Econometric Theory, **38**, 2022, 596–619.
doi:10.1017/S0266466621000207

CONSTRAINT QUALIFICATIONS IN PARTIAL IDENTIFICATION

HIROAKI KAIDO
Boston University

FRANCESCA MOLINARI
Cornell University

JÖRG STOYE
Cornell University

The literature on stochastic programming typically restricts attention to problems that fulfill constraint qualifications. The literature on estimation and inference under partial identification frequently restricts the geometry of identified sets with diverse high-level assumptions. These superficially appear to be different approaches to closely related problems. We extensively analyze their relation. Among other things, we show that for partial identification through pure moment inequalities, numerous assumptions from the literature essentially coincide with the Mangasarian–Fromowitz constraint qualification. This clarifies the relation between well-known contributions, including within econometrics, and elucidates stringency, as well as ease of verification, of some high-level assumptions in seminal papers.

1. INTRODUCTION

This paper connects two related but largely separate literatures, namely statistical analysis of stochastic programs and estimation and inference for (functions of) partially identified parameters. The bounds that are pervasive in the latter literature are often expressed as values of constrained optimization problems. As such, some similarity to stochastic programming is rather apparent and has been observed before. However, we uncover much deeper connections between these literatures.

Our discussion starts from the econometrics literature. In a seminal paper, Chernozhukov, Hong, and Tamer (2007, CHT henceforth) provide a comprehensive analysis of consistency of criterion function-based set estimators and their convergence rates in partially identified models. Their work highlights the challenges a researcher faces in this context and puts forward possible solutions

We are grateful to Ivan Canay for conversations that helped bring this paper into focus and to three anonymous referees as well as the editor and co-editor for extremely careful readings of the manuscript. We also thank Isaiah Andrews, Lixiong Li, and seminar audiences at the Bristol Econometrics Study Group, Bristol/Warwick joint seminar, Columbia, and Duke for their feedback. Any and all errors are our own. We gratefully acknowledge financial support through NSF Grants SES-1824344 and SES-2018498 (Kaido) as well as SES-1824375 (Molinari and Stoye). Address correspondence to Jörg Stoye, Department of Economics, Cornell University, Ithaca, NY, USA; e-mail: stoye@cornell.edu.

© The Author(s), 2021. Published by Cambridge University Press

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

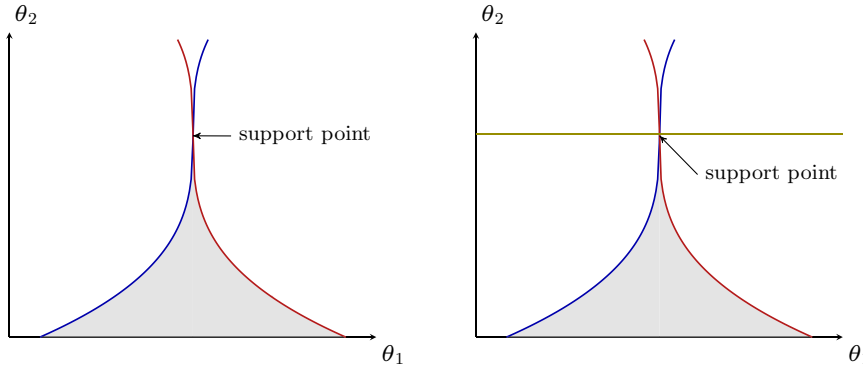


FIGURE 1. Two examples of irregular support points. Each curve represents an inequality constraint that excludes everything above it, so that Θ_I is the shaded region. Estimation of Θ_I or its support function in direction $(0, 1)$ ("up") will be difficult. Note that individual constraints may be well-behaved. The right panel differs from the left one by the addition of a constraint which ensures that a polynomial minorant condition holds. This illustrates that such a condition does not rule out the problem.

in the form of assumptions under which specific rates of convergence attain. While these assumptions can be dispensed with when the researcher's goal is to obtain a confidence set for the partially identified parameter vector that is—pointwise or uniformly—consistent in level (e.g., Andrews and Soares, 2010), related assumptions reappear when the aim is to obtain a confidence interval for a smooth function of the partially identified parameter vector that is—pointwise or uniformly—consistent in level (e.g., Pakes et al., 2011, PPHI henceforth Bugni, Canay, and Shi, 2017, BCS henceforth).[1] Some more recent contributions (Cho and Russell, 2019; Gafarov, 2019) observe a connection to stochastic programming and show that inference becomes much more tractable under the so-called Linear Independence constraint qualification. Some obvious questions arise: How do all these assumptions relate? What is the trade-off between them and possibly other assumptions from the statistics literature?

For a sense of why consistent estimation of identified sets or their projections can be hard even in otherwise well-behaved moment inequality settings, consider Figure 1. Both panels illustrate a detail of an *identified set* Θ_I defined as the collection of parameter values satisfying a finite number of moment inequalities (as in (2.1) below, but without equality constraints). Each restriction is represented by a curve and excludes everything above that curve. The resulting Θ_I is shaded.[2]

[1] We cite PPHI because the published version (Pakes et al., 2015) does not contain the inference procedure. However, this procedure has been used in influential papers (Eizenberg, 2014; Ho and Pakes, 2014; Holmes, 2011).

[2] An exact algebraic description of the example in Figure 1 is as follows. Let $\Theta = \mathbb{R}^2$ with typical element $\theta = (\theta_1, \theta_2)$ and let $\Theta_I = \{\theta : (\theta_2 - 1)^3 + \theta_1 E(X_1) \leq 0, (\theta_2 - 1)^3 - \theta_1 E(X_1) \leq 0\}$ for the left panel; for the right panel, additionally require $\theta_2 \leq E(X_3)$, where $E(X_1) = E(X_2) = E(X_3) = 1$. Looking ahead, the right panel of Figure 2 and both panels of Figure 4 are qualitative representations.

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

Suppose now that a researcher wants to find either (i) a set estimator of Θ_I that is consistent in Hausdorff distance (defined later) or (ii) an estimator of a linear projection of Θ_I (e.g., the identified set for a component of θ), represented through the support function

$$s(p, \Theta_I) \equiv \max\{p'\theta : \theta \in \Theta_I\} \tag{1.1}$$

in predetermined direction p. (In the figure, $p = (0, 1)$, i.e., we maximize θ_2.) Even if all constraints' graphs can be estimated at a specific—e.g., parametric—rate, it does not follow that their intersection estimates either object of interest at the same or indeed at any rate. For example, if estimators approximate the true constraints from below, the support function may be underestimated including in the limit.

The examples may appear "knife-edge." However, note that: (i) While we will, in this paper, take a pointwise perspective to simplify the analysis, the literature on partial identification is usually concerned with inference that is uniformly valid *near* such irregular cases because asymptotic approximations may otherwise be misleading. Indeed, this is emphasized in the abstract of Canay and Shaikh (2017). (ii) Inference methods that are uniformly valid typically use "Generalized Moment Selection" methods (see again Canay and Shaikh, 2017 for details) that account for statistical uncertainty not only of moment conditions that are violated in sample, but also of ones that are local-to-binding. In these methods, the "overidentified" feature of the right-hand panel, i.e., the intersection of more than d constraints at one point in $\mathbb{R}^d$, becomes typical of bootstrap d.g.p.'s. In addition, this feature characterizes the boundary case of overidentifying moment conditions, though some assumptions discussed below will exclude that case anyway.

A reader familiar with constraint qualifications in optimization problems will recognize that both panels of Figure 1 violate some of these qualifications. Conversely, a reader who is very familiar with the partial identification literature might recognize that they violate assumptions in CHT, PPHI, and elsewhere. We ask if this reflects deeper relations between these bodies of literature. The answer will be affirmative: Under a background assumption of continuous differentiability of moment conditions and abstracting from details of "how uniformly" the assumptions are stated, the literature on partial identification already invokes constraint qualifications; for examples, we show that both papers just cited rely on the Mangasarian–Fromowitz constraint qualification. This implies some previously unrecognized (to the best of our knowledge) logical relations between assumptions made in econometrics.

