of λ with scalar factor in $(0, 1)$; the fourth inequality follows from the fact that $\Gamma(\hat{\Delta}) \subseteq \mathcal{D}$; and the last inequality follows from the fact that as in equation 7.36 we can show that $\sup_{\lambda \in \mathcal{D}} \|\Pi^u(\lambda)\| < 1/\underline{\sigma}$, and that $M_N = o(\kappa_N)$. As $\kappa_N = o(\sqrt{N})$, equation 7.63 implies that for any $\epsilon > 0$, and for all sufficiently large N, we have

$$
\Gamma(\hat{\Delta}) \subseteq \{\lambda \in \mathcal{D} \mid \langle b, \lambda \rangle \geq -\epsilon\}.
$$

As a consequence, Lemma 7.5 then implies that on $E_N(Q)$ we have

$$
\vec{d}_H(\Gamma(\hat{\Delta}), \Delta_0) = o(1). \tag{7.64}
$$

Combining equations 7.62 and 7.64, we conclude that, on the event $E_N(Q)$ and for all sufficiently large N, we have

$$
\begin{aligned}
\vec{d}_H(\Pi^u(\hat{\Delta}), \Pi^u(\Delta_0)) &\leq \vec{d}_H(\Pi^u(\hat{\Delta}), \Pi^u(\Gamma(\hat{\Delta}))) + \vec{d}_H(\Gamma(\hat{\Delta}), \Delta_0) \\
&C\alpha_N + o(1)
\end{aligned}
$$

which establishes equation 7.61, and completes the proof of inequality 4.16.

□

Proof of Theorem 4.21. The proof of equation 4.22 is identical to that of equation 4.16 in Theorem 4.14. For the proof of equation 4.21, we can proceed, with very minor changes, as in Steps 1 and 2 of the proof of Section 7.0.1 (the proof of Theorem 4.14). However, Step 3 in Section 7.0.1 is different in the present context, in view of the new complications that arise from the sets $\mathcal{D}(Q)$ potentially not having extreme points. As the proof of the analogue of Step 3 in the present context is more similar to the one in Section 7.0.1 under assumption 4.19, we only establish Step 3 under Assumption 4.20, which is more closely related to the assumptions in Fang et al. (2023). Also, since $\mathcal{D}(Q)$ is no longer restricted to be a subset of the positive orthant, the constraint $\|D\lambda\|_1 \leq 1$ is now equivalent to a system of inequalities $HD\lambda \leq \mathbb{1}$, for some matrix H and conformable vector $\mathbb{1}$, as in Lemma 7.6.

This yields a Lagrangian for the unperturbed LP (with a similar form for the perturbed case):

$$L(\lambda, x, y) = b^\top \lambda + x^\top (\mathbb{1} - HD\lambda) - y^\top A^\top \lambda,$$

where $x, y \geq 0$ are Lagrange multipliers. With these modifications, we redefine the events $E_N(Q)$ as follows:[37]

$$E_N(Q) := \left\{ \mathbb{W} \,\middle|\, \begin{array}{ll} \|\hat{D} - D\| \leq \alpha_N, & d_{\mathrm{Pr}}\big(\mathcal{L}(\zeta^* \mid \mathbb{W}), \mathcal{N}(0, \Sigma(Q))\big) \leq \alpha_N, \\ \sqrt{N}\|\hat{b} - b(Q)\| \leq M_N, & \hat{b} - b(Q) \in \mathrm{Range}(\Sigma(Q)) \end{array} \right\}.$$

Steps 1 and 2 of Theorem 4.14 go through. We now turn to Step 3, which verifies the anti-concentration condition (equation 7.44 in the proof of Theorem 4.14) needed in Step 2, in the setting of the present theorem. Throughout the argument below, we use C to represent a generic constant that does not depend on $Q \in \mathcal{Q}$ or on N, for all sufficiently large N. The value of C may vary from line to line.

Step 3 We now establish equation 7.44 under Assumptions 4.2, 4.5,4.12, 4.18, and 4.20. Suppose that the event $\{\hat{T}_N > \hat{c}_N(1-\alpha)\} \cap E_N(Q)$ occurs, where $E_N(Q)$ is as defined above. Then, we necessarily have $\hat{v} > 0$, with $\hat{v}$ finite.[38]

We now show that it must be the case that the optimal value is achieved at an extreme point of $\hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^\perp$. Indeed, by the Weyl-Minkowski decomposition theorem, the polyhedron $\hat{\mathcal{D}}(Q)$ can be decomposed as the Minkowski sum of a pointed polyhedron (i.e., a polyhedron with at least one extreme point, or equivalently a polyhedron that does not contain a line), and the lineality space of $\hat{D}(Q)$ (defined as the set of all $z \in \mathbb{R}^p$ such that for all $\lambda \in \hat{\mathcal{D}}$, $\lambda + tz \in \hat{\mathcal{D}}$ for all $t \in \mathbb{R}$). In the present context, the decomposition can be explicitly given by

$$\hat{\mathcal{D}}(Q) = [\hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^\perp] + \mathcal{L}(Q),$$

where $+$ denotes the Minkowski sum, and $\mathcal{L}(Q)$ is the lineality space of $\mathcal{D}(Q)$ (and $\hat{\mathcal{D}}(Q)$) [39]given by the subspace

$$\mathcal{L}(Q) = \big\{\lambda \in \mathbb{R}^p \mid A^\top \lambda = 0,\ \Pi^u(\lambda) = 0\big\}.$$

Indeed, for $\lambda \in \mathbb{R}^p$, if we let $\Pi_{\mathcal{L}}(\lambda)$ denote the projection of λ on the subspace $\mathcal{L}(Q)$, then $\lambda - \Pi_{\mathcal{L}}(\lambda) \in \hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^\perp$ for all $\lambda \in \hat{\mathcal{D}}(Q)$, yielding the desired decomposition. We now show that since $\hat{v}$ is finite, we must have

$$\hat{v} = \max\{\hat{b}^\top \lambda \mid \lambda \in \hat{\mathcal{D}}(Q)\} = \max\{\hat{b}^\top \lambda \mid \lambda \in \hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^\perp\}.$$

This follows since $\hat{v}$ being finite implies that we must have $\hat{b}^\top z = 0$ for all $z \in \mathcal{L}(Q)$, since we must have $\hat{b}^\top(\lambda + tz) \leq \hat{v}$ for all $\lambda \in \hat{\mathcal{D}}(Q)$ and $t \in \mathbb{R}$. Hence, the set of optimal values contains the set

$$\arg\max\{\hat{b}^\top \lambda \mid \lambda \in \hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^\perp\},$$

[37]This redefinition effectively appends the event $\{\hat{b} - b(Q) \in \mathrm{Range}(\Sigma(Q))\}$ to the previously defined event $E_N(Q)$. Assumption 4.20 ensures that we continue to have $\underline{\lim}_{N\to\infty} \inf_{Q \in \mathcal{Q}} Q(E_N(Q)) = 1$.

[38]We recall here that from the discussion in the paragraph that follows equation 4.20, if $\hat{v} = \infty$, then $Q(\hat{v} = +\infty) = 1$ and we always reject the null. Thus it suffices to only consider DGPs such that $\hat{v}$ is always finite.

[39]Note that since D and $\hat{D}$ have the same nonzero diagonal elements, the conditions $\|D\lambda\|_1 = 0$, $\|\hat{D}\lambda\|_1 = 0$, and $\Pi^u(\lambda) = 0$ are equivalent.

which is a face of $\hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^{\perp}$. As $\hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^{\perp}$ is a pointed polyhedron,[40] and as such, any face of $\hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^{\perp}$ contains an extreme point, which cannot be zero, since $\hat{\nu} > 0$. In conclusion, on the event $\{\hat{T}_N > \hat{c}_N(1-\alpha)\} \cap E_N(Q)$, there must exist $\hat{\lambda} \in \text{extr}(\hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^{\perp}) \setminus \{0\}$ such that $\hat{b}^{\top}\hat{\lambda} = \hat{\nu}$. We will work with such a $\hat{\lambda}$ below.

Now since $\mathcal{L}(Q)^{\perp}$ is a subspace, it admits a polyhedral representation of the form

$$\mathcal{L}(Q)^{\perp} = \{\lambda \in \mathbb{R}^p \mid M\lambda = 0\},$$

for some matrix M, leading to a representation of $\hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^{\perp}$ given by

$$\hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^{\perp} = \{\lambda \in \mathbb{R}^p \mid A^{\top}\lambda \leq 0,\ M\lambda = 0,\ \|\hat{D}\lambda\|_1 \leq 1\}.$$

The latter representation comforms with the representation of the polytopes discussed in Lemma 7.6, and equation 7.130 implies that for all sufficiently large N, for all $Q \in \mathcal{Q}_0$, and for all $\lambda \in \text{extr}(\hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^{\perp}) \setminus \{0\}$, there exists $\lambda' \in \text{extr}(\mathcal{D}(Q) \cap \mathcal{L}(Q)^{\perp}) \setminus \{0\}$ such that

$$\lambda - \lambda' = (1 - \frac{1}{\|D\lambda\|_1})\lambda \qquad \text{and} \qquad \|\Pi^u(\lambda) - \Pi^u(\lambda')\| \leq C\alpha_N, \tag{7.65}$$

where we have used the fact that $\|\hat{D} - D\| \leq \alpha_N$ on $E_N(Q)$, and that $\sup_{\lambda \in \hat{\mathcal{D}}} \|\Pi^u(\lambda)\| \leq C$, as in equation 7.36. Thus our in sample optimal solution $\hat{\lambda}$ is the positive scalar multiple of an extreme point $\tilde{\lambda}$ of $\mathcal{D}(Q) \cap \mathcal{L}(Q)^{\perp} \setminus \{0\}$, and

$$\|\Pi^u(\hat{\lambda}) - \Pi^u(\tilde{\lambda})\| \leq C\alpha_N. \tag{7.66}$$

We claim that Assumption 4.20 implies that on the event $E_N(Q)$, we must have $\tilde{\lambda}^{\top}\Sigma(Q)\tilde{\lambda} \geq \rho$, with ρ as in Assumption 4.20. Indeed, by the definition of the event $E_N(Q)$, we have $\hat{b} - b \in \text{Range}(\Sigma(Q))$ on $E_N(Q)$. But then $\tilde{\lambda}$ cannot be an extreme point of $\mathcal{D}(Q) \cap \mathcal{L}(Q)^{\perp} \setminus \{0\}$ such that

$$\tilde{\lambda}^{\top}\Sigma(Q)\tilde{\lambda} = 0 \tag{7.67}$$

as the latter would imply that $\tilde{\lambda} \in \text{Range}(\Sigma(Q))^{\perp}$ and that $\tilde{\lambda}^{\top}\hat{b} = \tilde{\lambda}^{\top}b \leq 0$, since $b^{\top}\lambda \leq 0$ for all $\lambda \in \mathcal{D}(Q)$ (and thus $\hat{\mathcal{D}}(Q)$, by homogeneity). But the latter would contradict $\hat{b}^{\top}\tilde{\lambda} > 0$, as $\hat{\nu} = \hat{\lambda}^{\top}\hat{b} > 0$, and $\tilde{\lambda}$ is a positive scalar multiple of $\hat{\lambda}$. We thus conclude that for all large N, for $Q \in \mathcal{Q}_0$, and on the event $\{\hat{T}_N > \hat{c}_N(1-\alpha)\} \cap E_N(Q)$, there must exist a $\hat{\lambda} \in \text{extr}(\hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^{\perp}) \setminus \{0\}$ and $\tilde{\lambda} \in \text{extr}(\mathcal{D}(Q) \cap \mathcal{L}(Q)^{\perp}) \setminus \{0\}$, such that equations 7.66 and 7.67 hold. As in equation 7.49, we then have

$$\begin{aligned}
\hat{\lambda}^{\top}\Sigma(Q)\hat{\lambda} &= \Pi^u(\hat{\lambda})^{\top}\Sigma(Q)\Pi^u(\hat{\lambda}) = \|\Sigma(Q)^{1/2}\Pi^u(\hat{\lambda})\|^2 \\
&\geq (1/2)\|\Sigma(Q)^{1/2}\Pi^u(\tilde{\lambda})\|^2 - \|\Sigma(Q)^{1/2}(\Pi^u(\hat{\lambda}) - \Pi^u(\tilde{\lambda}))\|^2 \\
&= (1/2)\tilde{\lambda}^{\top}\Sigma(Q)\tilde{\lambda} - \|\Sigma(Q)^{1/2}(\Pi^u(\hat{\lambda}) - \Pi^u(\tilde{\lambda}))\|^2 \\
&\geq (1/2)\tilde{\lambda}^{\top}\Sigma(Q)\tilde{\lambda} - C\alpha_N^2.
\end{aligned}$$

[40] The set $\hat{\mathcal{D}}(Q) \cap \mathcal{L}(Q)^{\perp}$ is a pointed polyhedron, since any vector in its lineality is necessarily in the lineality of $\hat{\mathcal{D}}$, and thus in $\mathcal{L}(Q) \cap \mathcal{L}(Q)^{\perp} = \{0\}$

Since $\hat{\lambda}$ is in $\hat{\Delta}$, the latter inequality implies that for all sufficiently large N, for $Q \in \mathcal{Q}_0$, and on the event $\{\hat{T}_N > \hat{c}_N(1-\alpha)\} \cap E_N(Q)$, the set $\hat{\Delta}$ contains an element $\hat{\lambda}$ that satisfies

$$\hat{\lambda}^\top \Sigma(Q)\hat{\lambda} \geq \rho/4. \tag{7.68}$$

The anti-concentration condition 7.44 (in the present context) then follows from Proposition 7.2, using the fact that the sets $\Pi^u(\hat{\Delta})$ are eventually uniformly bounded on the events $E_N(Q)$, and

$$\sup_{\lambda \in \hat{\Delta}} \langle \zeta, \lambda \rangle = \sup_{\lambda \in \hat{\Delta}} \langle \Pi^u(\zeta), \Pi^u(\lambda) \rangle,$$

where $\zeta \sim \mathcal{N}(0, \Sigma(Q))$.

□

Proof of Theorem 4.29. In this section, we provide proofs of equations (4.31) and (4.32). The dual to the LP 4.25 is given by

$$\min\{t \mid t\Omega(Q)\mathbb{1} + A(Q)\eta \geq b(Q),\ t \geq 0\}. \tag{7.69}$$

Let the Lagrangian associated with the LP 4.25 be given by

$$L(\lambda, \eta, t) = b^\top \lambda - \eta^\top A^\top \lambda + t\big(1 - \mathbb{1}^\top \Omega\lambda\big).$$

where $t \geq 0$ and η are the Lagrange multipliers.

Let $\xi = (\xi_A, \xi_b, \xi_\Omega)$ denote generic pertubations of the input parameters $\mu = (A, b, \Omega)$ of our LPs, where $\xi_A \in \mathbb{R}^{p\times d}$, $\xi_b \in \mathbb{R}^p$, and $\xi_\Omega \in \mathbb{R}^{p\times p}$ is a diagonal matrix. Given ξ, let the resulting perturbed inputs $\mu_N(\xi) = (A_N(\xi), b_N(\xi), \Omega_N(\xi))$ be defined by

$$\mu_N(\xi) = \Big(A + \frac{\xi_A}{\sqrt{N}}, b + \frac{\xi_b}{\sqrt{N}}, \Omega + \frac{\xi_\Omega}{\sqrt{N}}\Big).$$

We assume here that the pertubation matrices ξ_Ω, like the matrices Ω, have all their (i,i)-th diagonal entries that correspond to indices $i \in K$ equal to zero.[41] Let $\phi_N(\mu)$ denote the value of the LP 4.25, for arbitrary inputs $\mu = (A, b, \Omega)$. Let $L_N(\lambda, \eta, t; \xi)$ denote the Lagrangian for the LP 4.25 when the inputs of the LP are $\mu_N(\xi)$, and given by

$$L_N(\lambda, \eta, t; \xi) = b_N(\xi)^\top \lambda - \eta^\top A_N(\xi)^\top \lambda + t\big(1 - \mathbb{1}^\top \Omega_N(\xi)\lambda\big).$$

Let $\Delta_N(0, N)(\xi)$ and $S_{0,N}(\xi)$ denote respectively the set of optimal solutions for the LP 4.25 and its dual, when the inputs are given by $\mu_N(\xi)$. Let Δ_0 and S_0 denote the analoguous objects when $\xi = 0$.