Some references to the literatures that we connect are as follows. Molinari (2020) gives a current overview over the field of partial identification. Canay and Shaikh (2017) provide a definitive treatment of the literature on moment inequalities. We will define constraint qualifications below but refer to Bazaraa, Sherali, and Shetty (2006) for a textbook treatment and to Bonnans and Shapiro (2000) for a textbook on perturbation analysis of stochastic programs.

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

TABLE 1. Assumptions and their relation as established in Lemma 3.1. The term "criterion" refers to $\max\left\{0, \max_{j=1,\ldots,J_1} E(m_j(X,\theta))/\sigma_j(\theta), \max_{j=J_1+1,\ldots,J} |E(m_j(X,\theta))/\sigma_j(\theta)|\right\}$.

Assumption	Intuition
1	Background assumption.
2	All moments strictly negative at some $\tilde{\theta}$ close to θ^*.
3 $\Leftrightarrow$ 3'	Criterion increases with distance from Θ_I.
4 $\Leftrightarrow$ 4'	Criterion increases with distance from support set on support plane.
5 $\Leftrightarrow$ 5'	Θ_I is locally contained in a pointy cone.
6 $\Leftrightarrow$ 6'	There exists a direction from θ^* in which every constraint becomes slack.
7 $\Leftrightarrow$ 7'	Criterion increases outside a pointy cone locally containing Θ_I.
8	Linear Independence Constraint Qualification (LICQ).
9	Mangasarian–Fromowitz Constraint Qualification (MFCQ).
10	Abadie Constraint Qualification (ACQ).

2. ASSUMPTIONS

This section clarifies the general setup and introduces a broad array of assumptions from the literature. For the reader's convenience, these assumptions, and some results that we will explain later, are summarized in Table 1.

We assume throughout that the model is correctly specified and that individual moment conditions are well-behaved. Specifically, the identified set Θ_I is characterized by J constraints, namely $J_1 \leq J$ moment inequalities and $J_2 \equiv J - J_1$ moment equalities:

$$\Theta_I \equiv \{\theta \in \Theta : E(m_j(X,\theta)) \leq 0,\ j = 1,\ldots,J_1; E(m_j(X,\theta)) = 0,\ j = J_1+1,\ldots,J\}, \tag{2.1}$$

where the functions $(m_1(\cdot),\ldots,m_J(\cdot))$ are known up to $\theta \in \Theta$. Then we impose:

Assumption 1. (a) $\Theta \subset \mathbb{R}^d$ is compact convex with nonempty interior.
(b) $\Theta_I \neq \emptyset$ and $\Theta_I \subset \text{int}(\Theta)$.
(c) $\sigma_j^2(\theta) \equiv \text{Var}(m_j(X,\theta)) \in (0,\infty)$ for $j = 1,\ldots,J$.
(d) The gradients $D_j(\cdot) \equiv \nabla_\theta\{E[m_j(X,\cdot)]/\sigma_j(\cdot)\}, j = 1,\ldots,J$ exist and are continuous.

These assumptions are standard in the literature. The requirement that $\Theta_I \subset \text{int}(\Theta)$ may appear stronger than the comparable one in CHT, i.e., their condition M2. However, the latter imposes continuous differentiability of moment conditions on a small enlargement of Θ, so it does constrain Θ_I to be interior to the set on which local linear approximation of moment conditions is valid. This is how we use the assumption, and we could analogously weaken it.

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

We next state numerous assumptions that are inspired by the aforementioned literature in econometrics. We first state them in a way that maximizes resemblance to the original formulation, subject to the unification that assumptions are stated pointwise (not uniformly) over d.g.p.'s and that their local implications near extreme points of Θ_I are extracted. Universal constants invoked in assumptions need not take the same value across appearances.

Define the support set of Θ_I as

$$S(p, \Theta_I) \equiv \{\theta \in \Theta_I : p'\theta = s(p, \Theta_I)\},$$

where the support function $s(\cdot)$ is defined in (1.1), and the supporting hyperplane as

$$H(p, \Theta_I) \equiv \{\theta \in \Theta : p'\theta = s(p, \Theta_I)\}.$$

Any element of $S(p, \Theta_I)$ is also called a support point. We use θ^* to denote a generic support point; to economize on subscripts and because we consider p fixed, we suppress dependence of θ^* on p. Recall also that a constraint is *active* at θ if $E(m_j(X, \theta)) = 0$. Let

$$\mathcal{J}_1^*(\theta) \equiv \left\{j \in \{1, \ldots, J_1\} : E(m_j(X, \theta)) = 0\right\} \tag{2.2}$$

denote the set of active inequality constraints and

$$\mathcal{J}^*(\theta) \equiv \left\{j : E(m_j(X, \theta)) = 0\right\} = \mathcal{J}_1^*(\theta) \cup \{J_1 + 1, \ldots, J\} \tag{2.3}$$

the full *active set* of (equality or inequality) constraints at θ.

We first adapt two assumptions, "degeneracy" and "polynomial minorant," from CHT. These are essential for getting rate results for consistency of analog estimators of Θ_I; in particular, Polynomial Minorant ensures rate results for a relaxed sample analog of Θ_I, and degeneracy allows one to drop the relaxation. We weaken the assumptions insofar as they are only imposed at support points. Also, we do not adapt the high-level Conditions C.2 and C.3 from CHT because, being about sample objects, they restrict the sampling process and not just population moments. To keep these issues separate, we focus on the sufficient conditions that stand in for the assumptions in a moment inequalities setting (i.e., CHT's displays 4.5 and 4.6).

Assumption 2. For each support point $\theta^* \in S(p, \Theta_I)$, there exist constants $\delta > 0$, $\eta > 0$, $M > 0$, and $C > 0$ s.t.

$$\max_{j=1,\ldots,J} E(m_j(X, \theta))/\sigma_j(\theta) \leq -C\varepsilon \quad \text{for all } \theta \in \Theta_I^{-\varepsilon} \cap B(\theta^*, \eta) \tag{2.4}$$

$$\max_{\theta \in \Theta_I \cap B(\theta^*, \eta)} d(\theta, \Theta_I^{-\varepsilon}) \leq M\varepsilon \quad \text{for all } \varepsilon \in [0, \delta], \tag{2.5}$$

where $\Theta_I^{-\varepsilon} = \{\theta \in \Theta_I : d(\theta, \Theta \setminus \Theta_I) \geq \varepsilon\}$ and $B(\theta^*, \eta) \equiv \{\theta \in \Theta : \|\theta - \theta^*\| \leq \eta\}$.

This "degeneracy" assumption ensures that for any support point θ^*, there exists a nearby point where moment inequalities hold with slack and whose membership in Θ_I is therefore easy to determine. In particular, we show in the proof of Theorem 3.1 that Assumption 2 precludes the existence of equality constraints.

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

In the original version, the equivalent of (2.5) was stated using Hausdorff distance, i.e., $d_H(\Theta_I, \Theta_I^{-\varepsilon}) \leq M\varepsilon$, where $d_H(A,B) = \max\{\max_{a\in A} d(a,B), \max_{b\in B} d(b,A)\}$ for generic sets A,B with typical elements a,b. However, $\max_{\theta\in\Theta_I^{-\varepsilon}} d(\theta,\Theta_I) = 0$ because $\Theta_I^{-\varepsilon} \subset \Theta_I$, so here and in the original, only the implied restriction on $\max_{\theta\in\Theta_I} d(\theta,\Theta_I^{-\varepsilon})$ is nonvacuous. We localize it by restricting θ to a neighborhood of the support point.