For $(\lambda_\xi, \eta_\xi, t_\xi) \in \Delta_{0,N}(\xi) \times S_{0,N}(\xi)$ and $(\lambda_0, \eta_0, t_0) \in \Delta_0 \times S_0$, the saddle point property of the Lagrangian

[41]See the paragraph that precedes Assumption 4.22, for a definition of the set K, U, and for a definition of the projection operator Π^u.

implies that, for all sufficiently small perturbations ξ such that $\phi_N(\mu_N(\xi))$ is finite, we have

$$\begin{aligned}\sqrt{N}\big(\phi_N(\mu_N(\xi))-\phi(\mu)\big) &= \sqrt{N}\big(L_N(\lambda_\xi,\eta_\xi,t_\xi;\xi)-L(\lambda_0,\eta_0,t_0)\big)\\ &\leq \sqrt{N}\big(L_N(\lambda_\xi,\eta_0,t_0;\xi)-L(\lambda_\xi,\eta_0,t_0)\big)\\ &= \xi_b^\top\lambda_\xi-\eta_0^\top\xi_A^\top\lambda_\xi-t_0\mathbb{1}^\top\xi_\Omega\lambda_\xi.\end{aligned}$$

Now, for any $Q\in\mathcal{Q}_0$, $\phi(\mu)=0$ and the condition $(\eta_0,t_0)\in S_0$ implies $t_0=0$. In fact, $S_0=\{(t,\eta)\mid t=0,\ b(Q)-A(Q)\eta\leq 0\}$. Therefore, the inequality above yields, for all sufficiently large N, for all sufficiently small perturbations ξ[42], and for $Q\in\mathcal{Q}_0$:

$$\sqrt{N}\phi_N(\mu_N(\xi))\leq \xi_b^\top\lambda_\xi-\eta_0^\top\xi_A^\top\lambda_\xi=\langle\xi_b-\xi_A\eta_0,\lambda_\xi\rangle. \tag{7.70}$$

Assumptions 4.22 through 4.25 imply the existence of $\alpha_N\downarrow 0$, such that

$$\sup_{Q\in\mathcal{Q}} d_{\Pr}\Big(\sqrt{N}(\hat{d}-d(Q)),\mathcal{N}(0,\Sigma(Q))\Big)\leq\alpha_N,$$

and such that for any $M_N\uparrow\infty$, and for all $Q\in\mathcal{Q}_0$, the event $E_N(Q)$ defined by[43]

$$E_N(Q):=\left\{\mathbb{W}\;\middle|\;\begin{array}{ll}\|\hat{\Omega}-\Omega(Q)\|\leq\alpha_N, & d_{\Pr}\big(\mathcal{L}(\zeta^*\mid\mathbb{W}),\mathcal{N}(0,\Sigma(Q))\big)\leq\alpha_N,\\ \sqrt{N}\|\hat{b}-b(Q)\|\leq M_N, & \sqrt{N}\|\hat{A}-A(Q)\|\leq M_N\end{array}\right\}.$$

satisfies

$$\underline{\lim}_{N\to\infty}\inf_{Q\in\mathcal{Q}} Q(E_N(Q))=1.$$

We choose $M_N\uparrow\infty$ such that $M_N=o(\kappa_N)$, $M_N\alpha_N=o(1)$, and $M_N^2\Big(\kappa_N/\sqrt{N}\Big)=o(1)$. Such a choice is possible since $\kappa_N=o(\sqrt{N})$, where κ_N, here and in what follows, denotes either κ_{1N}, κ_{2N}, or κ_{3N}.

Let $\mathcal{X}_N$ denote the set of pertubation ξ defined by

$$\mathcal{X}_N=\Big\{\xi\mid\|\xi_b\|\leq M_N,\ \|\xi_A\|\leq M_N,\ \|\xi_\Omega\|\leq\sqrt{N}\alpha_N\Big\}.$$

We now proceed to derive inequalities (4.31) and (4.32) from inequality (7.70). The proof is structured in four steps, which mirror the steps used in the derivation of the main results in Sections 4.1 through 4.3. In particular, inequality (4.31) is established in Steps 1 through 3, while Step 4 contains the proof of inequality (4.32). Throughout the argument, we let C denote a generic constant that is uniform over all data-generating processes $Q\in\mathcal{Q}_0$, for sufficiently large N. The constant C may vary from one equation to another.

Step 1 In this step, we show that there exists a sequence $\beta_N=o(1)$ such that, for all sufficiently large

[42] For small perturbations ξ such that Ω_ξ has strictly positive diagonal entries for all $i\in U$, it follows from the discussion preceding Assumption 4.28 that, for all $Q\in\mathcal{Q}_0$, the function $\phi(\mu_N(\xi))$ remains finite.

[43] Since assumption 4.22 implies that the number of estimated components of b and A are uniformly bounded w.r.t. N, the uniform consistency of the estimator $\hat{D}$ of the standard deviation matrices implies the uniform consistency of the normalizing matrices $\hat{\Omega}$.

N and for all $Q \in \mathcal{Q}_0$, the following inequality holds on the event $E_N(Q)$:

$$\sup_{\xi \in \mathcal{X}_N} \sup_{\lambda \in \Delta_{0,N}(\xi)} \left\{ \langle \xi_b - \xi_A \eta_0, \lambda \rangle - \sup_{\lambda \in \hat{\Delta}} \langle \xi_b - \xi_A \hat{\eta}, \lambda \rangle \right\} \leq \beta_N, \tag{7.71}$$

where $\hat{\Delta}$ is defined in equation 4.29, and $\hat{\eta}$ is as given in equation 4.30.

Let $\mathcal{D}(Q)$ and $\hat{\mathcal{D}}$ denote respectively the feasible region of the LPs 4.25 and 4.26. Note that on the event $E_N(Q)$, the condition $\|\hat{D} - D\| \leq \alpha_N$ implies $\|\hat{\Omega} - \Omega\| \leq \alpha_N$. Consequently, for all sufficiently large N such that $\alpha_N \leq \underline{\sigma}/2$, we have $[\hat{\Omega}]_{ii} > \underline{\sigma}/2$ for all $i \in U$. The same reasoning shows that for all such N and for all $\xi \in \mathcal{X}_N$, we also have $[\Omega_N(\xi)]_{ii} \geq \underline{\sigma}/2$ for all $i \in U$. As a consequence, since $\|\Omega_N(\xi)\lambda\|_1 \leq 1$ for all $\lambda \in \Delta_{0,N}(\xi)$, and $\|\hat{\Omega}\lambda\|_1 \leq 1$ for all $\lambda \in \hat{\mathcal{D}}$, it follows that for all N such that $\alpha_N \leq \underline{\sigma}/2$, we have

$$\sup_{\xi \in \mathcal{X}_N} \sup_{\lambda \in \Delta_{0,N}(\xi)} \|\Pi^u(\lambda)\|_1 \leq \frac{2}{\underline{\sigma}}, \qquad \text{and} \qquad \sup_{\lambda \in \hat{\mathcal{D}}} \|\Pi^u(\lambda)\|_1 \leq 2/\underline{\sigma} \quad \text{on the event } E_N(Q). \tag{7.72}$$

In the derivations that follow, we assume that N is sufficiently large that $\alpha_N \leq \underline{\sigma}/2$.

As η_0 in equation 7.70 is any solution to $b(Q) - A(Q)\eta \leq 0$, we can assume that $\eta_0 = \eta_Q$, where η_Q is as in Assumption 4.27. We first show that on the events $E_N(Q)$, for all sufficiently large N and for $Q \in \mathcal{Q}_0$, we can replace η_Q by $\hat{\eta}$, and we have

$$\sup_{\xi \in \mathcal{X}_N} \sup_{\lambda \in \Delta_{0,N}(\xi)} \{\langle \xi_b - \xi_A \eta_Q, \lambda \rangle - \langle \xi_b - \xi_A \hat{\eta}, \lambda \rangle\} \leq C M_N \left(\kappa_N/\sqrt{N}\right)^{1/2}, \tag{7.73}$$

We first show that on the event $E_N(Q)$, for all sufficiently large N and for $Q \in \mathcal{Q}_0$, $\eta(Q)$ is in the feasible region of the quadratic program 4.30, and satisfies

$$\hat{b} - \hat{A}\eta(Q) \leq \frac{\kappa_N}{\sqrt{N}} \hat{\Omega} \mathbb{1}. \tag{7.74}$$

Indeed, we have

$$\begin{aligned} \hat{b} - \hat{A}\eta_Q &\leq \hat{b} - \hat{A}\eta_Q + A\eta(Q) - b \\ &\leq \left(\|\hat{b} - b\| + \|\hat{A} - A\|\|\eta_Q\|\right)\Pi^u(\mathbb{1}). \end{aligned}$$

Since Assumption 4.27 implies that η_Q is uniformly (in $Q \in \mathcal{Q}_0$) bounded, we conclude that on the event $E_N(Q)$, we have

$$\hat{b} - \hat{A}\eta(Q) \leq \frac{CM_N}{\sqrt{N}} \Pi^u(\mathbb{1}),$$

which yields equation 7.74 for all sufficiently large N as $M_N = o(\kappa_N)$, and $[\hat{\Omega}]_{ii} \geq \underline{\sigma}/2$ for all $i \in U$. In turn, this implies that for all sufficiently large N and all $Q \in \mathcal{Q}_0$, we have

$$\|\hat{\eta}\| \leq \|\eta_Q\| \leq M \tag{7.75}$$

where $M > 0$ is such that $\sup_{Q \in \mathcal{Q}_0} \|\eta_Q\| \leq M$.

By Lemma 7.8, if $\mathcal{C}(Q) := \{\eta \mid b(Q) - A(Q)\eta \leq 0\}$, then inequality 7.75 and Assumption 4.27 imply that

$$d\big(\hat{\eta}, \mathcal{C}_Q\big) \leq C\|[b(Q) - A(Q)\hat{\eta}]_+\| \leq C\kappa_N/\sqrt{N}. \tag{7.76}$$

Indeed, we have

$$\begin{aligned}
[b - A\hat{\eta}]_+ &\leq \Big[b - A\hat{\eta} + \hat{A}\hat{\eta} - \hat{b} + (\hat{T}_N + \frac{\kappa_N}{\sqrt{N}})\hat{\Omega}\mathbb{1}\Big]_+ \\
&\leq [b - \hat{b}]_+ + [(\hat{A} - A)\hat{\eta}]_+ + C\frac{\kappa_N}{\sqrt{N}}\hat{\Omega}\mathbb{1} \\
&\leq C\frac{\kappa_N}{\sqrt{N}}\Pi^u(\mathbb{1}),
\end{aligned}$$

where we have used the fact that $\hat{\eta}$ is in the feasible region of the quadratic program 4.30, equation 7.75, the fact that $M_N = o(\kappa_N)$, and the additional fact, which we derive further below, that $T_N = o(\kappa_N/\sqrt{N})$.

Inequalities 7.75 and 7.76 imply that

$$\|\hat{\eta} - \eta(Q)\| \leq C\Big(\kappa_N/\sqrt{N}\Big)^{1/2}. \tag{7.77}$$

Indeed, if $\Pi_{\mathcal{C}}(\cdot)$ denotes the (Euclidean) projection operator on $\mathcal{C}(Q)$, then equations 7.75 and 7.76 imply that

$$\langle \eta(Q), \eta\rangle \geq \|\eta(Q)\|^2 \quad \forall \eta \in \mathcal{C}(Q) \qquad \text{and} \qquad \|\Pi_{\mathcal{C}}(\hat{\eta})\| \leq \|\eta(Q)\| + C\frac{\kappa_N}{\sqrt{N}}.$$

This gives

$$\|\eta(Q) - \Pi_{\mathcal{C}}(\hat{\eta})\|^2 = \|\eta(Q)\|^2 + \|\Pi_{\mathcal{C}}(\hat{\eta})\|^2 - 2\langle \eta(Q), \Pi_{\mathcal{C}}(\hat{\eta})\rangle \leq C\frac{\kappa_N}{\sqrt{N}},$$

and equation 7.77 then follows from the latter inequality, and equation 7.76, noting that $\|\hat{\eta} - \eta(Q)\| \leq \|\hat{\eta} - \Pi_{\mathcal{C}}(\hat{\eta})\| + \|\Pi_{\mathcal{C}}(\hat{\eta}) - \eta(Q)\|$.

Equation 7.72 implies that $\|\xi_A^\top \lambda\| = \|\xi_A^\top \Pi^u(\lambda)\| \leq CM_N$ for $\xi \in \mathcal{X}_N$ and $\lambda \in \Delta_{0,N}(\xi)$, and equation 7.77 then implies that on the events $E_N(Q)$, for all sufficiently large N and for $Q \in \mathcal{Q}_0$, we have

$$\sup_{\xi \in \mathcal{X}_N} \sup_{\lambda \in \Delta_{0,N}(\xi)} \Big|\{\langle \xi_b - \xi_A \eta_0, \lambda\rangle - \langle \xi_b - \xi_A \hat{\eta}, \lambda\rangle\}\Big| \leq CM_N\Big(\kappa_N/\sqrt{N}\Big)^{1/2}.$$

Note also that, since M_N is chosen such that $M_N^2(\kappa_N/\sqrt{N}) = o(1)$, the right-hand side of the latter inequality is asymptotically negligible. This yields equation 7.73.

We now show that for all sufficiently large N, and for all $Q \in \mathcal{Q}_0$, the following holds on the event $E_N(Q)$:

$$\sup_{\xi \in \mathcal{X}_N} \vec{d}_H\big(\Pi^u(\Delta_{0,N}(\xi)), \Pi^u(\hat{\Delta})\big) \leq C\alpha_N. \tag{7.78}$$

Assuming that inequality 7.78 holds, and invoking equation 7.72, it follows that on the event $E_N(Q)$ we have

$$\sup_{\xi \in \mathcal{X}_N} \sup_{\lambda \in \Delta_{0,N}(\xi)} \Big\{\langle \xi_b - \xi_A \hat{\eta}, \lambda\rangle - \sup_{\lambda \in \hat{\Delta}} \langle \xi_b - \xi_A \hat{\eta}, \lambda\rangle\Big\} \leq C\alpha_N M_N. \tag{7.79}$$

Combining inequalities 7.73 and 7.79 yields inequality 7.71 with

$$\beta_N = CM_N \left(\frac{\kappa_N}{\sqrt{N}}\right)^{1/2} + C\alpha_N M_N = o(1).$$

It therefore remains to establish inequality 7.78 in order to complete Step 1.

We first show that on the event $E_N(Q)$, for all sufficiently large N and for $Q \in \mathcal{Q}_0$, we have

$$\forall \xi \in \mathcal{X}_N, \quad \Delta_{0,N}(\xi) \subseteq \left\{\lambda \in \mathbb{R}^p_+ \mid -\frac{\kappa_N}{\sqrt{N}}\mathbb{1} \leq \hat{A}^\top\lambda \leq \frac{\kappa_N}{\sqrt{N}}\mathbb{1}\right\}. \tag{7.80}$$

Indeed, on $E_N(Q)$ we have

$$\sup_{\lambda\in\Delta_{0,N}(\xi)} \|\hat{A}^\top\lambda\|_\infty = \sup_{\lambda\in\Delta_{0,N}(\xi)} \|(\hat{A} - A_N(\xi))^\top\Pi^u(\lambda)\|_\infty \leq CM_N/\sqrt{N},$$

where we have use equation 7.72. Equation 7.80 then follows since $M_N = o(\kappa_N)$. Arguing as in 7.38, we easily show that on $E_N(Q)$ and for all N sufficiently large that $\alpha_N < \underline{\sigma}/4$, we have

$$(1 - 4\alpha_N/\underline{\sigma})\|\hat{\Omega}\lambda\|_1 \leq \|\Omega_N(\xi)\lambda\|_1 \leq (1 + 4\alpha_N/\underline{\sigma})\|\hat{\Omega}\lambda\|_1,$$

for all $\lambda \in \mathbb{R}^p$. Hence if $\hat{\Gamma}(\lambda) := (1 - 4\alpha_N/\underline{\sigma})\lambda$, then the latter inequality and equation 7.80 imply that on the event $E_N(Q)$, for all sufficiently large N and for $Q \in \mathcal{Q}_0$, we have

$$\forall \xi \in \mathcal{X}_N, \quad \hat{\Gamma}(\Delta_{0,N}(\xi)) \subseteq \left\{\lambda \in \mathbb{R}^p_+ \mid -\frac{\kappa_N}{\sqrt{N}}\mathbb{1} \leq \hat{A}^\top\lambda \leq \frac{\kappa_N}{\sqrt{N}}\mathbb{1},\ \|\hat{\Omega}(\lambda)\|_1 \leq 1\right\}.$$

Moreover, equation 7.72 implies that

$$\forall \xi \in \mathcal{X}_N, \quad \sup_{\lambda\in\Delta_{0,N}(\xi)} \|\Pi^u(\lambda - \hat{\Gamma}(\lambda))\| \leq C\alpha_N.$$

We now show that for all sufficiently large N, we have

$$\inf_{\lambda\in\Delta_{0,N}(\xi)} \hat{b}^\top\hat{\Gamma}(\lambda) \geq \hat{T}_N - \kappa_N/\sqrt{N},$$

which when combined with the preceding two inequalities yields 7.78, and completes the proof of Step 1. We first show that $\hat{T}_N = o(\kappa_N/\sqrt{N})$. Indeed, first note that Assumption 4.27 implies that there exists a constant $M > 0$ such that for all $Q \in \mathcal{Q}_0$, we have

$$b^\top\lambda \leq \langle\lambda, A\eta_Q\rangle = \langle A^\top\lambda, \eta_Q\rangle \leq M\|A^\top\lambda\|, \quad \forall\lambda \geq 0. \tag{7.81}$$

On the event $E_N(Q)$, for all sufficiently large N and for $Q \in \mathcal{Q}_0$, we have

$$\begin{aligned}
0 \le \hat{T}_N = \sup_{\lambda \in \hat{\mathcal{D}}} \langle \hat{b}, \lambda \rangle &\le \sup_{\lambda \in \hat{\mathcal{D}}} \langle \hat{b} - b, \lambda \rangle + \sup_{\lambda \in \hat{\mathcal{D}}} \langle b, \lambda \rangle \\
&\le \sup_{\lambda \in \hat{\mathcal{D}}} \langle \hat{b} - b, \Pi^u(\lambda) \rangle + \sup_{\lambda \in \hat{\mathcal{D}}} M \|A^\top \lambda\| \\
&= \sup_{\lambda \in \hat{\mathcal{D}}} \langle \hat{b} - b, \Pi^u(\lambda) \rangle + \sup_{\lambda \in \hat{\mathcal{D}}} M \|(A - \hat{A})^\top \Pi^u(\lambda)\| \\
&\le C M_N / \sqrt{N} = o\big(\kappa_N / \sqrt{N}\big),
\end{aligned} \tag{7.82}$$

where we have used equations 7.72 and 7.81, and the fact that $M_N = o(\kappa_N)$.

Now on $E_N(Q)$ and for $\xi \in \mathcal{X}_N$, we have

$$\begin{aligned}
\inf_{\lambda \in \Delta_{0,N}(\xi)} \hat{b}^\top \hat{\Gamma}(\lambda) &\ge \inf_{\lambda \in \Delta_{0,N}(\xi)} \langle \hat{b} - b_N(\xi), \Pi^u(\hat{\Gamma}(\lambda)) \rangle + \inf_{\lambda \in \Delta_{0,N}(\xi)} \langle b_N(\xi), \hat{\Gamma}(\lambda) \rangle \\
&\ge \inf_{\lambda \in \Delta_{0,N}(\xi)} \langle \hat{b} - b_N(\xi), \Pi^u(\hat{\Gamma}(\lambda)) \rangle \\
&\ge -C M_N / \sqrt{N}
\end{aligned} \tag{7.83}$$

where we have used the fact that $\langle b_N(\xi), \hat{\Gamma}(\lambda) \rangle \ge 0$ for all $\lambda \in \Delta_{0,N}(\xi)$, as $\hat{\Gamma}(\lambda)$ is a positive scalar multiple of λ and $\inf_{\lambda \in \Delta_{0,N}(\xi)} \langle b_N(\xi), \lambda \rangle \ge 0$ (since 0 is always feasible). Combining equations 7.82 and 7.83, and the fact that $M_N = o(\kappa_N)$, implies that on $E_N(Q)$ and for all sufficiently large N we have

$$\inf_{\lambda \in \Delta_{0,N}(\xi)} \hat{b}^\top \hat{\Gamma}(\lambda) \ge \hat{T}_N - \kappa_N / \sqrt{N},$$

thus completing the proof of Step 1.

Step 2 In this step, we use equation (7.71) and a coupling argument to upper bound the test statistic by another statistic whose distribution can be uniformly estimated for all $Q \in \mathcal{Q}_0$. The proof closely parallels Step 2 in the proof of Theorems 4.7 through 4.21. Let $\hat{\xi}^* = (\sqrt{N}(\hat{A}^* - A(Q)), \sqrt{N}(\hat{b}^* - b(Q)), \sqrt{N}(\hat{\Omega}^* - \Omega(Q)))$ be an independent and identically distributed version of the root $\hat{\xi} = (\sqrt{N}(\hat{A} - A(Q)), \sqrt{N}(\hat{b} - b(Q)), \sqrt{N}(\hat{\Omega} - \Omega(Q)))$, computed from a sample $\mathbb{W}_N^*$, such that $\mathbb{W}_N^*$ and $\mathbb{W}_N$ are independent and identically distributed. For $Q \in \mathcal{Q}$, let the event $F_N^* = F_N^*(Q)$ be defined by

$$F_N^*(Q) := \left\{ \mathbb{W}^* \;\middle|\; \|\hat{\xi}_A^*\| > M_N \text{ or } \|\hat{\xi}_b^*\| > M_N \text{ or } \|\hat{\xi}_\Omega^*\| > \sqrt{N} \alpha_N \right\}.$$

By our choice of M_N and α_N, we have

$$\delta_{1N} := \sup_{Q \in \mathcal{Q}} Q\big(F_N^*(Q)\big) \quad \text{satisfies} \quad \lim_{N \to \infty} \delta_{1N} = 0$$

Equations 7.70 and 7.71 then imply that for all sufficiently large N and on the event $E_N(Q)$, for $Q \in \mathcal{Q}_0$, we have

$$\sqrt{N} \phi(\mu_N(\hat{\xi}^*)) \le \sup_{\lambda \in \hat{\Delta}} \langle \hat{\xi}_b^* - \hat{\xi}_A^* \hat{\eta}, \lambda \rangle + \beta_N + \infty \cdot \mathbb{1}_{\{F_N^*\}}. \tag{7.84}$$

Let $J_N(\cdot;Q)$ denote the CDF of $\sqrt{N}\hat{T}_N$. Define

$$\hat{G}_N(\cdot;Q) := \mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\hat{\xi}_b^* - \hat{\xi}_A^*\hat{\eta},\lambda\rangle + \beta_N + \infty\cdot\mathbb{1}_{\{F_N^*\}}\,\middle|\,\mathbb{W}_N\right)$$

as the conditional (on $\mathbb{W}_N$) CDF of the right-hand side of inequality (7.84), viewed as a random element taking values in $\mathbb{R}\cup\{\infty\}$. Similarly, define

$$\hat{G}_\infty(\cdot;Q) := \mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\zeta_b - \zeta_A\eta_Q,\lambda\rangle\,\middle|\,\mathbb{W}_N\right),$$

where $\zeta\sim N(0,\Sigma(Q))$ (as in Assumption 4.23).[44] And let $\hat{H}_N(;Q)$ be defined as in Theorem 4.29

$$\hat{H}_N(\cdot;Q) := \mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\zeta_b^* - \zeta_A^*\hat{\eta},\lambda\rangle\,\middle|\,\mathbb{W}_N\right),$$

where ζ^* is the bootstrap estimate of of ζ that is given in Assumption 4.25. Note that Assumptions 4.22 and 4.24 (part (i)) imply that the random variables $\left\{\|\zeta\|\mid\zeta\sim\mathcal{N}(0,\Sigma(Q)),\ Q\in\mathcal{Q}_0\right\}$ are tight. As a consequence, Strassen's coupling, and equations 7.72 and 7.77, imply that the sequence

$$\delta_{2N} = \sup_{Q\in\mathcal{Q}_0}\sup_{\mathbb{W}_N\in E_N(Q)} d_{Pr}\left(\mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\zeta_b - \zeta_A\eta_Q,\lambda\rangle\,\middle|\,\mathbb{W}_N\right),\mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\hat{\xi}_b^* - \hat{\xi}_A^*\hat{\eta},\lambda\rangle\,\middle|\,\mathbb{W}_N\right)\right)$$

satisfies

$$\lim_{N\to\infty}\delta_{2N} = 0.$$

The preceding equation, together with Strassen's theorem, implies that on the event $E_N(Q)$, we have

$$\begin{aligned}
d_{Pr}\left(\hat{G}_N,\hat{G}_\infty\right) &\le d_{Pr}\left(\hat{G}_N,\mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\hat{\xi}_b^* - \hat{\xi}_A^*\hat{\eta},\lambda\rangle\,\middle|\,\mathbb{W}_N\right)\right)\\
&\quad + d_{Pr}\left(\hat{G}_\infty,\mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\hat{\xi}_b^* - \hat{\xi}_A^*\hat{\eta},\lambda\rangle\,\middle|\,\mathbb{W}_N\right)\right)\\
&\le \beta_N + Q(F_N^*) + \delta_{2N}.
\end{aligned}$$

Similarly, by Strassen's theorem and equations 7.72 and 7.77, it follows that on the event $E_N(Q)$, we have

$$\lim_{N\to\infty}\delta_{3N} = 0,$$

where the sequence δ_{3N} is defined by

$$\delta_{3N} = \sup_{Q\in\mathcal{Q}_0}\sup_{\mathbb{W}_N\in E_N(Q)} d_{Pr}\left(\mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\zeta_b - \zeta_A\eta_Q,\lambda\rangle\,\middle|\,\mathbb{W}_N\right),\mathcal{L}\left(\sup_{\lambda\in\hat{\Delta}}\langle\zeta_b^* - \zeta_A^*\hat{\eta},\lambda\rangle\,\middle|\,\mathbb{W}_N\right)\right).$$

[44] Here, ζ_A denotes the matrix formed by extracting the components corresponding to A from the random vector ζ, arranged in a manner compatible with the dimensions of A. ζ_b is defined analogously.

We have

$$d_{\mathrm{Pr}}\left(\hat{G}_\infty, \hat{H}_N\right) \leq \delta_{3N}$$

on $E_N(Q)$.

In conclusion, there exists $\gamma_N = o(1)$ such that for all sufficiently large N and all $Q \in \mathcal{Q}_0$, on the event $E_N(Q)$ we have

$$d_{\mathrm{Pr}}(\hat{G}_N, \hat{G}_\infty) \vee d_{\mathrm{Pr}}(\hat{H}_N, \hat{G}_\infty) \leq \gamma_N. \tag{7.85}$$

Let the significance level α and the critical value $\hat{c}_N(1-\alpha)$ be as in Theorem 4.29. For now, assume that there exists a constant C such that for all sufficiently large N, for all $Q \in \mathcal{Q}_0$, and for all sample realizations $\mathbb{W}_N$ on the event $E_N(Q) \cap \{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}$, we have the anti-concentration condition

$$\hat{G}_\infty(\hat{c}_N(1-\alpha) + \gamma_N) - \hat{G}_\infty(\hat{c}_N(1-\alpha) - \gamma_N) \leq C\gamma_N. \tag{7.86}$$

We prove in the next step that inequality (7.86) indeed holds. For $\mathbb{W}_N \in E_N(Q) \cap \{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}$, and $Q \in \mathcal{Q}_0$, arguing as in 7.24, we have

$$J_N(\hat{c}_N(1-\alpha)) \geq 1 - \alpha - C\gamma_N. \tag{7.87}$$

Let $q_N(\cdot)$ denote the quantiles of $J_N(\cdot)$. Inequality 7.87 then implies that on the event $E_N(Q) \cap \{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}$, with $Q \in \mathcal{Q}_0$, we have

$$q_N(1 - \alpha - C\gamma_N) \leq \hat{c}_N(1-\alpha). \tag{7.88}$$

Arguing as in equation 7.26, we have

$$\sup_{Q \in \mathcal{Q}_0} Q\Big(\{\sqrt{N}\hat{T}_N > \hat{c}_N(1-\alpha)\}\Big) \leq \alpha + C\gamma_N + \delta_{1N}. \tag{7.89}$$

which yields equation 4.31 from Theorem 4.29, as $\gamma_N = o(1)$ and $\delta_{1N} = o(1)$.

Step 3 In this step, we establish inequality 7.86, which is required in the derivation of equation 7.89. We begin by noting that for all sufficiently large N and for all $Q \in \mathcal{Q}_0$, equation 7.82 implies that $0 \in \hat{\Delta}$, as $\hat{T}_N = o(\sqrt{N}/N)$. Hence for all such sufficiently large N, the distributions $\hat{H}(\cdot)$ are supported on $\mathbb{R}_+$, and rejection occurs, i.e. $\{\hat{T}_N > \hat{c}_N(1-\alpha)\}$, only if $\hat{T}_N > 0$.

We first consider the case where $\Delta_0(Q) \neq \{0\}$, with

$$\Delta_0(Q) := \arg\max\left\{b(Q)^\top\lambda \;\middle|\; \lambda \geq 0,\ A(Q)^\top\lambda = 0,\ \|\Omega(Q)\lambda\|_1 \leq 1\right\}.$$

Equation 7.78, with $\xi = 0$, then implies that on the event $E_N(Q)$, we have

$$\vec{d}_H(\Pi^u(\Delta_0(Q)), \Pi^u(\hat{\Delta})) \leq C\alpha_N.$$

By Assumption 4.28, there exists $\lambda_Q \in \Delta_0(Q)$ such that

$$\lambda_Q^\top V(Q)\lambda_Q \geq \rho.$$

This, combined with inequality the preceding inequality, implies that there exists $\hat{\lambda} \in \hat{\Delta}$ such that

$$\left\|\Pi^u(\hat{\lambda}) - \Pi^u(\lambda_Q)\right\| \leq C\alpha_N.$$

Hence, for all sufficiently large N, we have

$$\sup_{\lambda \in \hat{\Delta}} \lambda^\top V(Q)\lambda = \sup_{\lambda \in \hat{\Delta}} \Pi^u(\lambda)^\top V(Q)\Pi^u(\lambda) \geq \Pi^u(\hat{\lambda})^\top V(Q)\Pi^u(\hat{\lambda}) \geq \rho/2. \tag{7.90}$$

We now consider the case where $\Delta_0(Q) = \{0\}$. Since for all sufficiently large N, rejection implies that $\hat{T}_N > 0$, with $\hat{T}_N$ finite, there must exist $\hat{\lambda} \in \text{extr}(\hat{D}) \setminus \{0\}$ such that $\|\hat{\Omega}\hat{\lambda}\|_1 = 1$, and

$$\hat{\lambda} \in \arg\max\left\{\hat{b}^\top\lambda \;\middle|\; \lambda \geq 0,\; \hat{A}^\top\lambda = 0,\; \|\hat{\Omega}\lambda\|_1 \leq 1\right\}.$$

Note that we have $\hat{T}_N = \hat{b}^\top\hat{\lambda}$, and we clearly have that $\hat{\lambda} \in \hat{\Delta}$. By Assumption 4.28, for all $\lambda \geq 0$ such that $\|\Pi^u(\lambda)\|_1 \geq 1$ we have

$$\lambda^\top V(Q)\lambda \geq \rho. \tag{7.91}$$

For all sufficiently large N, we eventually have $\alpha_N \leq \overline{\sigma}$. Therefore, on the event $E_N(Q)$, it follows that $\|\hat{\Omega}\| \leq 2\overline{\sigma}$, which implies

$$1 = \|\hat{\Omega}\hat{\lambda}\|_1 \leq 2\overline{\sigma}\|\Pi^u(\hat{\lambda})\|_1.$$

Using equation 7.91, we then obtain

$$\hat{\lambda}^\top V(Q)\hat{\lambda} \geq \rho\left(\frac{1}{2\overline{\sigma}}\right)^2. \tag{7.92}$$

In conclusion, equations 7.90 and 7.92 together imply that for all sufficiently large N, and for all $Q \in \mathcal{Q}_0$, whenever the event $E_N(Q) \cap \{\hat{T}_N > \hat{c}_N(1-\alpha)\}$ occurs, there exists some $\lambda \in \hat{\Delta}$ such that

$$\lambda^\top V(Q)\lambda \geq \rho\left(\frac{1}{2\overline{\sigma}}\right)^2 \wedge \frac{\rho}{2}.$$

Inequality 7.86 then follows from Proposition 7.2, by applying the above inequality together with the identity

$$\langle\zeta_b - \zeta_A\eta_Q, \lambda\rangle = \langle\Pi^u(\zeta_b - \zeta_A\eta_Q), \Pi^u(\lambda)\rangle,$$

and equation 7.72.