Assumption 3. For each support point $\theta^* \in S(p,\Theta_I)$, there exist constants $\eta > 0$, $c > 0$, and $C > 0$ such that, for all $\theta \in B(\theta^*,\eta)$,

$$\max\Big\{0, \max_{j=1,\ldots,J_1} E(m_j(X,\theta))/\sigma_j(\theta), \max_{j=J_1+1,\ldots,J} |E(m_j(X,\theta))/\sigma_j(\theta)|\Big\}$$
$$\geq C \times \min\{d(\theta,\Theta_I), c\}.$$

This "minorant" assumption ensures that the population criterion increases not too slowly as one moves away from Θ_I. Loosely speaking, it prevents "weak identification" problems. Next, BCS impose a polynomial minorant condition as well.[3]

Assumption 4. For each support point $\theta^* \in S(p,\Theta_I)$, there exist constants $\eta > 0$, $c > 0$, and $C > 0$ such that, for all $\theta \in B(\theta^*,\eta) \cap H(p,\Theta_I)$,

$$\max\Big\{0, \max_{j=1,\ldots,J_1} E(m_j(X,\theta))/\sigma_j(\theta), \max_{j=J_1+1,\ldots,J} |E(m_j(X,\theta))/\sigma_j(\theta)|\Big\}$$
$$\geq C \times \min\{d(\theta, S(p,\Theta_I)), c\}.$$

It bears emphasis that the last two assumptions are logically independent. Assumption 3 forces the population criterion to increase (to first order) as we move away from Θ_I. Assumption 4 enforces an analogous increase as we move away from the support set $S(p,\Theta_I)$ *along the supporting hyperplane*. This does not imply Assumption 3 because it only applies as one leaves Θ_I in selective directions. But it also is not implied because the directions considered in Assumption 4 may be tangential to Θ_I, in which case the increase in $d(\theta,\Theta_I)$ may be of low order. Indeed, Assumption 4 may be considered restrictive: By not allowing directions to be tangential to Θ_I without also being tangential to $S(p,\Theta_I)$, it excludes smooth maxima, e.g., any identified set whose entire boundary is a smooth manifold.[4]

We next adapt (in this order) assumptions 3, 4(a), and 4(b) from PPHI. These are modified in a few ways: PPHI assume that the support point θ^* is unique; this assumption is removed. Assumptions are also localized by only looking at θ near θ^*; this makes the first of them meaningfully weaker. At the same time,

[3] In the respective originals, both assumptions raise the r.h.s. to a power $\gamma > 0$, hence "polynomial minorant." However, in both cases, γ is restricted in ways that imply $\gamma = 1$ in the present setting.

[4] Consider the single moment inequality $\theta_1^2 + \theta_2^2 - E(X) \leq 0$, with $E(X) = 1$ and $p = (0,1)$. Then $S(p,\Theta_I)$ is a singleton at $\theta^* = (0,1)$. Let $\theta = (\zeta,1)$, then $\theta_1^2 + \theta_2^2 - E(X) = \zeta^2$, whereas $d(\theta, S(p,\Theta_I)) = \zeta$, violating Assumption 4 as $\zeta \to 0$; at the same limit, we have $d(\theta,\Theta_I) = \sqrt{1+\zeta^2} - 1 > \zeta^2/4$, so that Assumption 3 holds. We note that this and other easy counterexamples to Assumption 4 do not seem to be counterexamples to the BCS profiling method. We leave to future research the question whether this illustrates the possibility of relaxing their assumptions.

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

PPHI impose assumptions uniformly over d.g.p.'s; to keep this paper focused, such uniform statements are removed throughout.[5]

Assumption 5. For each support point $\theta^* \in S(p, \Theta_I)$, there exist $\delta, \eta > 0$ s.t. $\sup_{t \in T(\theta^*, \eta)} p't \leq -\delta$, where

$$T(\theta^*, \eta) \equiv \left\{ t = \frac{\theta - \theta^*}{\|\theta - \theta^*\|} : \theta \in \Theta_I \cap B(\theta^*, \eta), \theta \neq \theta^* \right\}.$$

Assumption 6. There are no equality constraints, i.e., $J_1 = J$. Furthermore, for each support point $\theta^* \in S(p, \Theta_I)$, there exist constants $\delta, \varepsilon > 0$ as well as direction (i.e., unit vector) $t \in \mathbb{R}^d$ s.t.

$$\max_{j: E(m_j(X, \theta^*))/\sigma_j(\theta^*) \geq -\delta} D_j(\theta^*)t \leq -\varepsilon.$$

Assumption 7. For each support point $\theta^* \in S(p, \Theta_I)$, there exist constants $\delta < 0$, $\varepsilon > 0$ s.t.

$$\inf_{t \in \mathbb{R}^d, \|t\| = 1, p't \geq \delta} \max \left\{ \max_{j \in \mathcal{J}_1^*(\theta^*)} D_j(\theta^*)t, \max_{j \in \{J_1 + 1, \ldots J\}} |D_j(\theta^*)t| \right\} > \varepsilon.$$

Brief intuitions for these are as follows. Assumption 5 ensures that Θ_I is contained in a cone that has θ^* as apex and does not otherwise intersect $H(p, \Theta_I)$. In particular, this implies uniqueness of θ^* (although we will not use this feature) and pointiness of the tangent cone (defined later) at θ^*. Assumption 6 ensures that locally to θ^*, there exists a direction in which all moment expectations, hence their maximum, strictly decrease. Note in particular that this excludes equality constraints—while this implication is not explicit in PPHI, they treat equalities as conjunctions of two inequalities, and the assumption cannot possibly hold for both. PPHI point out that their assumption is inspired by CHT's degeneracy (i.e., our Assumption 2), which also excludes equalities. Indeed, both assumptions force Θ_I to have an interior, and we will have more to say about their relation later. Assumption 7 enforces that the criterion is strictly increasing in many directions from θ^*, including all that have positive inner product, and some that have negative inner product, with p.[6]

We finally recall some classic constraint qualifications. This requires some standard notation that we will also use later. For any $\theta \in \Theta_I$, define the tangent cone

$$\mathcal{T}(\theta) \equiv \left\{ t \in \mathbb{R}^d : \exists \{\theta_m\}_{m=1}^{\infty} \subset \Theta_I, \theta_m \to \theta, \lim_{m \to \infty} \frac{\theta_m - \theta}{\|\theta_m - \theta\|} = \frac{t}{\|t\|} \right\} \cup \{0\}$$

[5] To compare the statements of assumptions, also keep in mind notational conventions: PPHI write moment conditions as $E(m_j(\cdot)) \geq 0$, set $p = (-1, 0, \ldots, 0)$, and use subscripts to denote components of vectors, so that t_1 there would be $-p't$ here.

[6] The restriction $\delta < 0$ in Assumption 7 correctly reflects our source. However, we considered the possibility that (in our notation) $\delta > 0$ was intended. The assumption then becomes weaker. Specifically, along the lines of our main result below, one can show that it is then equivalent to $\max_{j \in \mathcal{J}^*(\theta^*)} D_j(\theta^*)p > 0$ and is implied by Assumption 10.

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

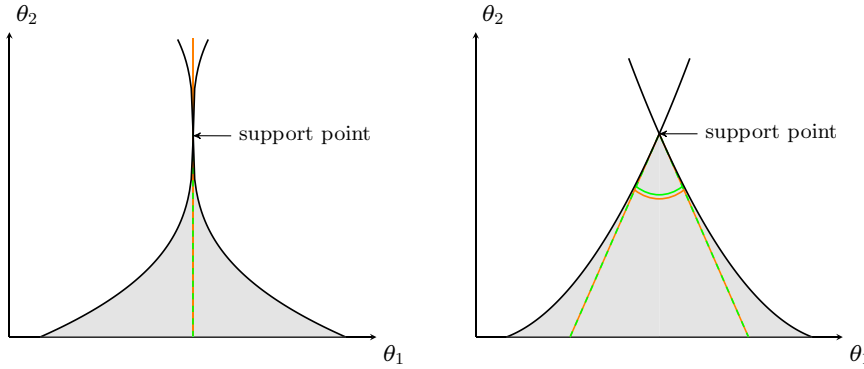


FIGURE 2. An illustration of linear cone (orange) versus tangent cone (green). Left panel: The two disagree in the example from Figure 1; note that the linear cone goes both "up" and "down" from the support point. Right panel: The two agree in a well-behaved setting that cannot be statistically distinguished from the ill-behaved one.

as well as the linearized cone

$$\mathcal{L}(\theta) \equiv \left\{t \in \mathbb{R}^d : D_j(\theta)t \leq 0, j \in \mathcal{J}_1^*(\theta); D_j(\theta)t = 0, j \in \{J_1+1, \ldots, J\}\right\}.$$

We will only invoke these objects for support points $\theta^* \in S(p, \Theta_I)$. Recall that $\mathcal{T}(\theta) \subseteq \mathcal{L}(\theta)$ (Bazaraa et al. 2006, chap. 5).