Step 4 In this step, we establish equation (4.32) from Theorem 4.29. Fix $Q \in \mathcal{Q}_0$, and suppose that η_Q is the unique solution to the inequality system

$$b(Q) - A(Q)\eta \leq 0.$$

Lemma 7.4 implies that

$$\sqrt{N}\hat{T}_N \xrightarrow{d} \max_{\lambda\in\Delta_0}\langle \zeta_b - \zeta_A\eta_Q, \Pi^u(\lambda)\rangle,$$

where Δ_0 is the set of optimal solutions to the LP given by

$$\Delta_0 := \left\{\lambda \in \mathbb{R}^p \mid \lambda \geq 0,\ \langle b, \lambda\rangle \geq 0,\ A^\top\lambda = 0, \mathbb{1}\Omega\lambda \leq 1\right\}.$$

Assumption 4.25 and equation 7.77 imply that

$$d_{Pr}\Big(\mathcal{L}\Big(\zeta_b^* - \zeta_A^*\hat{\eta} \mid \mathbb{W}_N\Big), \mathcal{L}(\zeta_b - \zeta_A\eta_Q)\Big) = o_p(1).$$

Reasoning using an almost sure representation as in the proof of Theorem 3.4, equation 4.32 then follows if we show that

$$d_H(\Pi^u(\hat{\Delta}), \Pi^u(\Delta_0)) = o_p(1).$$

Equation 7.78, with $\xi = 0$, implies that

$$\vec{d}_H(\Pi^u(\Delta_0), \Pi^u(\hat{\Delta})) = o_p(1).$$

It thus remains to show that

$$\vec{d}_H(\Pi^u(\hat{\Delta}), \Pi^u(\Delta_0)) = o_p(1). \tag{7.93}$$

Using the definition of $\hat{\Delta}$ and equations 7.72 and 7.82, on the event $E_N(Q)$, we have

$$\begin{aligned}\sup_{\lambda\in\hat{\Delta}}\|A^\top\lambda\| &\leq \sup_{\lambda\in\hat{\Delta}}\|(A-\hat{A})^\top\Pi^u(\lambda)\| + (\kappa_N/\sqrt{N})\|\mathbb{1}\| \\ &\leq C\kappa_N/\sqrt{N} = o(1),\end{aligned} \tag{7.94}$$

and

$$\begin{aligned}\sup_{\lambda\in\hat{\Delta}}(\mathbb{1}^\top\Omega\lambda - 1)_+ &\leq \sup_{\lambda\in\hat{\Delta}}(\mathbb{1}^\top(\Omega-\hat{\Omega})\lambda)_+ \\ &\leq C\alpha_N = o(1),\end{aligned} \tag{7.95}$$

and

$$\begin{aligned}\sup_{\lambda\in\hat{\Delta}}(b^\top\lambda)_- &\leq \sup_{\lambda\in\hat{\Delta}}\left[(b-\hat{b})^\top\Pi^u(\lambda) + \hat{T}_N - \frac{\kappa_N}{\sqrt{N}}\right]_- \\ &\leq \sup_{\lambda\in\hat{\Delta}}\|b-\hat{b}\|\cdot\|\Pi^u(\lambda)\| + |\hat{T}_N| + \frac{\kappa_N}{\sqrt{N}} \\ &\leq C\frac{\kappa_N}{\sqrt{N}} = o(1).\end{aligned} \tag{7.96}$$

Equations 7.94, 7.94, and 7.94, in combination with Lemma 7.8, then imply that we have

$$\vec{d}_H(\hat{\Delta}, \Delta_0) = o_p(1),$$

which in turn implies equation 7.93. This completes the proof. □

The following proposition builds on Theorems 1 and 2 of Tsirel'son (1975) to derive an anti-concentration result, which is useful for establishing our uniformity results and may be of independent interest.

Proposition 7.2 (Anti-concentration). *Let* T *be an arbitrary index set. Let* $\{K_t \mid t \in T\}$ *be a family of compact subsets of* $\mathbb{R}^p$ $(K_t \subseteq \mathbb{R}^p)$*, and let* $\{Z_t \mid t \in T\}$ *be a family of mean-zero Gaussian vectors in* $\mathbb{R}^p$*, with* $Z_t \sim \mathcal{N}(0, \Sigma_t)$*. Suppose there exist strictly positive constants* M *and* σ *such that*

$$\sup_{t\in T} \|\Sigma_t\| \le M \quad \text{and} \quad \sup_{t\in T} \max_{\lambda\in K_t} \|\lambda\| \le M, \tag{7.97}$$

and that for all $t \in T$*, there exists* $\lambda_t \in K_t$ *such that*

$$\lambda_t^\top \Sigma_t \lambda_t > \sigma^2. \tag{7.98}$$

Define random variables $\{Y_t\}_{t\in T}$ *by*

$$Y_t = \max_{\lambda\in K_t} Z_t^\top \lambda.$$

Then, for any $\delta > 0$*, the distributions* $\{F_t\}_{t\in T}$ *of the random variables* $\{Y_t\}_{t\in T}$ *are absolutely continuous on the interval* $[\delta, +\infty)$ *and possess densities* $\{f_t\}_{t\in T}$ *that are uniformly bounded on* $[\delta, +\infty)$*; that is,*

$$\sup_{t\in T} \sup_{x\in[\delta,+\infty)} f_t(x) < \infty. \tag{7.99}$$

As a consequence, if $q_t(\cdot)$ *denotes the quantile function of* F_t*, then for any* $\overline{\alpha} \in (0, 1/2)$*, there exists* $\underline{x} > 0$ *such that*

$$\inf_{\alpha\in(0,\overline{\alpha})} \inf_{t\in T} q_t(1-\alpha) > 2\underline{x}, \quad \text{and} \quad \sup_{t\in T} \sup_{x\in[\underline{x},+\infty)} f_t(x) < \infty. \tag{7.100}$$

Proof of Proposition 7.2. We proceed in two steps. In the first step, we establish inequality (7.99) using (7.97) and Theorem 2 of Tsirel'son (1975). In the second step, we use (7.98) and (7.99) to establish inequality (7.100).

Step 1: Proof of (7.99). Given $\delta > 0$ as in the statement of the proposition, let $C_\delta > 0$ be a constant satisfying $MC_\delta < \delta/2$. We first show that inequality (7.98) implies

$$\inf_{t\in T} P(\|Z_t\| < C_\delta) > \eta \tag{7.101}$$

for some $\eta > 0$. Indeed, since $Z_t \sim \Sigma_t^{1/2} Z$, where $Z \sim \mathcal{N}(0, \mathbb{I}_p)$ is a standard normal vector in $\mathbb{R}^p$, we have

$$\inf_{t\in T} P(\|Z_t\| < C_\delta) \ge P(\|Z\| < C_\delta/\sqrt{M}),$$

where M is as in (7.97), and we have used the fact that $\|\Sigma_t^{1/2}\| = \|\Sigma_t\|^{1/2} \le M^{1/2}$, as Σ_t is symmetric and positive semidefinite. We can thus set $\eta = P(\|Z\| < C_\delta/\sqrt{M})$ in (7.101).

Now, by inequality (7.101) and the definition of Y_t, it follows that

$$F_t(\delta/2) = P(Y_t \le \delta/2) \ge P(|Y_t| \le \delta/2) \ge P(\|Z_t\| \le C_\delta) \ge \eta. \tag{7.102}$$

Let $\Phi(\cdot)$ denote the CDF of a standard normal variable, and for each $t \in T$, define τ_t by

$$\tau_t = \Phi^{-1}(F_t(\delta/2)).$$

Assuming without loss of generality that $\eta < 1/2$, inequality (7.102) yields

$$(\tau_t)_- := \max\{-\tau_t, 0\} \le -\Phi^{-1}(\eta). \tag{7.103}$$

Applying Theorem 2 of Tsirel'son (1975), for each $t \in T$ and for $x \in [\delta, +\infty)$, we obtain

$$f_t(x) \le \big((\tau_t)_- + 2\big)^2 \frac{x}{(x-\delta/2)^2} + \big((\tau_t)_- + 2\big)\frac{1}{x-\delta/2}.$$

Combining this with (7.103), we deduce that

$$\sup_{t\in T}\sup_{x\in[\delta,+\infty)} f_t(x) \le \big(-\Phi^{-1}(\eta)+2\big)^2\frac{4}{\delta} + \big(-\Phi^{-1}(\eta)+2\big)\frac{2}{\delta} < \infty,$$

which establishes (7.99).

Step 2: Proof of (7.100). We now prove inequality (7.100). Observe that by the definition of Y_t, it stochastically dominates $\lambda_t^\top Z_t$, where $\lambda_t \in K_t$ is as in the statement of the proposition. Hence, the $1-\alpha$ quantiles of Y_t are at least as large as those of $\lambda_t^\top Z_t$.

Moreover, for each $\alpha \in (0, 1/2)$, inequality (7.98) implies that the $1-\alpha$ quantiles of $\lambda_t^\top Z_t$ are bounded below by those of a $\mathcal{N}(0, \sigma^2)$ distribution, with σ as in (7.98). Therefore,

$$\inf_{\alpha\in(0,\overline{\alpha})}\inf_{t\in T} q_t(1-\alpha) \ge \inf_{t\in T} q_t(1-\overline{\alpha}) \ge \sigma\Phi^{-1}(1-\overline{\alpha}) := 2\underline{x}.$$

The second part of (7.100) then follows from inequality (7.99), with $\delta = \underline{x}$. □

The following two lemmas are useful in deriving the (pointwise) asymptotic distributions of our test statistics. The first lemma gives the asymptotic distribution of the test statistics in Theorems 4.7, 4.14, and 4.21, while the second lemma is used to derive the asymptotic distribution of the test statistic in 4.29. Foth are variations on the results of Shapiro (1991).

Lemma 7.3. *Consider the linear program* $\nu = \max\{b^\top\lambda \mid A^\top\lambda \le 0,\ \|D\lambda\|_1 \le 1\}$. *Let* $\hat{b}$ *and* $\hat{D}$ *denote estimators of* b *and* D, *respectively, such that* $\sqrt{N}(\hat{b} - b) \Rightarrow \zeta^b$ *and* $\hat{D} - D = o_p(1)$. *Let* K *and* U *form a partition of* $[p]$, *where* p *denotes the dimension of* b. *We allow* K *to be empty, but require* U *to be nonempty. The set* K *corresponds to the indices of the deterministic or known components of* b, *and we assume that*

$$\forall i \in K,\quad \hat{b}_i = b_i \quad \textit{and} \quad \hat{D}_{ii} = D_{ii} = 0, \qquad \textit{and} \qquad \exists\, \underline{\sigma} > 0 \ \textit{s.t.}\ D_{ii} > \underline{\sigma} \quad \forall i \in U. \tag{7.104}$$

Define $\Delta_0 := \arg\max\{b^\top\lambda \mid A^\top\lambda \le 0,\ \|D\lambda\|_1 \le 1\}$. *If* $\nu = 0$, *then*

$$\sqrt{N}(\hat{\nu} - \nu) = \sqrt{N}\hat{\nu} \xrightarrow{d} \max_{\lambda\in\Delta_0}\langle\zeta^b, \lambda\rangle. \tag{7.105}$$

Here $\hat{\nu} = \max\{\hat{b}^\top\lambda \mid A^\top\lambda \le 0,\ \|\hat{D}\lambda\|_1 \le 1\}$.

Proof of Lemma 7.3. Let $\hat{\zeta} = (\sqrt{N}(\hat{b} - b), \hat{D} - D)$. The hypothesis of the theorem implies that $\hat{\zeta} \xrightarrow{d} \zeta := (\zeta^b, 0)$. By the Skorokhod representation theorem, there exists a sufficiently rich probability space, and versions $\xi_N \overset{d}{=} \hat{\zeta}$ and $\xi := (\xi^b, 0) \overset{d}{=} \zeta$ such that $a.s. - \lim_{N\to\infty}\xi_N = \xi$. Using the notation used in the derivation of equation 4.8, let $\mu = (b, D)$ and $\mu_N(\xi_N) = (b_N(\xi_N), D_N(\xi_N)) = (b + \frac{\xi_N^b}{\sqrt{N}}, D + \xi_N^D)$. We have $\nu = \phi(\mu)$ and $\hat{\nu} \overset{d}{=} \phi(\mu_N(\xi_N))$. For $\lambda \in \mathbb{R}^p$, let $\Pi^u : \mathbb{R}^p \to \mathbb{R}^p$ be the projection defined by

$$[\Pi^u(\lambda)]_i = \lambda_i \quad \forall i \in U \qquad \text{and} \qquad [\Pi^u(\lambda)]_i = 0 \quad \forall i \in K,$$

where K and U are as in equation 7.104. That is, $\Pi^u(\cdot)$ is the projection on the subspace spanned by the cannonical basis vectors with indices in U. Condition 7.104 then implies that we necessarily have (with prob. 1)

$$[\xi^b]_i = [\xi_N^b]_i = 0 \quad \forall i \in K \qquad \text{and} \qquad [\xi_N^D]_{ii} = 0 \quad \forall i \in K.$$

We show below, using an argument similar to that of Theorem 17 in Gol'shtein (1972), that

$$a.s. - \lim_{N\to\infty}\sqrt{N}\phi(\mu_N(\xi_N)) = \max_{\lambda\in\Delta_0}\langle\lambda, \xi^b\rangle$$

from which the conclusion of the lemma follows.

Note that the value of our LP has the following equivalent representations

$$\begin{aligned}\nu &= \max\{b^\top\lambda \mid A^\top\lambda \le 0,\ D\lambda = u - v,\ \mathbb{1}^\top(u + v) \le 1,\ u, v \ge 0\} \\ &= \min\{x \mid x \ge 0,\ z \ge 0,\ Az - b = Dy,\ \|y\|_\infty \le x\}.\end{aligned} \tag{7.106}$$

Let $L(\lambda, u, v, x, y, z)$ denote the Lagrangian associated with the unperturbed LP, and given by

$$L(\lambda, u, v, x, y, z) = \langle\lambda, b\rangle + y^\top(D\lambda - u + v) + x(1 - \mathbb{1}^\top(u + v)) - z^\top A^\top\lambda.$$

Similarly, let $L_N(\lambda, u, v, x, y, z; \xi_N)$ denote the Lagrangian associated with the perturbed LP, and defined by

$$L_N(\lambda, u, v, x, y, z; \xi_N) = \langle\lambda, b_N(\xi_N)\rangle + y^\top(D_N(\xi_N)\lambda - u + v) + x(1 - \mathbb{1}^\top(u + v)) - z^\top A^\top\lambda.$$

Let $(\lambda_0, u_0, v_0, x_0, y_0, z_0)$ (resp. $(\lambda(\xi_N), u(\xi_N), v(\xi_N), x(\xi_N), y(\xi_N), z(\xi_N))$) denote the optimal primal and dual solutions to the unperturbed (resp. perturbed) linear programs. Then, by the saddle point property of the Lagrangian, and by noting that the condition $\nu = 0$ impilies that $x_0 = 0$ and $y_0 = 0$, we have

$$\begin{aligned}
\sqrt{N}\phi(\mu_N(\xi_N)) &= \sqrt{N}\Big(L(\lambda(\xi_N), u(\xi_N), \nu(\xi_N), x(\xi_N), y(\xi_N), z(\xi_N); \xi_N) - L(\lambda_0, u_0, \nu_0, x_0, y_0, z_0)\Big) \\
&\leq \sqrt{N}\Big(L(\lambda(\xi_N), u(\xi_N), \nu(\xi_N), x_0, y_0, z_0; \xi_N)) - L(\lambda(\xi_N), u(\xi_N), \nu(\xi_N), x_0, y_0, z_0)\Big) \\
&= \langle \xi_N^b, \lambda(\xi_N)\rangle = \langle \xi_N^b, \Pi^u(\lambda(\xi_N))\rangle.
\end{aligned} \tag{7.107}$$

We now work toward establishing equation 7.115. To that end, we want to show that

$$\vec{d}_H\big(\Pi^u(\Delta_{0,N}(\xi_N)), \Pi^u(\Delta_0)\big) \overset{a.s.}{\to} 0, \tag{7.108}$$

where

$$\Delta_{0,N}(\xi_N) := \arg\max\{b_N(\xi_N)^\top\lambda \mid A^\top\lambda \leq 0,\ \|D_N(\xi_N)\lambda\|_1 \leq 1\}.$$

Note that, almost surely, for all sufficiently large N such that $\|\xi_N^D\| \leq \underline{\sigma}/2$, we have

$$\|D_N(\xi_N)\lambda\|_1 = \|\Pi^u(D_N(\xi_N)\lambda)\|_1 \geq (\underline{\sigma}/2)\|\Pi^u(\lambda)\|_1.$$

with

$$\mathcal{D}_N(\xi_N) := \big\{\lambda \in \mathbb{R}^p \mid A^\top\lambda \leq 0,\ \|D_N(\xi_N)\lambda\|_1 \leq 1\big\}.$$

As a consequence, we have

$$a.s. - \overline{\lim}_{N\to\infty} \sup_{\lambda\in\mathcal{D}_N(\xi_N)} \|\Pi^u(\lambda)\|_1 \leq 2/\underline{\sigma}. \tag{7.109}$$

A similar argument gives

$$\sup_{\lambda\in\mathcal{D}} \|\Pi^u(\lambda)\|_1 \leq 1/\underline{\sigma}. \tag{7.110}$$

Since, the smallest nonzero diagonal element of D is bounded away from zero, for all N sufficiently large that $\|\xi_N^D\| < \underline{\sigma}$, we have

$$\big(1 - \|\xi_N^D\|/\underline{\sigma}\big)\|D\lambda\|_1 \leq \|D_N(\xi_N)\lambda\|_1 \leq (1 + \|\xi_N^D\|/\underline{\sigma})\|D\lambda\|_1. \tag{7.111}$$

Therefore, as the feasible regions are star-shaped w.r.t the origin, if $\lambda \in \Delta_{0,N}(\xi_N)$, then

$$\Gamma_N(\lambda) := (1 - \|\xi_N^D\|/\underline{\sigma})\lambda \in \mathcal{D} \qquad \text{and} \qquad \|\Pi^u(\lambda) - \Pi^u(\Gamma_N(\lambda))\| \leq \|\Pi^u(\lambda)\| \times \|\xi_N^D\|/\underline{\sigma}.$$

The latter inequality and equation 7.109 then yields

$$\vec{d}_H(\Pi^u(\Delta_{0,N}(\xi_N)), \Pi^u(\Gamma_N(\Delta_{0,N}(\xi_N)))) \xrightarrow{a.s.} 0. \tag{7.112}$$

We have

$$
\begin{aligned}
0 \geq \sup_{\lambda\in\Delta_{0,N}(\xi_N)} \langle b, \Gamma_N(\lambda)\rangle &= \inf_{\lambda\in\Delta_{0,N}(\xi_N)} \langle b, \Gamma_N(\lambda)\rangle \\
&= \inf_{\lambda\in\Delta_{0,N}(\xi_N)} \langle b - b_N(\xi_N), \Gamma_N(\lambda)\rangle + \langle b_N(\xi_N), \Gamma_N(\lambda)\rangle \\
&\geq \inf_{\lambda\in\Delta_{0,N}(\xi_N)} \langle b - b_N(\xi_N), \Gamma_N(\lambda)\rangle = \inf_{\lambda\in\Delta_{0,N}(\xi_N)} \langle \xi_N^b/\sqrt{N}, \Pi^u(\Gamma_N(\lambda))\rangle \\
&\geq \inf_{\lambda\in\mathcal{D}} \langle \xi_N^b/\sqrt{N}, \Pi^u(\lambda)\rangle \\
&\geq -\sup_{\lambda\in\mathcal{D}} \|\Pi^u(\lambda)\| \frac{\|\xi_N^b\|}{\sqrt{N}} \xrightarrow{a.s.} 0
\end{aligned} \tag{7.113}
$$

where the first inequality follows from the condition $\nu = 0$; the second inequality follows from the fact that $\langle b_N(\xi_N), \lambda\rangle \geq 0$ for all $\lambda \in \Delta_{0,N}(\xi_N)$, since $0 \in \mathcal{D}_N(\xi_N)$ and $\Gamma_N(\lambda)$ is a positive scalar multiple of λ; the third inequality follows from the inclusion $\Gamma_N(\Delta_{0,N}(\xi_N)) \subseteq \mathcal{D}$; and the almost sure convergence follows from equation 7.110 and the fact that $\xi_N^b \overset{a.s.}{\to} 0$. By Lemma 7.5, Inequality 7.113 and the fact that $\nu = 0$ then imply

$$
\vec{d}_H(\Gamma_N(\Delta_{0,N}(\xi_N)), \Delta_0) \xrightarrow{a.s.} 0,
$$

and we thus have

$$
\vec{d}_H(\Pi^u(\Gamma_N(\Delta_{0,N}(\xi_N))), \Pi^u(\Delta_0)) \xrightarrow{a.s.} 0. \tag{7.114}
$$

Equation 7.108 then follows from equations 7.112 and 7.114.

We can now take the limit of both sides of equation 7.107, and using equation 7.108 gives

$$
\begin{aligned}
a.s. - \overline{\lim}_{N\to\infty} \sqrt{N}\phi(\mu_N(\xi_N)) &\leq a.s. - \overline{\lim}_{N\to\infty} \langle \xi_N^b, \Pi^u(\lambda(\xi_N))\rangle \\
&\leq a.s. - \overline{\lim}_{N\to\infty} \sup_{\lambda\in\Delta_0} \langle \xi_N^b, \Pi^u(\lambda)\rangle + \|\xi_N^b\| \vec{d}_H(\Pi^u(\Delta_{0,N}(\xi_N)), \Pi^u(\Delta_0)) \\
&= \sup_{\lambda\in\Delta_0} \langle \xi^b, \lambda\rangle.
\end{aligned} \tag{7.115}
$$

To complete the argument, it remains to establish the converse of inequality 7.115. To that end, using the saddle point property and adopting the same notation as in the derivation of inequality 7.107, we have

$$
\begin{aligned}
\sqrt{N}\phi(\mu_N(\xi_N)) &= \sqrt{N}\Big(L(\lambda(\xi_N), u(\xi_N), v(\xi_N), x(\xi_N), y(\xi_N), z(\xi_N); \xi_N) - L(\lambda_0, u_0, v_0, x_0, y_0, z_0)\Big) \\
&\geq \sqrt{N}\Big(L(\lambda_0, u_0, v_0, x(\xi_N), y(\xi_N), z(\xi_N); \xi_N) - L(\lambda_0, u_0, v_0, x(\xi_N), y(\xi_N), z(\xi_N))\Big) \\
&= \langle \xi_N^b, \lambda_0\rangle + \sqrt{N}\langle y(\xi_N), \xi_N^D \lambda_0\rangle.
\end{aligned} \tag{7.116}
$$

The dual representation of our LP (equation 7.106) and inequality 7.115 imply that

$$
a.s. - \overline{\lim}_{N\to\infty} \sqrt{N}\, \|y(\xi_N)\|_\infty = a.s. - \overline{\lim}_{N\to\infty} \sqrt{N}\, x(\xi_N) = a.s. - \overline{\lim}_{N\to\infty} \sqrt{N}\phi(\mu_N(\xi_N)) < \infty,
$$

and since $\xi_N^D \xrightarrow{a.s.} 0$, we conclude that

$$a.s. - \lim_{N\to\infty} \sqrt{N}\langle y(\xi_N), \xi_N^D \lambda_0\rangle = 0.$$

Equation 7.116 thus gives

$$a.s. - \underline{\lim} \sqrt{N}\phi(\mu_N(\xi_N)) \geq \max_{\lambda\in\Delta_0}\langle \xi^b, \lambda\rangle, \tag{7.117}$$

which establishes the converse to 7.115, and proves the claim. □

Lemma 7.4. *Consider the linear program* $\nu = \max\{b^\top\lambda \mid \lambda \geq 0,\ A^\top\lambda = 0,\ \|\Omega\lambda\|_1 \leq 1\}$ *in 4.25. And let* $\hat{d}$ *be an estimator of* $d = \mathrm{vec}([A \quad b])$ *such that*

$$\sqrt{N}(\hat{d} - d) \xrightarrow{d} \mathcal{N}(0, \Sigma) \qquad \textit{and} \qquad \hat{\Omega} - \Omega \xrightarrow{p} 0 \tag{7.118}$$

where Ω *and* $\hat{\Omega}$ *are as defined in Section 4.4. Let* $U \subset [p]$ *denote the set of indices of positive diagonal elements of* Ω. *Assume that the system*

$$b - A\eta \leq 0$$

has a unique solution η^*. *For* $\hat{T}_N$ *as defined in equation 4.26, we have*

$$\sqrt{N}\hat{T}_N \xrightarrow{d} \max_{\lambda\in\Delta_0}\langle \zeta_b - \zeta_A\eta^*, \lambda\rangle, \tag{7.119}$$

where $\zeta = \mathrm{vec}([\zeta_A \quad \zeta_b]) \sim \mathcal{N}(0, \Sigma)$, *and* $\Delta_0 = \arg\max\{b^\top\lambda \mid \lambda \geq 0,\ A^\top\lambda = 0,\ \|\Omega\lambda\|_1 \leq 1\}$.

Proof of Lemma 7.4. We proceed using an almost sure representation argument as in the preceding lemma. Let $\hat{\zeta} = (\sqrt{N}(\hat{A} - A), \sqrt{N}(\hat{b} - b),\ \hat{\Omega} - \Omega)$. The hypothesis of the theorem implies that $\hat{\zeta} = (\hat{\zeta}_A, \hat{\zeta}_b, \hat{\zeta}_\Omega) \xrightarrow{d} \zeta := (\zeta_A, \zeta_b, 0)$ where $\mathrm{vec}([\zeta_A \quad \zeta_b]) \sim \mathcal{N}(0, \Sigma)$. By the Skorokhod representation theorem, there exists a sufficiently rich probability space, and versions $\xi_N \overset{d}{=} \hat{\zeta}$ and $\xi := (\xi_A, \xi_b, 0) \overset{d}{=} \zeta$ such that $a.s.{-}\lim_{N\to\infty} \xi_N = \xi$. Using the notation used the proof of Theorem 4.29 (Section 7.0.1), let $\mu = (A, b, \Omega)$ and $\mu_N(\xi_N) = (A_N(\xi_N), b_N(\xi_N), \Omega_N(\xi_N)) = \left(A + \frac{\xi_{AN}}{\sqrt{N}}, b + \frac{\xi_{bN}}{\sqrt{N}}, \Omega + \xi_{\Omega N}\right)$. We have $\nu = \phi(\mu)$ and $\hat{\nu} \overset{d}{=} \phi(\mu_N(\xi_N))$. Let the Lagrangians $L_N(\lambda, \eta, t; \xi)$ and $L(\lambda, \eta, t)$ be as defined in Section 7.0.1.

Reasoning as in the derivation of equation 7.70, we have

$$\sqrt{N}\phi_N(\mu_N(\xi_N)) \leq \langle \xi_{bN} - \xi_{AN}\eta^*, \lambda(\xi_N)\rangle \tag{7.120}$$

where $\lambda(\xi_N)$ is any element in $\Delta_{0,N}(\xi_N) = \arg\max\{b_N(\xi_N)^\top\lambda \mid \lambda \geq 0,\ A_N(\xi_N)^\top\lambda = 0,\ \mathbb{1}^\top\Omega_N(\xi_N)\lambda \leq 1\}$, and η^* is as in the statement of the lemma.

We now show that

$$a.s. - \lim_{N\to\infty} \vec{d}_H(\Delta_{0,N}(\xi_N), \Delta_0) = 0. \tag{7.121}$$

We have

$$\sup_{\lambda\in\Delta_{0,N}(\xi_N)} \|A^\top\lambda\| = \sup_{\lambda\in\Delta_{0,N}(\xi_N)} \|(A - A_N(\xi_N))^\top\Pi^u(\lambda)\| \leq \|A - A_N(\xi_N)\| \sup_{\lambda\in\Delta_{0,N}(\xi_N)} \|\Pi^u(\lambda)\| \xrightarrow{a.s.} 0,$$

and

$$\sup_{\lambda\in\Delta_{0,N}(\xi_N)} (\mathbb{1}^\top\Omega\lambda - 1)_+ \le \sup_{\lambda\in\Delta_{0,N}(\xi_N)} (\mathbb{1}^\top\Omega\lambda - 1 + 1 - \mathbb{1}^\top\Omega_N(\xi_N)\lambda)_+$$
$$\le \|\Pi^u(\mathbb{1})\|\|\Omega - \Omega_N(\xi_N)\| \sup_{\lambda\in\Delta_{0,N}(\xi_N)} \|\Pi^u(\lambda)\| \xrightarrow{a.s.} 0,$$

where we have used equation 7.72. The latter two inequalities and Lemma 7.8 yield equation 7.121.

Equations 7.120 and 7.121 imply that

$$\begin{aligned} a.s. - \overline{\lim}_{N\to\infty}\sqrt{N}\phi_N(\mu_N(\xi_N)) &\le a.s. - \overline{\lim}_{N\to\infty} \sup_{\lambda\in\Delta_0} \langle \xi_{bN} - \xi_{AN}\eta^*, \lambda\rangle + \\ &\quad \vec{d}_H(\Delta_{0,N}(\xi_N), \Delta_0)\|\xi_{bN} - \xi_{AN}\eta^*\| \\ &= \sup_{\lambda\in\Delta_0} \langle \xi_b - \xi_A\eta^*, \lambda\rangle. \end{aligned} \tag{7.122}$$

We now derive the converse to equation 7.122. Reasoning as in the derivation of equation 7.116, we have

$$\begin{aligned} \sqrt{N}\,\phi(\mu_N(\xi_N)) &\ge \sqrt{N}\big(L_N(\lambda_0, \eta(\xi_N), t(\xi_N); \xi_N) - L(\lambda_0, \eta(\xi_N), t(\xi_N))\big) \\ &= \xi_{bN}^\top\lambda_0 - \eta(\xi_N)^\top\xi_{AN}^\top\lambda_0 - t(\xi_N)\mathbb{1}^\top\xi_{\Omega N}\lambda_0, \end{aligned} \tag{7.123}$$

where $\lambda_0 \in \Delta_0$ and $(\eta(\xi_N), t(\xi_N)) \in S_{0,N}(\xi_N)$, the set of optimal solutions to the (perturbed) dual LP 7.69 with input $\mu_N(\xi_N)$ (see Section 7.0.1). Note that all elements of the set $S_{0,N}(\xi_N)$ have the same second component, given by the value of the perturbed LP: i.e., $t(\xi_N) = \phi(\mu_N(\xi_N))$. We first show that

$$a.s. - \overline{\lim}_{N\to\infty} t(\xi_N)) = 0. \tag{7.124}$$

Indeed, we have

$$\phi(\mu_N(\xi_N)) = \min\{t \mid t\Omega_N(\xi_N)\mathbb{1} + A_N(\xi_N)\eta \ge b_N(\xi_N),\ t \ge 0\}.$$

Since $b - A\eta^* \le 0$, we have

$$\begin{aligned} b_N(\xi_N) - A_N(\xi_N)\eta^* &\le \big(b_N(\xi_N) - b\big) - \big(A_N(\xi_N) - A\big)\eta^* \\ &\le \Big(\|b - b_N(\xi_N)\| + \|\eta^*\|\|A - A_N(\xi_N)\|\Big)\Pi^u(\mathbb{1}). \end{aligned}$$

Equation 7.124 then follows since $\|b - b_N(\xi_N)\| + \|\eta^*\|\|A - A_N(\xi_N)\| \xrightarrow{a.s.} 0$ and $[\Omega]_{ii} > 0$ for all $i \in U$.