Both cones are illustrated in Figure 2. They coincide in "nice" examples like the right panel of Figure 2 (indeed, this "niceness" is the Abadie constraint qualification defined below) but they may disagree, as in the left panel which reprises the left panel from Figure 1.

We can then state the following constraint qualifications in decreasing order of restrictiveness.

Assumption 8 Linear Independence Constraint Qualification (LICQ). For each support point $\theta^* \in S(p, \Theta_I)$, the active (equality or inequality) constraints have linearly independent gradients $D_j(\theta^*)$.

Of course, the LICQ requires at most d active constraints, making it quite restrictive.

Assumption 9 Mangasarian–Fromowitz Constraint Qualification (MFCQ). At each support point $\theta^* \in S(p, \Theta_I)$, the gradients of the equality constraints are linearly independent and there exists $t \in \mathbb{R}^d$ s.t. $D_j(\theta^*)t < 0$ for $j \in \mathcal{J}_1^*(\theta^*)$ and $D_j(\theta^*)t = 0$ for $j \in \{J_1+1, \ldots, J\}$.

Assumption 10 Abadie Constraint Qualification (ACQ). For each support point $\theta^* \in S(p, \Theta_I)$, we have $\mathcal{L}(\theta^*) = \mathcal{T}(\theta^*)$.

These conditions are frequently invoked in the statistical literature. For example, Shapiro (1990, 1991, 1993) uses either LICQ or uniqueness of Lagrange multipli-

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

ers, and these two assumptions are essentially the same (Wachsmuth, 2013). In econometrics, Cho and Russell (2019) and Kaido and Santos (2014) use LICQ; Andrews, Roth, and Pakes (2019) and Gafarov (2019) restrict attention to linear constraints and thereby impose ACQ.

3. RESULTS

3.1. Restating Some Assumptions, and a First Set of Equivalences

We next restate some of these assumptions, exploiting their localization or using the language of constraint qualifications. Between our localization of assumptions and the language of tangent cones, several of the assumptions we just introduced can be stated more succinctly. In what follows, recall that $\mathcal{J}^*(\theta)$, defined in (2.3), is the active set of (equality or inequality) constraints at θ. Specifically, define:

Assumption 3'. For each support point $\theta^* \in S(p, \Theta_I)$, there exist constants $\eta > 0$ and $C > 0$ such that, for all $\theta \in B(\theta^*, \eta)$,

$$\max\Big\{0, \max_{j=1,\ldots,J_1} E(m_j(X,\theta))/\sigma_j(\theta), \max_{j=J_1+1,\ldots,J} |E(m_j(X,\theta))/\sigma_j(\theta)|\Big\} \geq C \times d(\theta, \Theta_I).$$

Assumption 4'. For each support point $\theta^* \in S(p, \Theta_I)$, there exist constants $\eta > 0$ and $C > 0$ such that, for all $\theta \in B(\theta^*, \eta) \cap H(p, \Theta_I)$,

$$\max\Big\{0, \max_{j=1,\ldots,J_1} E(m_j(X,\theta))/\sigma_j(\theta), \max_{j=J_1+1,\ldots,J} |E(m_j(X,\theta))/\sigma_j(\theta)|\Big\} \geq C \times d(\theta, S(p, \Theta_I)).$$

Assumption 5'. For each support point $\theta^* \in S(p, \Theta_I)$, $\max\{p't/\|t\| : t \in \mathcal{T}(\theta^*) \setminus \{0\}\} < 0$.

Assumption 6'. There are no equality constraints. Furthermore, for each support point $\theta^* \in S(p, \Theta_I)$, $\min_{t \in \mathbb{R}^d} \max_{j \in \mathcal{J}^*(\theta^*)} D_j(\theta^*)t/\|t\| < 0$.

Assumption 7'. For each support point $\theta^* \in S(p, \Theta_I)$, $\max\{p't/\|t\| : t \in \mathcal{L}(\theta^*) \setminus \{0\}\} < 0$.

Then we have:

LEMMA 3.1. *Suppose that Assumption 1 holds. Then the following assumptions are equivalent: 3⇔3', 4⇔4', 5⇔5', 6⇔6', and 7⇔7'.*

Proof. Regarding Assumptions 3 and 3', $\Leftarrow$ is obvious and $\Rightarrow$ holds because in Assumption 3, one can choose $\eta = c$, ensuring $\min\{d(\theta, \Theta_I), c\} = d(\theta, \Theta_I)$. The argument for 4⇔4' is the same.

Next, 5⇒5' holds because $T(\theta^*, \eta)$ shrinks toward $\mathcal{T}(\theta^*) \setminus \{0\}$ as $\eta \to 0$. Also, suppose $\mathcal{T}(\theta^*) \setminus \{0\} = \emptyset$, then Assumption 5' holds vacuously; but in this case, $T(\theta^*, \eta) = \emptyset$ for small enough η and so Assumption 5 holds as well. It remains to show that, if $\mathcal{T}(\theta^*) \setminus \{0\} \neq \emptyset$ and therefore $T(\theta^*, \eta) \neq \emptyset$ for any η,

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

then failure of Assumption 5 implies failure of Assumption 5'. Now, failure of Assumption 5 and nonemptiness of $T(\theta^*, \eta)$ jointly imply existence of sequences $(\delta_n, \eta_n) \downarrow (0,0)$ and $\theta_n \in B(\theta^*, \eta_n) \setminus \{\theta^*\}$ s.t. $p'(\theta_n - \theta^*)/\|\theta_n - \theta^*\| > -\delta_n$. But then any accumulation point t of $(\theta_n - \theta^*)/\|\theta_n - \theta^*\|$ is in $\mathcal{T}(\theta^*)$ and has $p't \geq 0$, contradicting Assumption 5'.

Assumption 6 obviously implies 6'. To see the reverse implication, suppose 6' holds, then one can verify Assumption 6 by choosing δ to be half the slack of the tightest inactive inequality (or arbitrarily if all inequalities are active).

Next, note first that

$$\max\{p't/\|t\| : t \in \mathcal{L}(\theta^*) \setminus \{0\}\} < 0 \tag{3.1}$$
$$\iff \min_{t \in \mathbb{R}^d : \|t\|=1, p't \geq 0} \max\Big\{ \max_{j \in \mathcal{J}_1^*(\theta^*)} D_j(\theta^*)t, \max_{j \in \{J_1+1,\ldots,J\}} |D_j(\theta^*)t| \Big\} > 0.$$

For example, it is easy to see that the above minimum is attained. If it equalled 0, the vector t attaining it would be in $\mathcal{L}(\theta^*)$, so the first maximum would be at least 0. The converse argument is similar.

The left-hand side of (3.1) is Assumption 7'. We next show that its right-hand side is equivalent to Assumption 7. It is implied by Assumption 7 because the minimization is over a smaller set. To see the converse, suppose Assumption 7 fails, then there exist sequences $\delta_n \uparrow 0$, $\varepsilon_n \downarrow 0$, and t_n with $p't_n \geq \delta_n$ and $\max\{\max_{j \in \mathcal{J}_1^*(\theta^*)} D_j(\theta^*)t_n, \max_{j \in \{J_1+1,\ldots,J\}} |D_j(\theta^*)t_n|\} \leq \varepsilon_n$. Any accumulation point of t_n then is a counterexample to the right-hand side of (3.1). ■

Remark 3.1. Some of these equivalences are due to localization of assumptions. For example, the original Assumption 3 in PPHI, but also both polynomial minorant conditions, are otherwise stronger than their simplifications. Indeed, regarding the equivalences 3⇔3' and 4⇔4', the real insight is that the original assumptions combine local and global conditions, e.g., (for Assumption 3)

$$d(\theta, \Theta_I) \leq \delta \tag{3.2}$$
$$\implies \max\Big\{0, \max_{j=1,\ldots,J_1} E(m_j(X,\theta))/\sigma_j(\theta), \max_{j=J_1+1,\ldots,J} |E(m_j(X,\theta))/\sigma_j(\theta)|\Big\} \geq C \times d(\theta, \Theta_I)$$

and

$$d(\theta, \Theta_I) \geq \delta \tag{3.3}$$
$$\implies \max\Big\{0, \max_{j=1,\ldots,J_1} E(m_j(X,\theta))/\sigma_j(\theta), \max_{j=J_1+1,\ldots,J} |E(m_j(X,\theta))/\sigma_j(\theta)|\Big\} \geq \varepsilon$$

for some $\varepsilon > 0$ (namely, setting $\varepsilon = C\delta$). Only the local condition (3.2) is related to constraint qualifications. The other part is really a global identification condition. We will revisit this distinction later.