We now show that

$$a.s. - \lim \vec{d}_H(\Pi_1(S_{0,N}(\xi_N)), \{\eta^*\}) = 0 \tag{7.125}$$

where $\Pi_1(S_{0,N(\xi_N)}) := \big\{\eta \mid (\eta, t(\xi_N)) \in S_{0,N}(\xi_N)\big\}$ is the projection of $S_{0,N}(\xi_N)$ on its first component. Indeed, for every $\eta \in \Pi_1(S_{0,N}(\xi_N))$, we have

$$b_N(\xi_N) - A_N(\xi_N)\eta \le t(\xi_N)\Omega_N(\xi_N)\mathbb{1}.$$

Note that if (A_N, b_N, t_N) is any sequence such that $A_N \to A$ and $b_N \to b$, and $t_N \to 0$, then the set

$\mathcal{C}_N = \{\eta \mid b_N - A_N\eta \le t_N\}$ satisfies

$$\overline{\lim}_{N\to\infty} \sup_{\eta\in\mathcal{C}_N} \|\eta\| < \infty.$$

If not, then, after passing to a subsequence, we can assume that there exist $\eta_N \in \mathcal{C}_N$ such that $\|\eta_N\| \to \infty$. But then we have

$$b_N/\|\eta_N\| - A_N(\eta_N/\|\eta_N\|) \le t_N/\|\eta_N\|.$$

After passing to a further subsequence, we can assume that $\eta_N/\|\eta_N\| \to \eta_0 \ne 0$, and we have

$$-A\eta_0 \le 0,$$

which contradicts the assumption of the lemma that the inequality system $b - A\eta \le 0$ has a unique solution η^*. Thus, we have

$$a.s. - \overline{\lim}_{N\to\infty} \sup_{\eta\in\Pi_1(S_{0,N}(\xi_N))} \|\eta\| < \infty,$$

and

$$\begin{aligned} a.s. - \overline{\lim} \sup_{\eta\in\Pi_1(S_{0,N}(\xi_N))} \|[b - A\eta]_+\| \le a.s. - \overline{\lim}\Big\{&\|b - b_N(\xi_N)\| + \|A - A_N(\xi_N)\| \sup_{\eta\in\Pi_1(S_{0,N}(\xi_N))} \|\eta\| + \\ &t(\xi_N)\|\Omega_N(\xi_N)\|\|\mathbb{1}\|\Big\} \\ &= 0. \end{aligned}$$

The latter inequality and Lemma 7.8 then imply equation 7.125.

Combining equations 7.123, 7.124, and 7.125, we get

$$\begin{aligned} a.s. - \underline{\lim}_{N\to\infty}\sqrt{N}\phi(\mu_N(\xi_N)) &\ge a.s. - \underline{\lim}_{N\to\infty} \sup_{\eta\in\Pi_1(S_{0,N}(\xi_N))} \sup_{\lambda\in\Delta_0} \left\{\xi_{bN}^\top\lambda - \eta^\top\xi_{AN}^\top\lambda - t(\xi_N)\mathbb{1}^\top\xi_{\Omega N}\lambda\right\} \\ &= \sup_{\lambda\in\Delta_0} \langle \xi_b - \xi_A\eta^*, \lambda\rangle. \end{aligned} \tag{7.126}$$

Combining equations 7.122 and 7.126 yield equation 7.119, and the proof is complete.

□

The following lemma establishes the set of maximizers of a LP is well separated in the sense that values of the objective function at feasible points that are at some fixed distance from the set of optimizers are strictly below the optimal value, even if the feasible region is not compact.

Lemma 7.5. *Consider the LP* $\nu = \max\{b^\top\lambda \mid A\lambda \le d\}$*, where* ν *is finite, and the feasible region* $\mathcal{D} := \{\lambda \in \mathbb{R}^p \mid A\lambda \le d\}$ *is not necessarily bounded. Let* $\Delta_0 = \arg\max\{b^\top\lambda \mid A\lambda \le d\}$ *denote the set of optimal solutions, and given* $\epsilon > 0$*, let* Δ_ϵ *denote the set of* $\epsilon-$*optimal solutions defined by* $\Delta_\epsilon := \{\lambda \in \mathcal{D} \mid b^\top\lambda \ge \nu - \epsilon\}$*. Then there exists a constant* C *that depends on the inputs* (A, b, d)*, such that*

$$\vec{d}_H(\Delta_\epsilon, \Delta_0) \le C\epsilon. \tag{7.127}$$

Proof of Lemma 7.5. The dual LP is given by

$$\min\{d^\top z \mid z \geq 0,\ A^\top z = b\}.$$

Let z^* denote an optimal dual solution, and let $J \subseteq [m]$ denote the set of nonzero indices of z^*, where m denotes the number of rows of the matrix A. If $J = \emptyset$, then $z^* = 0$, and since $A^\top z^* = b$, we have $b = 0$, from which we deduce that $\Delta_0 = \mathcal{D}$, and the claim trivially holds. So for the remainder of the proof, let us assume that $J \neq \emptyset$.

Let the face F of $\mathcal{D}$ be given by

$$F := \{\lambda \in \mathbb{R}^p \mid A_J\lambda = d_J \quad \text{and} \quad A_{J^c}\lambda \leq d_{J^c}\}$$

where $J^c = [m]\backslash J$, and A_J is the matrix obtained from A by deleting the rows not in J (with a similar definition for d_J, A_{J^c} and d_{J^c}). We assert that F is equal to the set of optimal solutions, i.e., $F = \Delta_0$. Indeed, if $\lambda \in F$ then

$$b^\top\lambda = \langle A^\top z^*, \lambda\rangle = \langle A^\top z^*, \lambda\rangle = \langle z_J^*, A_J\lambda\rangle = \langle z_J^*, d_J\rangle = \langle z^*, d\rangle = \nu,$$

and we have $F \subseteq \Delta_0$. For the reverse inclusion, note that if $\lambda^* \in \Delta_0$, then

$$\langle z^*, d\rangle = \nu = \langle \lambda^*, b\rangle = \langle \lambda^*, A^\top z^*\rangle \quad \Longrightarrow \quad \langle z^*, A\lambda^* - d\rangle = 0 \Longleftrightarrow \langle z_J^*, A_J\lambda^* - d_J\rangle = 0.$$

The last equation then implies that $A_J\lambda^* - d_J = 0$, since we necessarily have $A_J\lambda^* - d_J \leq 0$, and all entries of z_J* are strictly positive.[45]. As $\lambda^* \in \mathcal{D}$, we clearly have $A_{J^c}\lambda^* \leq d_{J^c}$. Therefore, $\lambda^* \in \Delta_0$ implies that $A_J\lambda^* - d_J = 0$ and $A_{J^c}\lambda^* \leq d_{J^c}$, which gives the second inclusion $\Delta_0 \subseteq F$, and we thus conclude that $F = \Delta_0$.

Now for $\lambda \in \Delta_\epsilon$ we have

$$\left[0 \leq \nu - b^\top\lambda \leq \epsilon\right] \Longleftrightarrow \left[0 \leq \langle z^*, d\rangle - \langle A^\top z^*, \lambda\rangle \leq \epsilon\right] \Longleftrightarrow \left[0 \leq \langle z^*, d - A\lambda\rangle \leq \epsilon\right] \Longleftrightarrow \left[0 \leq \langle z_J^*, d_J - A_J\lambda\rangle \leq \epsilon\right].$$

If $\underline{z} := \min\{z_j^* \mid j \in J\}$ denotes the smallest value of z^* on its support, then the last inequality implies that for all $\lambda \in \Delta_\epsilon$, we have

$$\|d_J - A_J\lambda\|_1 \leq \epsilon/\underline{z}. \tag{7.128}$$

We recall that Hoffman's Theorem (Lemma 7.8) implies that there exists a constant $C > 0$ such that for all $\lambda \in \mathbb{R}^p$ we have

$$d(\lambda, F) \leq C\Big\{\|d_J - A_J\lambda\| + \|(d_{J^c} - A_{J^c}\lambda)_-\|\Big\}.$$

For $\lambda \in \Delta_\epsilon$, we necessarily have $(d_{J^c} - A_{J^c}\lambda)_- = 0$, and Hoffman's Theorem and equation 7.128 then implies that

$$\sup_{\lambda \in \Delta_\epsilon} d(\lambda, F) \leq C\Big\{\|d_J - A_J\lambda\|\Big\} \leq (C/\underline{z})\epsilon$$

which completes the proof of inequality 7.127. □

[45] This corresponds to the complementary slackness condition in linear programming.

The following lemma establishes the Hausdorff continuity of the set of extreme points of our feasible regions under perturbations of the normalizing matrix D. It extends Lemma 4 in Andrews, Roth, and Pakes (2023) to settings in which the points λ in the feasible region are not restricted to lie in an orthant of $\mathbb{R}^p$, and where the constraint used for the normalization may correspond to intersecting the feasible region with more than one halfspace. [46]

Lemma 7.6. *Given a diagonal matrix* $D \in \mathbb{R}^{n\times n}$ *and a matrix* $B \in \mathbb{R}^{m_B\times n}$*, let* $\mathcal{D}(B, D)$ *denote a polyhedral set defined by*

$$\mathcal{D}(B, D) := \{\lambda \in \mathbb{R}^n \,|\, B\lambda \le 0,\ \ \|D\lambda\|_1 \le 1\}.$$

For D *a diagonal matrix, let* $\operatorname{supp}(D)$ *denote the support of* D*, defined as the set of indices corresponding to nonzero diagonal entries of* D*:*

$$\operatorname{supp}(D) := \{i \in [n] \,|\, D_{ii} \ne 0\}.$$

Let $\operatorname{extr}(\mathcal{D})$ *denote the (possibly empty) set of extreme points of* $\mathcal{D}$*. Let* $\underline{\sigma}$ *and* $\overline{\sigma}$ *be two fixed positive constants such that* $\underline{\sigma} < \overline{\sigma}$ *. Then for all matrices* B *and diagonal matrices* D *and* D' *satisfying*

$$\operatorname{supp}(D) = \operatorname{supp}(D'), \quad \|D - D'\| \le \underline{\sigma}/2, \quad \textit{and} \quad \underline{\sigma} \le D_{ii} \le \overline{\sigma}, \quad \forall i \in \operatorname{supp}(D)$$

the following hold:

i) For all $\lambda \in \operatorname{extr}(\mathcal{D}(B, D))\backslash\{0\}$*, we have* $\|D'\lambda\|_1 > 0$*, and if we set* $\lambda' := \lambda/\|D'\lambda\|_1$*, then* $\lambda' \in \operatorname{extr}(\mathcal{D}(B, D'))\backslash\{0\}$*, and we have*

$$\lambda - \lambda' = \Big(1 - \frac{1}{\|D'\lambda\|_1}\Big)\lambda \qquad \textit{and} \qquad \Big|1 - \frac{1}{\|D'\lambda\|_1}\Big| \le \frac{2}{\underline{\sigma}}\|D' - D\|. \tag{7.129}$$

ii) For all $\lambda' \in \operatorname{extr}(\mathcal{D}(B, D'))\backslash\{0\}$*, we have* $\|D\lambda'\|_1 > 0$*, and if we set* $\lambda := \lambda'/\|D\lambda'\|_1$*, then* $\lambda \in \operatorname{extr}(\mathcal{D}(B, D))\backslash\{0\}$*, and we have*

$$\lambda' - \lambda = \Big(1 - \frac{1}{\|D\lambda'\|_1}\Big)\lambda' \qquad \textit{and} \qquad \Big|1 - \frac{1}{\|D\lambda'\|_1}\Big| \le \frac{1}{\underline{\sigma}}\|D' - D\|. \tag{7.130}$$

Proof of Lemma 7.6. We establish 7.129. The proof of 7.130 is similar. Note that as the sets $\mathcal{D}(B, D)$ and $\mathcal{D}(B, D')$ are such that we have $\lambda \in \mathcal{D}(B, D)$ iff $\exists t > 0$ such that $t\lambda \in \mathcal{D}(B, D')$. Hence $0 \in \operatorname{extr}(\mathcal{D}(B, D))$ iff $0 \in \operatorname{extr}(\mathcal{D}(B, D'))$, and it suffices to consider the nonzero elements of $\operatorname{extr}(\mathcal{D}(B, D))$. To that end, we first note that the inequality $\|\lambda\|_1 \le 1$ can be expressed as a set of 2^n inequality constraints, given by $H\lambda \le 1$, where the rows of H are given by the elements of $\{-1, 1\}^n$. Thus the set $\mathcal{D}(B, D)$ has the equivalent representation

$$\mathcal{D}(B, D) = \big\{\lambda \in \mathbb{R}^n \mid B\lambda \le 0,\ HD\lambda \le \mathbb{1}_{2^n}\big\}$$

[46] When, as in Andrews, Roth, and Pakes (2023), $\lambda \in \mathbb{R}^p_+$ for all λ in the feasible region, the normalizing constraint $\|D\lambda\|_1 \le 1$ is equivalent to the inequality $\mathbb{1}^\top D\lambda \le 1$. When λ is not restricted to lie in a specific orthant, the constraint $\|D\lambda\|_1 \le 1$ is no longer equivalent to one inequality, and may be equivalent with up to 2^p inequalities.

where $\mathbb{1}_{2^n}$ represent the vector of all ones of dimension 2^n. For $\lambda \in \mathbb{R}^n$, define the support of λ as $\mathrm{supp}(\lambda) = \{i \in [n] \mid \lambda_i \neq 0\}$. Let $e_{n,j}$, for $j \in [n]$, denote the canonical basis vectors in $\mathbb{R}^n$.

By the characterization of extreme points of a polyhedron, we have $\lambda \in \mathcal{D}(B, D) \setminus \{0\}$ if and only if $\lambda \in \mathcal{D}(B, D)$ and there exist subsets $J_1 \subseteq [m_B]$ and $J_2 \subseteq [2^n]$ such that $J_2 \neq \emptyset$ and $n_1 + n_2 = n$, where $n_1 := |J_1|$ and $n_2 := |J_2|$, the matrix $\begin{pmatrix} B_{J_1} \\ H_{J_2} D \end{pmatrix}$ is invertible, and λ is the unique solution to the equation[47]

$$\begin{pmatrix} B_{J_1} \\ H_{J_2} D \end{pmatrix} \lambda = \begin{pmatrix} 0 \\ \mathbb{1}_{n_2} \end{pmatrix}. \tag{7.131}$$

For all $i \in [n_2]$, we have

$$e_{n_2,i}^\top H_{J_2} D\lambda = \|D\lambda\|_1 = 1.$$

As a consequence, when restricted to the columns in the support of $D\lambda$, the rows of H_{J_2} must all be equal, with entries equal to the sign of $D\lambda$; that is, for $i, j \in [n_2]$ and $k \in \mathrm{supp}(D\lambda)$, we have

$$e_{n_2,i}^\top H_{J_2} e_{n,k} = e_{n_2,j}^\top H_{J_2} e_{n,k}, \quad \text{and} \quad e_{n_2,i}^\top H_{J_2} e_{n,k} = \mathrm{sign}(e_{n,k}^\top D\lambda). \tag{7.132}$$

Note that as D' is also a positive matrix with same support as D, we have $\mathrm{supp}(D'\lambda) = \mathrm{supp}(D\lambda)$, $\mathrm{sign}(D\lambda) = \mathrm{sign}(D'\lambda)$[48], and the second part of equation 7.132 remains valid when D is replaced with D', which yields

$$H_{J_2} D'\lambda = \|D'\lambda\|_1 \mathbb{1}_{n_2}, \qquad \text{with} \qquad \|D'\lambda\|_1 > 0. \tag{7.133}$$

If we set $\lambda' = \frac{\lambda}{\|D'\lambda\|_1}$, then $B\lambda \leq 0$ (by homogeneity) and $\|D'\lambda'\|_1 = 1$, and thus $\lambda' \in \mathcal{D}(B, D')$. We now prove that $\lambda' \in \mathrm{extr}(\mathcal{D}(B, D'))$. Suppose that there exist $\alpha \in (0, 1)$ and $\lambda_1, \lambda_2 \in \mathcal{D}(B, D')$ such that

$$\lambda' = \alpha\lambda_1 + (1 - \alpha)\lambda_2. \tag{7.134}$$

For $i \in \{1, 2\}$, as $\lambda_i \in \mathcal{D}(B, D')$, we must have $\|D'\lambda_i\|_1 \leq 1$, and since $\|D'\lambda'\|_1 = 1$, equations 7.133 and 7.134 imply that

$$\|D'\lambda_i\|_1 = 1, \qquad \text{and} \qquad H_{J_2} D'\lambda_i = \mathbb{1}_{n_2}. \tag{7.135}$$

Also, since by homogeneity $B_{J_1}\lambda' = 0$, and, for $i \in \{1, 2\}$, $B_{J_1}\lambda_i \leq 0$, equation 7.134 implies that we must have

$$B_{J_1}\lambda_i = 0. \tag{7.136}$$

As in equation 7.132, equation 7.135 implies that, for each $i \in \{1, 2\}$, the entries of the rows of H_{J_2} coincide when restricted to the indices of columns corresponding to the support of $D'\lambda_i$, and these entries are equal to the signs of the corresponding elements in $D'\lambda_i$. That is, for $i \in \{1, 2\}$, and for any $j_1, j_2 \in [n_2]$ and

[47] Given a matrix $A \in \mathbb{R}^{m \times n}$ and an index set $J \subseteq [m]$ corresponding to rows of A, we let A_J denote the $|J| \times n$ matrix obtained by selecting only the rows of A indexed by J.

[48] Given a vector $x \in \mathbb{R}^n$, let $\mathrm{sign}(x)$ be the vector in $\mathbb{R}^n$ with i^{th} entry equal to 1 if $x_i > 0$, 0 if $x_i = 0$, and -1 if $x_i < 0$.

$k \in \mathrm{supp}(D'\lambda_i)$, we have

$$e_{n_2,j_1}^\top H_{J_2} e_{n,k} = e_{n_2,j_2}^\top H_{J_2} e_{n,k}, \quad \text{and} \quad e_{n_2,j_1}^\top H_{J_2} e_{n,k} = \mathrm{sign}(e_{n,k}^\top D'\lambda_i). \tag{7.137}$$

Since $\mathrm{supp}(D\lambda_i) = \mathrm{supp}(D'\lambda_i)$ and $\mathrm{sign}(D\lambda_i) = \mathrm{sign}(D'\lambda_i)$, equation 7.137 implies that, for $i \in \{1, 2\}$, we have

$$H_{J_2} D\lambda_i = \|D\lambda_i\|_1 \mathbb{1}_{n_2}. \tag{7.138}$$

As the matrix in equation 7.131 is invertible, equations 7.136 and 7.138 imply that $\lambda_i = \|D\lambda_i\|_1 \lambda$. thus λ_1, λ_2 and λ' are all positive scalar multiples of λ, with $\|D'\lambda_1\|_1 = \|D'\lambda_2\|_1 = \|D'\lambda'\|_1 = 1$, which implies that $\lambda' = \lambda_1 = \lambda_2$, and $\lambda' \in \mathrm{extr}(\mathcal{D}(B, D'))\backslash\{0\}$.

We have

$$\left|1 - \frac{1}{\|D'\lambda\|_1}\right| = \left|1 - \frac{\|D\lambda\|_1}{\|D'\lambda\|_1}\right| \le \frac{\|(D' - D)\lambda\|_1}{\|D'\lambda\|_1} \le (2/\underline{\sigma})\|D - D'\|$$

where we have used the fact that the smallest diagonal entry in the support of D' is bounded below by $2/\underline{\sigma}$ when $\|D - D'\| \le \underline{\sigma}/2$. This completes the proof of 7.129. □

7.1 Appendix B: Auxiliary lemmas

Lemma 7.7. *Let*

$$F, F_n : S \times \Delta \subset \mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}, \quad n \ge 1,$$

be a sequence of continuous functions defined on a neighborhood $S \times \Delta$ *of the compact set* $S_0 \times \Delta_0$*. Suppose that*

- F_n *converges uniformly to* F *on* $S \times \Delta$*,*
- $S_n \times \Delta_n \subset \mathbb{R}^d \times \mathbb{R}^p$ *are compact sets satisfying*

$$d_H(S_n, S_0) \to 0, \quad d_H(\Delta_n, \Delta_0) \to 0.$$

Then, the minimax values satisfy

$$\min_{\theta \in S_n} \max_{\lambda \in \Delta_n} F_n(\theta, \lambda) \to \min_{\theta \in S_0} \max_{\lambda \in \Delta_0} F(\theta, \lambda).$$

Proof of Lemma 7.7. For all sufficiently large n's, as S and Δ are neighborhoods of S_0 and Δ_0, we can assume for simplicity that $S_n \times \Delta_n \subset S \times \Delta$ for all $n \ge 1$, and that $S \times \Delta$ is compact. Let $\phi, \phi_n : S \to \mathbb{R}$, for $n \ge 1$, be defined by $\phi_n(\theta) = \max_{\lambda \in \Delta_n} F_n(\theta, \lambda)$ and $\phi(\theta) = \max_{\lambda \in \Delta_0} F(\theta, \lambda)$. Then by the theorem of the maximum, ϕ_n and ϕ are continuous.

Step 1We first claim that ϕ_n converges uniformly to ϕ on S. Indeed, as ϕ is continuous and S is compact, it suffices to show that $\lim_{n\to\infty} \phi_n(\theta_n) = \phi(\theta^*)$ whenever $\theta_n \to \theta^*$, $\theta_n, \theta^* \in S$. We establish the latter by showing that every subsequence of $\{\phi_n(\theta_n)\}_{n\ge1}$ has a further subsequence that converges to $\phi(\theta^*)$. For $\theta \in S$, let $\lambda_n(\theta)$ and $\lambda(\theta)$ be such that by $\lambda_n(\theta) \in \arg\max_{\lambda \in \Delta_n} F_n(\theta, \lambda)$ and $\lambda(\theta) \in \arg\max_{\lambda \in \Delta_0} F(\theta, \lambda)$.

Then $\phi_n(\theta_n) = F_n(\theta_n, \lambda_n(\theta_n))$. As $S \times \Delta$ is compact, every subsequence of $\{\phi_n(\theta_n)\}_{n\geq 1}$ has a further subsequence, say $\{\phi_{n_k}(\theta_{n_k})\}_{k\geq 1}$, such that $\lambda_{n_k}(\theta_{n_k}) \to \lambda^*$. By uniform convergence of F_n to F and the continuity of F, we have $\phi_{n_k}(\theta_{n_k}) = F_{n_k}(\theta_{n_k}, \lambda_{n_k}(\theta_{n_k})) \to F(\theta^*, \lambda^*)$. As $d_H(\Delta_n, \Delta_0) \to 0$, we have $\lambda^* \in \Delta_0$. We now show that $F(\theta^*, \lambda^*) = \phi(\theta^*)$. Indeed, for each $\lambda \in \Delta_0$, as $d_H(\Delta_{n_k}, \Delta_0) \to 0$ there exists $\lambda_k \in \Delta_{n_k}$ such that $\lambda_k \to \lambda$. Then such $\lambda \in \Delta_0$ we get

$$F(\theta^*, \lambda) = \lim_{n\to\infty} F_{n_k}(\theta_{n_k}, \lambda_k) \leq \lim_{n\to\infty} F_{n_k}(\theta_{n_k}, \lambda_{n_k}) = F(\theta^*, \lambda^*).$$

Thus $F(\theta^*, \lambda^*) = \phi(\theta^*)$, and we conclude that $\phi_n(\theta_n) \to \phi(\theta^*)$ and thus ϕ_n uniformly converges to ϕ on S.

Step 2 We now show that $\min_{\theta\in S_n} \phi_n(\theta) \to \min_{\theta\in S_0} \phi(\theta)$. Let $\phi_n(\theta_n^*) = \min_{\theta\in S_n} \phi_n(\theta)$, where $\theta_n^* \in S_n$. We establish the claim by showing that every subsequence of $\{\phi_n(\theta_n^*)\}$ has a further subsequence that converges to $\min_{\theta\in S_0} \phi(\theta)$. Indeed, given a subsequence of $\{\phi_n(\theta_n^*)\}$, consider a further subsequence, say $\{\phi_{n_k}(\theta_{n_k}^*)\}_{k\geq 1}$, such that $\theta_{n_k}^* \to \theta^*$. Then the preceding step implies that $\phi_{n_k}(\theta_{n_k}^*) \to \phi(\theta^*)$. Also, as $d_H(S_n, S_0) \to 0$, we have $\theta^* \in S_0$, and given any $\theta \in S_0$, there exists $\theta_k \in S_{n_k}$ such that $\theta_k \to \theta$. Hence for $\theta \in S_0$, we have

$$\phi(\theta) = \lim_{k\to\infty} \phi_{n_k}(\theta_k) \geq \lim_{k\to\infty} \phi_{n_k}(\theta_{n_k}^*) = \phi(\theta^*).$$

Thus $\phi(\theta^*) = \min_{\theta\in S_0} \phi(\theta)$ and we conclude that $\phi_n(\theta_n^*) \to \min_{\theta\in S_0} \phi(\theta)$.

□

The general form of the following lemma, stated with an implicit constant, is known as Hoffman's Theorem Hoffman (1952). Here, we present a proof that includes an explicit constant, as understanding its dependence on the inputs (A, b, c) is crucial for our uniformity results. Our proof modifies the argument in Hoffman (1952) slightly to obtain an explicit Lipschitz constant.

Lemma 7.8 (Hoffman's Theorem). *Consider the polyhedron $\mathcal{P} = \{z \in \mathbb{R}^n \mid A_E z = b_E,\ A_I z \leq b_I\}$, where $A_E \in \mathbb{R}^{m_E\times n}$ and $A_I \in \mathbb{R}^{m_I\times n}$. Then, for all $x \in \mathbb{R}^n$, we have*

$$d(x, \mathcal{P}) \leq C\Big(\|A_E x - b_E\| + \|(A_I x - b_I)_+\|\Big), \tag{7.139}$$

where the constant C can be chosen as $C = \big(\Lambda(A)\big)^{1/2}$, with $\Lambda(A)$ defined in Definition 4.26.

Here, $(x)_+$ denotes the positive part of x, obtained by replacing each entry x_i with $\max\{x_i, 0\}$, and $\|\cdot\|$ represents the Euclidean norm in the corresponding space.

Proof of Lemma 7.8. Suppose $x \notin \mathcal{P}$, and let y be the projection of x onto $\mathcal{P}$ with respect to the Euclidean norm. Define $J \subset I$ as the subset of rows of the matrix A_I that are active at y. The tangent cone of $\mathcal{P}$ at y is given by

$$\mathcal{C} = \{z \in \mathbb{R}^n \mid A_E z = 0,\ A_J z \leq 0\}.$$

Since y is the closest point in $\mathcal{P}$ to x, the first-order optimality condition (FOC) implies that $\langle x - y, z\rangle \leq 0$ for all $z \in \mathcal{C}$. By Farkas' Lemma, there exist vectors λ_E and λ_J with $\lambda_J \geq 0$ such that

$$x - y = A_E^T \lambda_E + A_J^T \lambda_J.$$

Let $\tilde{E} \subset E$ be the subset of indices corresponding to nonzero entries of λ_E, and define $\tilde{J} \subset J$ similarly for λ_J. By an argument similar to the one used in proving Carathéodory's convex hull theorem, we can choose λ_E and λ_J so that the set of rows of A_E and A_I corresponding to the indices in $F := \tilde{E} \cup \tilde{J}$, $\{a_i \mid i \in F\}$, is linearly independent. Set $\lambda_F = (\lambda'_{\tilde{E}}, \lambda'_{\tilde{J}})'$ and $A_F = (A'_{\tilde{E}}, A'_{\tilde{J}})'$. Then, we have

$$x - y = A_F^T \lambda_F,$$

where the rows of A_F are linearly independent.[49]

Since $x \notin \mathcal{P}$, it follows that $x - y \neq 0$ and $x - y \notin \mathcal{C}$ (otherwise, the inequality $\langle x - y, z\rangle \leq 0$ for all $z \in \mathcal{C}$ would imply $\|x - y\| = 0$, a contradiction).

Define the polyhedral cone K by

$$K = \{z \in \mathbb{R}^{|F|} \mid z_i = 0 \ \forall i \in \tilde{E},\ z_j \leq 0 \ \forall j \in \tilde{J}\}.$$

Setting $w = x - y$, our goal is to bound $\|w\|$ in terms of $d(A_F w, K)$. Let Π_K denote the projection matrix onto K, and define

$$\delta = A_F w - \eta, \quad \text{where} \quad \eta = \Pi_K(A_F w).$$

We then obtain

$$\lambda_F^T \delta = \lambda_F^T A_F w - \lambda_F^T \eta = \lambda_F^T A_F A_F^T \lambda_F - \lambda_F^T \eta.$$

Since $\lambda_{\tilde{J}} \geq 0$ and by the definition of K, we have $\lambda_F^T \eta \leq 0$, leading to

$$\|A_F^T \lambda_F\|^2 \leq |\lambda_F^T \delta| \Rightarrow (\lambda_{\min}(A_F A_F^T))^{1/2} \|\lambda_F\| \|A_F^T \lambda_F\| \leq \|\lambda_F\| \|\delta\|. \tag{7.140}$$

Using the definition of K, it can be easily verified that $\eta = (0', -(A_{\tilde{J}} w)'_-)'$. Consequently, we obtain

$$\delta = \left((A_{\tilde{E}} w)', (A_{\tilde{J}} w)'_+\right)',$$

and $\|\delta\|^2 = \|A_{\tilde{E}} x - b_{\tilde{E}}\|^2 + \|(A_{\tilde{J}} x - b_{\tilde{J}})_+\|^2$. Since

$$\|A_{\tilde{E}} x - b_{\tilde{E}}\|^2 + \|(A_{\tilde{J}} x - b_{\tilde{J}})_+\|^2 \leq \|A_E x - b_E\|^2 + \|(A_I x - b_I)_+\|^2,$$

combining this inequality with the final implication in (7.140), we conclude that

$$d(x, \mathcal{P}) \leq \frac{1}{\sqrt{\lambda_{\min}(A_F A_F^T)}} \Big(\|A_E x - b_E\| + \|(A_I x - b_I)_+\| \Big),$$

which proves the lemma. □

The following lemma is based on Lemma 2 in Robinson (1977) and establishes that the feasible region $\mathcal{P}$ is Lipschitz upper hemicontinuous with respect to its inputs if $\mathcal{P}$ is bounded, and locally Lipschitz upper

[49] Alternatively, we can let F be the support of a vertex solution of the dual to the linear program $\max\{\langle x - y, z\rangle \mid A_J z \leq 0,\ A_E z = 0\}$. Note that by assumption, the latter LP has value zero.

hemicontinuous otherwise. Throughout this lemma, let $\Lambda(A)$ be as in Definition 4.26, and define $b = (b_E', b_I')'$, $A = (A_E', A_I')'$.

Lemma 7.9 (Lipschitz Upper Hemicontinuity). *Consider the polyhedron*

$$\mathcal{P} = \{z \in \mathbb{R}^n \mid A_E z = b_E,\ A_I z \le b_I\},$$

and let

$$\tilde{\mathcal{P}} = \{z \in \mathbb{R}^n \mid \tilde{A}_E z = \tilde{b}_E,\ \tilde{A}_I z \le \tilde{b}_I\}$$

denote a perturbed version of $\mathcal{P}$, obtained by replacing its input parameters (A, b) with $(\tilde{A}, \tilde{b})$.

If $\tilde{\mathcal{P}}$ is nonempty, then for all $R > 0$, we have [50]

$$\sup_{x \in \tilde{\mathcal{P}},\ \|x\| \le R} d(x, \mathcal{P}) \le 2(R+1)\Lambda(A)^{-1/2}\Big(\|A - \tilde{A}\| + \|b - \tilde{b}\|\Big). \tag{7.141}$$

Moreover, if $\mathcal{P}$ is bounded by some $M > 0$, i.e.,

$$\max_{x \in \mathcal{P}} \|x\| \le M,$$

and the perturbation satisfies

$$\|b - \tilde{b}\| \vee \|A - \tilde{A}\| \le [\Lambda(A)]^{1/2}/4,$$

then the set $\tilde{\mathcal{P}}$, if nonempty, remains bounded, with

$$\sup_{x \in \tilde{\mathcal{P}}} \|x\| \le 2M + 1,$$

and the deviation of points in $\tilde{\mathcal{P}}$ from $\mathcal{P}$ is bounded by

$$\sup_{x \in \tilde{\mathcal{P}}} d(x, \mathcal{P}) \le 2(2M+1)\Lambda(A)^{-1/2}\Big(\|A - \tilde{A}\| + \|b - \tilde{b}\|\Big). \tag{7.142}$$

Proof of Lemma 7.9. Let $x \in \tilde{\mathcal{P}}$. By Lemma 7.8, we have

$$d(x, \mathcal{P}) \le (\Lambda(A))^{-1/2}\Big(\|A_E x - b_E\| + \|(A_I x - b_I)_+\|\Big). \tag{7.143}$$

Since $x \in \tilde{\mathcal{P}}$, it follows that

$$A_E x - b_E = (A_E - \tilde{A}_E)x + (\tilde{b}_E - b_E)$$

and

$$A_I x - b_I \le (A_I - \tilde{A}_I)x + (\tilde{b}_I - b_I).$$

Substituting these expressions into inequality (7.143), we obtain

$$d(x, \mathcal{P}) \le 2(\Lambda(A))^{-1/2}\Big(\|A - \tilde{A}\|\|x\| + \|b - \tilde{b}\|\Big) \tag{7.144}$$

[50] By convention, the supremum over an empty set is equal to $-\infty$.

which readily yields inequality 7.141. Next, under the assumption that $\sup\{\|x\| \mid x \in \mathcal{P}\} \le M$, we have $\|x\| \le M + d(x, \mathcal{P})$ for any x, which combined with inequality 7.144 yields

$$\|x\|(1 - 2\Lambda(A)^{-1/2}\|A - \tilde{A}\|) \le 2\Lambda(A)^{-1/2}\|b - \tilde{b}\| + M.$$

Since the perturbation satisfies

$$\|b - \tilde{b}\| \vee \|A - \tilde{A}\| \le [\Lambda(A)]^{1/2}/4,$$

we conclude that $\|x\| \le 2M + 1$. Therefore, $\tilde{\mathcal{P}}$ remains bounded, and we have

$$\sup_{x \in \tilde{\mathcal{P}}} \|x\| \le 2M + 1.$$

Finally, substituting this bound into inequality (7.144) establishes the desired result, completing the proof.