3.2. Econometric Assumptions as Constraint Qualifications

We now present our main insight: Several of the above assumptions are equivalent, or very close to, constraint qualification assumptions, and there are numerous

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

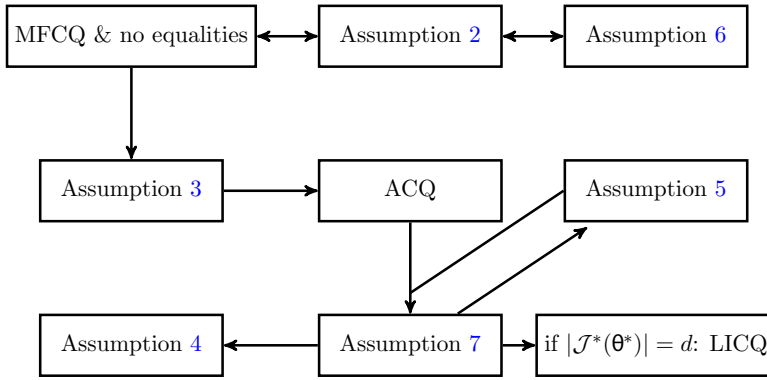


FIGURE 3. A flow chart illustrating Theorem 3.1. Examples for reading this: "3 implies Abadie Constraint Qualification (ACQ)"; "2 and 6 are equivalent"; "5 and ACQ jointly imply 7."

logical relations between them. Our main result, which is also visualized in Figure 3, follows.

THEOREM 3.1. *Suppose Assumption 1 holds. Then the following relations between assumptions hold true.*

1. *Assumptions 2 and 6 are equivalent. Furthermore, both are equivalent to jointly (i) excluding equality restrictions and (ii) imposing 9.*
2. *Any of Assumptions 2, 6, and 9 (the latter combined with excluding equality constraints) imply 3.*
3. *Assumption 3 implies 10.*
4. *Assumptions 5 and 10 jointly imply 7.*
5. *Assumption 7 implies 5.*
6. *Assumption 7 implies that the gradients of active constraints span $\mathbb{R}^d$. In particular, if $\mathcal{J}^*(\theta^*)$ has exactly d elements, 8 is implied.*
7. *Assumption 7 implies 4.*

Proof. Throughout this proof, consider a fixed θ^*. We invoke Lemma 3.1 and use the simplified versions of the assumptions.

1. If equalities are excluded, MFCQ reduces to

$$\min_{t\in\mathbb{R}^d:\|t\|=1} \max_{j\in\mathcal{J}^*(\theta^*)} D_j(\theta^*)t < 0. \tag{3.4}$$

This immediately clarifies that excluding inequalities and imposing MFCQ is just Assumption 6'. It remains to show equivalence with Assumption 2.

Assumption 2 excludes equalities because the presence of even one equality constraint implies that $\max_{j=1,\dots,J} E(m_j(X,\theta))/\sigma_j(\theta) = 0$ for all $\theta \in \Theta_I$. This leaves

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

two possibilities: Either $\Theta_I^{-\varepsilon} \neq \emptyset$, in which case (2.4) fails, or $\Theta_I^{-\varepsilon} = \emptyset$, in which case (2.5) fails because $d(\theta, \Theta_I^{-\varepsilon}) = \infty$.

To see that Assumption 2 also implies (3.4), consider a sequence $\varepsilon_n \to 0$. For n large enough we have $\varepsilon_n/M \leq \delta$, where M and δ are from Assumption 2. Then by (2.5) there exists $\theta_n \in \Theta_I^{-\varepsilon_n/M}$ with $\|\theta_n - \theta^*\| \leq \varepsilon_n$, and by (2.4), we have $\max_{j \in \mathcal{J}^*(\theta^*)}\{E(m_j(X,\theta_n))/\sigma_j(\theta_n)\} \leq -C\varepsilon_n/M$. Next, let t be any accumulation point of $(\theta_n - \theta^*)/\|\theta_n - \theta^*\|$, then by continuous differentiability one has $\max_{j \in \mathcal{J}^*(\theta^*)} D_j(\theta^*)t \leq -C/M < 0$.

To see the converse, let

$$t^* = \arg \min_{t \in \mathbb{R}^d : \|t\| = 1} \max_{j \in \mathcal{J}^*(\theta^*)} D_j(\theta^*)t \tag{3.5}$$

and let $\mu = \frac{1}{2}|\max_{j \in \mathcal{J}^*(\theta^*)} D_j(\theta^*)t^*|$. By (3.4), $\max_{j \in \mathcal{J}^*(\theta^*)} D_j(\theta^*)t^* = -2\mu < 0$.

We next argue why inactive constraints, i.e., $j \notin \mathcal{J}^*(\theta^*)$, can be ignored in what follows. Note first that by the Mean Value Theorem and Assumption 1, there exists $\tilde{\theta}$ componentwise between θ and θ^* s.t.

$$\begin{aligned}\frac{E(m_j(X,\theta))}{\sigma_j(\theta)} - \frac{E(m_j(X,\theta^*))}{\sigma_j(\theta^*)} &= D_j(\tilde{\theta})(\theta - \theta^*) \\ &= (D_j(\theta^*) + O(\|\theta - \theta^*\|))(\theta - \theta^*) = O(\|\theta - \theta^*\|) + O(\|\theta - \theta^*\|^2).\end{aligned}$$

Let $\rho \equiv \max_{j \notin \mathcal{J}^*(\theta^*)} E(m_j(X,\theta^*))/\sigma_j(\theta^*) < 0$, then it follows that

$$\begin{aligned}&\max_{\theta \in B(\theta^*,\eta)} \max_{j \notin \mathcal{J}^*(\theta)} E(m_j(X,\theta))/\sigma_j(\theta) \leq \rho \\ &\quad + \max_{\theta \in B(\theta^*,\eta)} \left\{ \frac{E(m_j(X,\theta))}{\sigma_j(\theta)} - \frac{E(m_j(X,\theta^*))}{\sigma_j(\theta^*)} \right\} < 0\end{aligned}$$

for a small enough choice of η. Thus, inactive constraints do not affect the value taken by $\max_{j=1,\ldots,J} E(m_j(X,\theta))/\sigma_j(\theta)$ anywhere on $B(\theta^*,\eta)$. (In the remainder of this proof, η is understood to be either the η just chosen or the η from Assumption 2, whichever is smaller.)

Fix $\delta > 0$ and consider any $\theta \in B(\theta^*,\eta)$ with $\max_{j \in \mathcal{J}^*(\theta^*)} E(m_j(X,\theta))/\sigma_j(\theta) > -\delta$. Then

$$\begin{aligned}&\max_{j \in \mathcal{J}^*(\theta^*)} E(m_j(X,\theta - \delta t^*/\mu))/\sigma_j(\theta - \delta t^*/\mu) \\ &= \max_{j \in \mathcal{J}^*(\theta^*)} \left\{ E(m_j(X,\theta))/\sigma_j(\theta) - \delta D_j(\bar{\theta}_j)t^*/\mu \right\} \\ &= \max_{j \in \mathcal{J}^*(\theta^*)} \left\{ E(m_j(X,\theta))/\sigma_j(\theta) - \delta D_j(\theta^*)t^*/\mu - o(\delta) \right\} \\ &\geq -\delta + 2\delta - o(\delta) > 0\end{aligned}$$

for δ small enough, implying that $\theta - 2\delta t^*/\mu \notin \Theta_I$ and therefore that $d(\theta, \Theta \setminus \Theta_I) < 2\delta/\mu$. (Here, $\bar{\theta}_j$ is componentwise between θ and $\theta - \delta t^*/\mu$ and may change with j; $\Theta_I \subset \operatorname{int}\Theta$ ensures $\theta - 2\delta t^*/\mu \in \Theta$ for δ small enough.)