□

The following lemma is a variation on the inequalities in Theorem 1 of Robinson (1975), adapted to our simpler setting with explicit constants. The explicit dependence of these constants on the inputs is crucial for our uniformity results.

Lemma 7.10 (Lipschitz Lower Hemicontinuity). *Let $\mathcal{P} = \{z \in \mathbb{R}^n \mid A_E z = b_E,\ A_I z \le b_I\}$ be a polyhedron, and assume that $\|A_I\| \le M'$ for some $M' > 0$. Assume A_E has full row rank and that there exists $x_0 \in \mathcal{P}$ and $\delta_0 > 0$ such that $A_I x_0 + \delta_0 \mathbb{1} \le b_I$. Let $\tilde{\mathcal{P}}$ be a perturbed version of $\mathcal{P}$, and let $M > \|x_0\|$ be a fixed constant. If the perturbation satisfies*

$$\|A - \tilde{A}\| \vee \|b - \tilde{b}\| < \min\left\{\frac{\lambda_{\min}(A_E A_E^T)^{1/2}}{2}, M', \delta_0\left[\frac{1 \wedge \lambda_{\min}(A_E A_E^T)^{1/2}}{3 + 3M + 24(M+1)M'}\right]\right\}, \tag{7.145}$$

then $\tilde{\mathcal{P}} \neq \emptyset$, and we have

$$\sup_{x \in \mathcal{P},\ \|x\| \le M} d(x, \tilde{\mathcal{P}}) \le \left\{\frac{2M(3 + 3M + 24(M+1)M')}{\delta_0(1 \wedge \lambda_{\min}(A_E A_E^T)^{1/2})} + \frac{2(M+1)}{\sqrt{\lambda_{\min}(A_E A_E^T)}}\right\}\Big(\|b - \tilde{b}\| + \|A - \tilde{A}\|\Big). \tag{7.146}$$

Proof of Lemma 7.10. We proceed in two steps. First, we show that any point in the relative interior of $\mathcal{P}$ remains close to $\tilde{\mathcal{P}}$ under small perturbations. Second, we extend this result to all points in $\mathcal{P}$.

Step 1: Suppose $x \in \mathcal{P}$ is such that $\|x\| \le M$ and $\delta \mathbb{1} + A_I x \le b_I$, for some $\delta_0 \ge \delta > 0$. In addition, assume that condition 7.145 holds with δ_0 replaced by δ:

$$\|A - \tilde{A}\| \vee \|b - \tilde{b}\| \le \min\left\{\frac{\lambda_{\min}(A_E A_E^T)^{1/2}}{2}, M', \delta\left[\frac{1 \wedge \lambda_{\min}(A_E A_E^T)^{1/2}}{3 + 3M + 24(M+1)M'}\right]\right\}. \tag{7.147}$$

Note that such points x exist, as x_0 is a candidate.

By Weyl's inequality,

$$\left|\lambda_{\min}(A_E A_E^T)^{1/2} - \lambda_{\min}(\tilde{A}_E \tilde{A}_E^T)^{1/2}\right| \le \|A_E - \tilde{A}_E\|.$$

As condition 7.147 ensures $\|A_E - \tilde{A}_E\| \le \lambda_{\min}(A_E A_E^T)^{1/2}/2$, we have $\lambda_{\min}(\tilde{A}_E \tilde{A}_E^T)^{1/2} \ge \lambda_{\min}(A_E A_E^T)^{1/2}/2$, implying $\tilde{A}_E$ has full row rank. The equation $\tilde{A}_E(x+h) = \tilde{b}_E$, rewritten as

$$\tilde{A}_E h = (\tilde{b}_E - b_E) - (\tilde{A}_E - A_E)x,$$

has a solution h satisfying

$$\|h\| \le \frac{2}{\sqrt{\lambda_{\min}(A_E A_E^T)}}\big(\|b_E - \tilde{b}_E\| + M\|A_E - \tilde{A}_E\|\big).$$

We now show, using condition 7.147, that $\tilde{A}_I(x+h) \le \tilde{b}_I$, so $x + h \in \tilde{\mathcal{P}}$, yielding

$$d(x, \tilde{\mathcal{P}}) \le \|h\| \le \frac{2(M+1)}{\sqrt{\lambda_{\min}(A_E A_E^T)}}\big(\|b_E - \tilde{b}_E\| + \|A_E - \tilde{A}_E\|\big). \tag{7.148}$$

Since

$$\tilde{A}_I(x+h) \le \tilde{b}_I + (b_I - \tilde{b}_I) + (\tilde{A}_I - A_I)x + \tilde{A}_I h - \delta \mathbb{1},$$

and condition 7.147 ensures

$$\|(\tilde{A}_I - A_I)x\| \le \delta/3, \quad \|\tilde{A}_I h\| \le \delta/3, \quad \|\tilde{b}_I - b_I\| \le \delta/3,$$

it follows that

$$\tilde{A}_I(x+h) \le \tilde{b}_I,$$

thus proving (7.148).

Step 2: Let $x \in \mathcal{P} \cap \{x \mid \|x\| \le M\}$ be arbitrary. For $t \in (0,1)$, define

$$x_t = tx + (1-t)x_0, \quad \delta_t = (1-t)\delta_0,$$

where x_0 and δ_0 are as in the statement of the lemma. Then,

$$\sup_{x \in \mathcal{P},\ \|x\| \le M} \|x - x_t\| \le 2M(1-t), \quad A_I x_t + \delta_t \mathbb{1} \le b_I.$$

Thus,

$$\begin{aligned}
\sup_{x \in \mathcal{P},\ \|x\| \le M} d(x, \tilde{\mathcal{P}}) &\le \sup_{x \in \mathcal{P},\ \|x\| \le M} d(x, x_t) + \sup_{x \in \mathcal{P},\ \|x\| \le M} d(x_t, \tilde{\mathcal{P}}) \\
&\le 2M(1-t) + \sup_{x \in \mathcal{P},\ \|x\| \le M} d(x_t, \tilde{\mathcal{P}}).
\end{aligned}$$

If condition 7.145 holds, there exists $t^* \in [0,1)$ such that

$$\|A - \tilde{A}\| \vee \|b - \tilde{b}\| = \delta_{t^*}\Big[\frac{1 \wedge \lambda_{\min}(A_E A_E^T)^{1/2}}{3 + 3M + 24(M+1)M'}\Big]. \tag{7.149}$$

Setting $\delta = \delta_{t^*}$ in 7.147, inequality 7.148 gives

$$\sup_{x\in\mathcal{P},\ \|x\|\le M} d(x_t,\tilde{\mathcal{P}}) \le \frac{2(M+1)}{\sqrt{\lambda_{\min}(A_E A_E^T)}}\left(\|b_E - \tilde{b}_E\| + \|A_E - \tilde{A}_E\|\right). \tag{7.150}$$

Thus,

$$\sup_{x\in\mathcal{P},\ \|x\|\le M} d(x,\tilde{\mathcal{P}}) \le 2M(1-t^*) + \frac{2(M+1)}{\sqrt{\lambda_{\min}(A_E A_E^T)}}\left(\|b_E - \tilde{b}_E\| + \|A_E - \tilde{A}_E\|\right).$$

Substituting $(1-t^*)$ using 7.149 yields

$$\sup_{x\in\mathcal{P},\ \|x\|\le M} d(x,\tilde{\mathcal{P}}) \le \left\{\frac{2M(3+3M+24(M+1)M')}{\delta_0(1\wedge\lambda_{\min}(A_E A_E^T)^{1/2})} + \frac{2(M+1)}{\sqrt{\lambda_{\min}(A_E A_E^T)}}\right\}\left(\|b-\tilde{b}\| + \|A-\tilde{A}\|\right),$$

which completes the proof. □

The following lemma is based on parts (a) and (b) of Theorem 1 in Robinson (1977) and provides an alternative characterization of regularity. This alternative characterization will be used in parts to show how to regulararize irregular LPs.

Lemma 7.11 (Dual Characterization of Regularity). *Suppose that the primal linear program* $\min\{c^T x \mid A_E x = b_E,\ A_I x \le b_I\}$ *is feasible and with finite value. Then,* $\mathcal{P} = \{x \in \mathbb{R}^d \mid A_E x = b_E,\ A_I x \le b_I\}$ *satisfies the MFCQ if and only if the set*

$$\Delta_0 = \arg\max_{\lambda}\left\{\lambda_E^T b_E + \lambda_I^T b_I \mid \lambda_I \le 0,\ A_E^T\lambda_E + A_I^T\lambda_I = c\right\}$$

is bounded.

Proof of Lemma 7.11. Necessity Suppose, for the sake of contradiction, that $\mathcal{P}$ is regular and Δ_0 is unbounded. Since the MFCQ holds, there exists a point x_0 satisfying $A_E x_0 = b_E$ and $A_I x_0 < b_I$. The unboundedness of Δ_0 implies the existence of a nonzero vector $\lambda = (\lambda_E', \lambda_I')'$ such that

$$\lambda_I \le 0, \qquad A_E^T\lambda_E + A_I^T\lambda_I = 0, \qquad b_I^T\lambda_I + b_E^T\lambda_E = 0. \tag{7.151}$$

Since the MFCQ ensures that A_E has full row rank, we cannot have $\lambda_I = 0$ in (7.151). Consequently, the choice of λ in (7.151) leads to

$$\langle\lambda_E, A_E x_0\rangle + \langle\lambda_I, A_I x_0\rangle > \langle\lambda_E, b_E\rangle + \langle\lambda_I, b_I\rangle.$$

This results in $0 > 0$, a contradiction.

Sufficiency Suppose, for the sake of contradiction, that Δ_0 is bounded but the MFCQ does not hold. The failure of MFCQ implies the existence of a sequence of perturbations $b_n \to b$ (where only the b-vector is perturbed) such that

$$\mathcal{P}_n = \{x \in \mathbb{R}^d \mid A_E x = b_{n,E},\ A_I x \le b_{n,I}\} = \emptyset. \tag{7.152}$$

By assumption, since the value of the unperturbed program is finite, the set

$$\{\lambda \mid \lambda_I \le 0,\ A_E^T\lambda_E + A_I^T\lambda_I = c\}$$

is nonempty. Moreover, (7.152) implies that the corresponding perturbed dual programs are unbounded:

$$\max_{\lambda}\Big\{\lambda_E^T b_{n,E} + \lambda_I^T b_{n,I} \mid \lambda_I \le 0,\ A_E^T\lambda_E + A_I^T\lambda_I = c\Big\} = +\infty, \quad \forall n \ge 1.$$

This implies the existence of an unbounded sequence $\{\lambda_n\}$ such that

$$\lambda_n^T b_n \to +\infty, \qquad \lambda_{n,I} \le 0, \qquad A_E^T\lambda_{n,E} + A_I^T\lambda_{n,I} = c.$$

Dividing through by $\|\lambda_n\|$ and considering a limit $\lambda^* \ne 0$ of a convergent subsequence of $\{\lambda_n/\|\lambda_n\|\}_{n\ge1}$, we obtain

$$\langle\lambda^*, b\rangle \ge 0, \qquad \lambda_I^* \le 0, \qquad A_E^T\lambda_E^* + A_I^T\lambda_I^* = 0.$$

However, if $\lambda \in \Delta_0$, then for any $t > 0$, the vector $\lambda + t\lambda^*$ also belongs to Δ_0, contradicting the assumption that Δ_0 is bounded. This contradiction completes the proof. □

References

Andrews, Donald W. K. (2000). "Inconsistency of the Bootstrap when a Parameter Is on the Boundary of the Parameter Space". In: *Econometrica* 68.2, pp. 399–406.

Andrews, Donald W. K. and Gustavo Soares (2010). "Inference for Parameters defined by Moment Inequalities Using Generalized Moment Selection". In: *Econometrica* 78.1, pp. 119–157.

Andrews, Isaiah, Jonathan Roth, and Ariel Pakes (2023). "Inference for Linear Conditional Moment Inequalities". In: *The Review of Economic Studies* 90.6, pp. 2763–2791.

Bai, Yuehao, Andres Santos, and Azeem Shaikh (2022). *On testing systems of linear inequalities with known coefficients*. Working papers.

Brinch, Christian N., Magne Mogstad, and Matthew Wiswall (2017). "Beyond LATE with a Discrete Instrument". In: *Journal of Political Economy* 125.4, pp. 985–1039.

Bugni, Federico A., Ivan A. Canay, and Xiaoxia Shi (2017). "Inference for subvectors and other functions of partially identified parameters in moment inequality models". In: *Quantitative Economics* 8.1, pp. 1–38.

Cho, JoonHwan and Thomas M. Russell (2024). "Simple Inference on Functionals of Set-Identified Parameters Defined by Linear Moments". In: *Journal of Business & Economic Statistics* 42.2, pp. 563–578.

Cox, Gregory and Xiaoxia Shi (2023). "Simple Adaptive Size-Exact Testing for Full-Vector and Subvector Inference in Moment Inequality Models". In: *The Review of Economic Studies* 90.1, pp. 201–228.

Dupas, Pascaline (2014). "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence From a Field Experiment". In: *Econometrica* 82.1, pp. 197–228.

Fang, Zheng and Andres Santos (2019). "Inference on Directionally Differentiable Functions". In: *The Review of Economic Studies* 86.1, pp. 377–412.

Fang, Zheng et al. (2023). "Inference for Large-Scale Linear Systems With Known Coefficients". In: *Econometrica* 91.1, pp. 299–327.

Gafarov, Bulat (2025). "Simple subvector inference on sharp identified set in affine models". In: *Journal of Econometrics* 249.

Gol'shtein, Yegeniy (1972). *Theory of Convex Programming*. American Mathematical Society.

Hoffman, Alan (1952). "On Approximate Solutions of Systems of Linear Inequalities". In: *Journal of Research of the National Bureau of Standards* 49.4, pp. 263–265.

Imbens, Guido W. and Joshua D. Angrist (1994). "Identification and Estimation of Local Average Treatment Effects". In: *Econometrica* 62.2, pp. 467–475.

Lehman, E. L. and J. P. Romano (2005). *Testing Statistical Hypotheses*. Springer.

Mogstad, Magne, Andres Santos, and Alexander Torgovitsky (2017). *Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters*. Working Paper 23568. National Bureau of Economic Research.

— (2018). "Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters". In: *Econometrica* 86.5, pp. 1589–1619.

Robinson, Stephen (1975). "Stability Theory for Systems of Inequalities. Part I: Linear Systems". In: *Siam J. Numer. Anal.* 12, pp. 754–769.

— (1977). "A Characterization of Stability in Linear Programming". In: *Operations Research* 25, pp. 435–447.

Robinson, Stephen M. (1973). "Bounds for error in the solution set of a perturbed linear program". In: *Linear Algebra and its Applications* 6, pp. 69–81.

Romano, Joseph and Azeem Shaikh (2012). "On the Uniform Asymptotic Validity of Subsampling and the Bootstrap". In: *Ann. Stat.* 40.6, pp. 2798–2822.

Schrijver, Alexander (1999). *Theory of linear and integer programming*. John Wiley & Sons.

Shapiro, Alexander (1991). "Asymptotic analysis of stochastic programs". In: *Annals of Operations Research* 30, pp. 169–196.

Tsirel'son, V. S. (1975). "The density of the distribution of the maximum of a Gaussian random process". In: *Theory of Probability and its Applications* 16, pp. 847–856.

Voronin, Andrei (2025). *Linear Programming Approach to Partially Identified Econometric Models*. Working papers. Arxiv.

Vytlacil, Edward (2002). "Independence, Monotonicity, and Latent Index Models: An Equivalence Result". In: *Econometrica* 70.1, pp. 331–341.