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

Conversely, by setting $\varepsilon = 2\delta/\mu$, we find that (for ε small enough)

$$d(\theta, \Theta \setminus \Theta_I) \geq \varepsilon \implies \max_{j \in \mathcal{J}^*(\theta^*)} E(m_j(X,\theta))/\sigma_j(\theta) \leq -\mu\varepsilon/2,$$

verifying (2.4) with $C = \mu/2$. The requirement that ε be small enough can be enforced by choosing η low enough.

Next, for any $\theta \in \Theta_I \cap B(\theta^*, \eta)$, we similarly have

$$\begin{aligned}
& \max_{j \in \mathcal{J}^*(\theta^*)} E(m_j(X, \theta + \delta t^*/\mu))/\sigma_j(\theta + \delta t^*/\mu) \\
= & \max_{j \in \mathcal{J}^*(\theta^*)} \left\{E(m_j(X,\theta))/\sigma_j(\theta) + \delta D_j(\bar{\theta}_j) t^*/\mu\right\} \\
= & \max_{j \in \mathcal{J}^*(\theta^*)} \left\{E(m_j(X,\theta))/\sigma_j(\theta) + \delta D_j(\theta^*) t^*/\mu + o(\delta)\right\} \\
\leq & \; 0 - 2\delta + o(\delta)
\end{aligned}$$

for δ small enough. (Again, $\theta + \delta t^*/\mu \in \Theta$ because $\Theta_I \subset \operatorname{int}\Theta$. The interpretation, though not the value taken, of $\bar{\theta}_j$ is as before.)

Now, set $\overline{M} = \max_{j \in \mathcal{J}^*(\theta^*)} \|D_j(\theta^*)\|$ and let t be any unit vector, then

$$\begin{aligned}
& \max_{j \in \mathcal{J}^*(\theta^*)} E(m_j(X, \theta + \delta t^*/\mu + \delta t/\overline{M}))/\sigma_j(\theta + \delta t^*/\mu + \delta t/\overline{M}) \\
= & \max_{j \in \mathcal{J}^*(\theta^*)} \left\{E(m_j(X, \theta + \delta t^*/\mu))/\sigma_j(\theta + \delta t^*/\mu) + \delta D_j(\bar{\theta}_j) t/\overline{M}\right\} \\
\leq & \; -2\delta + \delta + o(\delta) < 0
\end{aligned}$$

for δ small enough, thus

$$B(\theta + \delta t^*/\mu, \delta/\overline{M}) \subseteq \Theta_I \implies \theta + \delta t^*/\mu \in \Theta_I^{-\delta/\overline{M}} \implies d(\theta, \Theta_I^{-\delta/\overline{M}}) \leq \delta/\mu.$$

Setting $\varepsilon = \delta/\overline{M}$, we find $d(\theta, \Theta_I^{-\varepsilon}) \leq \varepsilon\overline{M}/\mu$, verifying (2.5) with $M = \overline{M}/\mu$.

2. Suppose that (3.4) applies and let μ and t^* be as in (3.5). Fix any scalar

$$\gamma \in \left(0, \frac{1}{\mu} \max_{\theta \in B(\theta^*, \eta)} \max_{j \in \mathcal{J}^*(\theta^*)} E(m_j(X,\theta))/\sigma_j(\theta)\right],$$

noting that by Assumption 1(d), the upper bound on γ vanishes as $\eta \to 0$. Consider any $\theta \in B(\theta^*, \eta)$ s.t. $\max_{j \in \mathcal{J}^*(\theta^*)} E(m_j(X,\theta))/\sigma_j(\theta) < \mu\gamma$. Then by a use of the mean value theorem very similar to preceding displays,

$$\max_{j \in \mathcal{J}^*(\theta^*)} E(m_j(X, \theta + \gamma t^*))/\sigma_j(\theta + \gamma t^*) < \mu\gamma - 2\mu\gamma + o(\gamma) < 0$$

for γ small enough (which can be ensured by choosing η small enough). It follows that $\theta + \gamma t^* \in \operatorname{int}\Theta_I$, hence $d(\theta, \Theta_I) < \gamma$. Conversely, $d(\theta, \Theta_I) = \gamma$ then implies $\max_{j \in \mathcal{J}^*(\theta^*)} E(m_j(X,\theta))/\sigma_j(\theta) \geq \mu\gamma$ and therefore $\max_j E(m_j(X,\theta))/\sigma_j(\theta) \geq \mu\gamma$. As γ was arbitrary, this verifies Assumption 3' with $C = \mu$.

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

3. Fix any $t \in \mathcal{L}(\theta^*)$. By differentiability of $E(m_j(X,\cdot))/\sigma_j(\cdot)$ and the definition of $\mathcal{L}(\cdot)$, we then have

$$\limsup\Big\{n \times \max\Big\{\max_{j\in\mathcal{J}_1^*(\theta^*)} D_j(\theta^*)t/n, \max_{j\in\{J_1+1,\ldots J\}} |D_j(\theta^*)t/n|\Big\}\Big\} = 0 \text{ as } n\to\infty.$$

But then Assumption 3' implies $n \times d(\theta^* + t/n, \Theta_I) \to 0$. Next, let $\theta_n = \arg\min_{\theta\in\Theta_I} \|\theta - (\theta^* + t/n)\|$ and therefore $\|\theta_n - (\theta^* + t/n)\| = d(\theta^* + t/n, \Theta_I)$, then

$$\begin{aligned}\lim \frac{\theta_n - \theta^*}{\|\theta_n - \theta^*\|} &= \lim \frac{\theta_n - (\theta^* + t/n) + (\theta^* + t/n) - \theta^*}{\|\theta_n - (\theta^* + t/n) + (\theta^* + t/n) - \theta^*\|} \\ &= \lim \frac{n \times (\theta_n - (\theta^* + t/n)) + t}{\|n \times (\theta_n - (\theta^* + t/n)) + t\|} = \frac{t}{\|t\|},\end{aligned}$$

hence $t \in \mathcal{T}(\theta^*)$.

4. Under ACQ, $\mathcal{L}(\theta^*) = \mathcal{T}(\theta^*)$, hence Assumptions 7' and 5' are then equivalent.

5. This holds because $\mathcal{T}(\theta^*) \subseteq \mathcal{L}(\theta^*)$.

6. Suppose the conclusion fails, thus no d gradients of active constraints span $\mathbb{R}^d$. Then there exists a unit vector t s.t. $D_j(\theta^*)t = 0$ for all $j \in \mathcal{J}^*(\theta^*)$. This implies

$$\min_{t\in\mathbb{R}^d:\|t\|=1, p't\geq 0} \max_{j\in\mathcal{J}^*(\theta^*)} D_j(\theta^*)t = 0$$

because at least one of $(t, -t)$ is feasible in this minimization problem, contradicting Assumption 7'; compare in particular the equivalent representation of this assumption on the right-hand side of (3.1).

7. Because $p't = 0 \Leftrightarrow \theta^* + t \in H(p, \Theta_I)$, the right-hand side of (3.1), and thereby Assumption 7', can be restated as

$$\begin{aligned}\min_{\theta\in H(p,\Theta_I)} \max\Big\{&\max_{j\in\mathcal{J}_1^*(\theta^*)} D_j(\theta^*)(\theta-\theta^*)/\|\theta-\theta^*\|, \\ &\max_{j\in\{J_1+1,\ldots J\}} |D_j(\theta^*)(\theta-\theta^*)|/\|\theta-\theta^*\|\Big\} > 0.\end{aligned}$$

By continuous differentiability (using arguments very similar to above), this then implies that, for small enough $\eta > 0$,

$$\begin{aligned}\max\Big\{0, \max_{j=1,\ldots,J_1} E(m_j(X,\theta))/\sigma_j(\theta), \max_{j=J_1+1,\ldots,J} |E(m_j(X,\theta))/\sigma_j(\theta)|\Big\} \geq C\|\theta-\theta^*\| \\ \text{for all } \theta \in B(\theta^*,\eta)\cap H(p,\Theta_I),\end{aligned}$$

where $C = \frac{1}{2}\min_{t\in\mathbb{R}^d:\|t\|=1, p't=0} \max\Big\{\max_{j\in\mathcal{J}_1^*(\theta^*)} D_j(\theta^*)t, \max_{j\in\{J_1+1,\ldots J\}} D_j(\theta^*)t\Big\} > 0$. ■

3.3. Tightness of Theorem 3.1

We next clarify that Theorem 3.1 is tight, i.e., no implication that is not indicated in Figure 3 holds true. This subsection can be skipped without loss of continuity.

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

Let $m_j(X,\theta) = \mu_j(\theta) + X_j$ for functions $\mu_j : \Theta \mapsto \mathbb{R}$ defined below and suppose that all X_j are standard normal; thus, $E(m_j(X,\theta))/\sigma_j(\theta) = \mu_j(\theta)$. Also, $\Theta = [-1,1]^2$. All examples are constructed s.t. the support point of interest is $\theta^* = (0,0)$. Direction of projection is $p = (0,1)$ unless explicitly indicated otherwise. There are no equality constraints, so for the purpose of these examples, "MFCQ and no equalities" is just MFCQ.

- **Neither Assumption 5' nor 4' imply either 7' or 10 (ACQ).**

$$\mu_1(\theta) = \theta_2^3 - \theta_1$$
$$\mu_2(\theta) = \theta_2^3 + \theta_1.$$

We start with this example because several others build on it. Qualitatively, it is the left panel of Figure 1. The linear cone $\mathcal{L}(\theta^*)$ is spanned by $\{(0,1),(0,-1)\}$, the tangent cone $\mathcal{T}(\theta^*)$ is spanned only by $\{(0,-1)\}$. The one direction, $t = (0,-1)$, that is in $\mathcal{T}(\theta^*)$ is tangential to all constraints. The example violates ACQ (and all stronger assumptions) as well as 7' but fulfills 5' as well as 4'.

- **Assumption 3' does not imply 9 (MFCQ).**

$$\mu_1(\theta) = \theta_2^3 - \theta_1$$
$$\mu_2(\theta) = \theta_2^3 + \theta_1$$
$$\mu_3(\theta) = \theta_2.$$

This example adds a third constraint to the first example; compare the right panel of Figure 1. Both $\mathcal{L}(\theta^*)$ and $\mathcal{T}(\theta^*)$ are spanned by $(0,-1)$. The example therefore fulfills 3' (and by implication ACQ) but not MFCQ.

- **Assumption 10 (ACQ) does not imply 3'.**

$$\mu_1(\theta) = \theta_1^2(\theta_2 + \theta_1^2)$$
$$\mu_2(\theta) = \theta_2.$$

Here, $\mathcal{L}(\theta^*) = \mathcal{T}(\theta^*) = \{t : p't \le 0\}$, so that ACQ is fulfilled. However, 3' is violated in direction $t = (1,0)$:

$$\begin{aligned} \mu_1(\gamma t) &= \gamma^4 \\ d(\gamma t, \Theta_I) &\le \gamma^2 \quad \text{(because } (\gamma, -\gamma^2) \in \Theta_I) \\ \Longrightarrow\ \mu_1(\gamma t)/d(\gamma t, \Theta_I) &\to 0 \quad \text{as } \gamma \to 0. \end{aligned}$$

- **Assumption 7' does not imply 10 (ACQ).**

$$\mu_1(\theta) = -\theta_1 + \theta_2$$
$$\mu_2(\theta) = \theta_1 + \theta_2$$
$$\mu_3(\theta) = (\theta_1\theta_2)^2.$$

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

In this example (which is inspired by a well-known counterexample to ACQ), we have that $\mathcal{L}(\theta^*)$ is spanned by $\{(-1,-1),(1,-1)\}$ but $\mathcal{T}(\theta^*)$ is spanned by $(0,-1)$ only.

- **Assumption 4' does not imply 7' or 5'.**

$$\mu_1(\theta) = \theta_2$$
$$\mu_2(\theta) = \theta_1 + \theta_2.$$

In this example, $\mathcal{L}(\theta^*)$ and $\mathcal{T}(\theta^*)$ coincide and are spanned by $\{(-1,0),(1,-1)\}$, contradicting both 7' and 5'. Assumption 4', which here only applies if we move in direction $(1,0)$ from θ^*, is fulfilled.

4. DISCUSSION

Our findings inform a number of clarifying remarks on the existing literature.

- As mentioned above, CHT's polynomial minorant condition can be disentangled into a local and a global identification condition. The local condition (3.2) is a mild strengthening of ACQ and is implied by degeneracy. The global condition (3.3) is essentially the weakest additional statement needed to ensure that Θ_I is a well-separated (if set-valued) minimum of $\max\{\max_{j=1,\dots,J_1} E(m_j(X,\theta))/\sigma_j(\theta), \max_{j=J_1+1,\dots,J} |E(m_j(X,\theta))/\sigma_j(\theta)|, 0\}$.[7] While the polynomial minorant condition is, therefore, not redundant, an instructive restatement of the assumptions is available.
- Regarding assumptions in PPHI, claim 5 of Theorem 3.1 owes to our simplification, but in view of the smoothness imposed in their Assumption 7, claim 4 also applies to the original versions.

  Also, if $J = d$, then the PPHI assumptions imply LICQ. This clarifies relation to recent work by Cho and Russell (2019) and Gafarov (2019): Both effectively impose LICQ and benefit from this by being able to propose relatively simple inference. However, while stronger than assumptions in CHT, BCS, and certainly Kaido, Molinari, and Stoye (2019), the assumptions exceed those in PPHI only in the sense of excluding "overidentified" support points, i.e., more than d active constraints, and are actually weaker in other respects.
- Cho and Russell (2019) furthermore impose Assumption 3, i.e., (by Theorem 3.1) ACQ. This is not redundant in their paper because 3 is imposed for all $\theta \in \Theta_I$ and with universal C, i.e., "more uniformly" than LICQ.
- Yildiz (2012) presents conditions for Hausdorff consistency of $\hat{\Theta}_I$. Some of these are in essence constraint qualifications and can be related to our analysis as follows. For the case of pure inequality constraints, her high-level Assumption 3.1 states that Θ_I is the closure of the *strict* level 0 lower contour set of

[7] A well-separated minimum requires that for each $\varepsilon > 0$, there exists $\delta > 0$ s.t. $d(\theta, \Theta_I) \geq \varepsilon$ implies $\max\{\dots\} \geq \delta$. Its role in ensuring "inner consistency" of sample analogs of identified sets is well understood (Newey and McFadden, 1994).

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

$\max_j E(m_j(X,\theta))/\sigma_j(\theta)$.[8] For the case of at most d active constraints, this is derived as an implication of the LICQ (Assumption 3.2). For the more general case, it is derived from an assumption (in Lemma 3.1) enforcing that, at any boundary point θ^* of Θ_I, the criterion function $\max_j E(m_j(X,\theta))/\sigma_j(\theta)$ is strictly increasing in some component of θ. To make it comparable to our assumptions, one would impose this to hold at any support point θ^*. By a minimal extension of step 2 of Theorem 3.1 (see expression (3.4)), it is then equivalent to Assumption 2. Therefore, Yildiz (2012) essentially imposes MFCQ for the pure inequality case.[9] For the case of mixed equalities and inequalities, she invokes a LICQ (Assumption 4.1(b)).

- We close with some remarks on why, for inference on projections $p'\theta$, certain approaches do not require constraint qualifications. Specifically, the profiling approach in BCS gets by with the relatively weak Assumption 4; Kaido et al. (2019) or projection of confidence regions for θ (such as those in Andrews and Soares, 2010, Bugni, 2009, or Canay, 2010) use no shape restrictions for Θ_I at all.

  The reason is that all of these approaches localize inference at a conjectured true value of the support function $s(p,\Theta_I)$ (in BCS) or parameter vector θ (in all others). Consistent estimation of identified sets for these objects is then not a concern. In particular, BCS need to ensure some form of consistency of a sample analog of Θ_I that is restricted to the true supporting hyperplane, and this is precisely what the minorant on the support plane achieves. The other approaches need no constraint qualifications at all.

  Of course, there is no free lunch. All the methods just alluded to are computationally expensive because they effectively invert tests whose critical value depends on the exact value of the parameter under the null. Thus, while some shortcuts may be available in practice (see in particular Kaido et al., 2017, 2019), critical values must in principle be recomputed at every conceivable value of θ or $s(p,\Theta)$.

5. VERIFYING ASSUMPTIONS IN SOME EXAMPLES

In this section, we discuss how these restrictions apply to two well-understood examples of partial identification. It will become clear that one cannot take for granted that they "typically" hold and that verifying them could be intricate in more involved examples. That is part of our message: We point out that many of them are basically constraint qualifications, and it is well known that constraint qualifications can be subtle to verify.

[8] Molchanov (1998) also uses this condition to ensure Hausdorff consistency of set estimators.

[9] Yildiz (2012) writes that her assumptions imply a degeneracy condition in CHT without claiming the reverse implication. This refers to their high-level degeneracy Assumption C.3, which is implied whenever the simple sample analog of Θ_I is consistent.

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

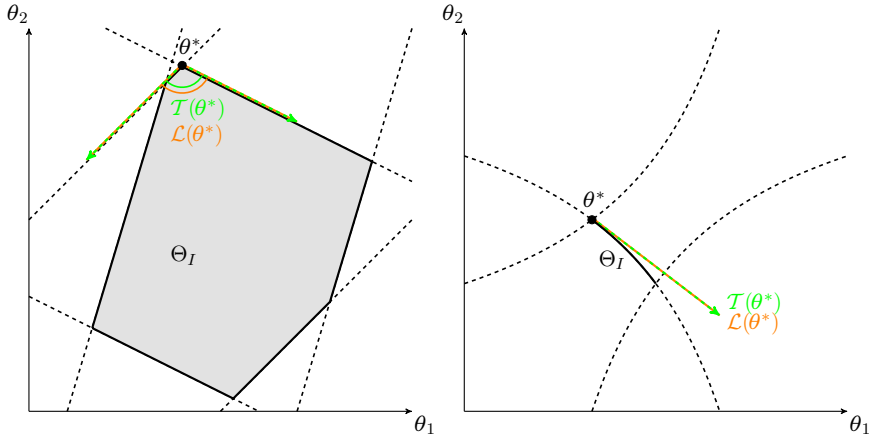


FIGURE 4. Visualization of the two examples from Section 5. Dashed lines denote (in)equality restrictions. The figures also display the identified set Θ_I, the support point θ^* in direction $p = (0, 1)$ (i.e., "up," as in previous figures), and the associated linear and tangent cones (which coincide). Left panel: Linear regression with interval outcome data (Section 5.1). Right panel: A simple entry game (Section 5.2).

The examples are visualized in Figure 4, which is designed to resemble Figure 2. Note that in the second example, Θ_I has zero measure.

5.1. Linear Regression with Interval Outcome Data and Discrete Regressors

Consider a linear regression model:

$$W = Z'\theta + u,$$

where $Z = (Z_1, \ldots, Z_d)$ is a $d \times 1$ random vector with $Z_1 = 1$. We assume that Z has k points of support denoted $\mathcal{Z} = \{z^1, \ldots, z^k \in \mathbb{R}^d\}$ with $\max_{r=1,\ldots,k} \|z^r\| < M < \infty$. The researcher observes $\{W_0, W_1, Z\}$ with $P(W_0 \leq W \leq W_1 | Z = z^r) = 1, r = 1, \ldots, k$, where $\pi^r = P(Z = z^r) > 0, r = 1, \ldots, k$ are assumed known. Suppose that W_0 and W_1 take values in $\mathcal{W} \subset \mathbb{R}$. An important special case is missing data, where W_0 and W_1 both equal the true W if the latter is observed and correspond to some bound on it otherwise.

The identified set is characterized by the following moment inequalities.

$$E(W_0 | Z = z^r) - z^{r\prime}\theta \leq 0, \; r = 1, \ldots, k,$$
$$z^{r\prime}\theta - E(W_1 | Z = z^r) \leq 0, \; r = 1, \ldots, k.$$

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

Equivalently,

$$E(W_0 1\{Z = z^r\})/\pi^r - z^{r\prime}\theta \le 0,\ r = 1, \ldots, k, \tag{5.1}$$

$$z^{r\prime}\theta - E(W_1 1\{Z = z^r\})/\pi^r \le 0,\ r = 1, \ldots, k. \tag{5.2}$$

Define the following objects.

$$\sigma_j = \begin{cases} \mathrm{Var}(W_0 1\{Z = z^j\})^{1/2}/\pi^j, & j = 1, \ldots, k \\ \mathrm{Var}(W_1 1\{Z = z^{j-k}\})^{1/2}/\pi^{j-k}, & j = k+1, \ldots, 2k, \end{cases}$$

$$D_j = \begin{cases} -z^{j\prime}/\sigma_j, & j = 1, \ldots, k \\ z^{j-k\prime}/\sigma_j, & j = k+1, \ldots, 2k, \end{cases}$$

$$b_j = \begin{cases} -\frac{E(W_0|Z=z^j)}{\mathrm{Var}(W_0|Z=z^j)^{1/2}} & j = 1, \ldots, k \\ \frac{E(W_1|Z=z^{j-k})}{\mathrm{Var}(W_1|Z=z^{j-k})^{1/2}} & j = k+1, \ldots, 2k. \end{cases}$$

Note that, since D_j does not depend on θ, we drop its argument.

By (5.1) and (5.2), the identified set is a polytope characterized by k pairs of parallel constraints

$$E(W_0 1\{Z = z^r\})/\pi^r \le z^{r\prime}\theta \le \theta - E_P(W_1 1\{Z = z^r\})/\pi^r,\ r = 1, \ldots, k. \tag{5.3}$$

See the left panel of Figure 4 for an illustration with $k = 3$. Each support point is either a vertex of the polytope or a point in the relative interior of a nonsingleton support set. We therefore analyze subcases below. Note that, since the gradients $z^1, \ldots, z^k$ of the constraints in (5.3) are known, one knows without data which case applies.

We first establish conditions implying MFCQ. For $j = 1, \ldots, 2k$, define:

$$H_j = \{\theta \in \Theta : D_j\theta = b_j\}$$
$$H_j^- = \{\theta \in \Theta : D_j\theta \le b_j\}.$$

We call H_j a *hyperplane* and H_j^- a *half space*. Let $\mathrm{ri}(A)$ denote the relative interior of A. A $(d-1)$-dimensional flat face in $\mathbb{R}^d$ is called a *facet*. Similarly, a ℓ-dimensional element of a d-dimensional polytope is called a ℓ-face, where $1 \le \ell \le d-2$. For example, a one-face is an edge in a three-dimensional polytope.

LEMMA 5.1. *Suppose (i) $\mathcal{W}$ is compact; (ii) $\Theta = \{\theta \in \mathbb{R}^d : \|\theta\|^2 \le B_0\}$ with $B_0 < \infty$ satisfying $C_0 B_0 > k \sup_{w\in\mathcal{W}} w^2$, where $C_0 \equiv \inf_{p\in\mathbb{S}^{d-1}} \sum_r (p^\prime z^r)^2$; (iii) $k \ge d$, and any subset $\mathcal{C} \subseteq \mathcal{Z}$ with $\#\mathcal{C} \le d$ is linearly independent; (iv) $E[W_1 - W_0|Z = z^r] > 0$ for all $r = 1, \ldots, k$. Let $\theta^* \in S(p, \Theta_I)$.*

1. *(Facet) If $S(p, \Theta_I)$ is not a singleton and is the intersection of a hyperplane and a finite number of half spaces, MFCQ holds at any $\theta^* \in ri(S(p, \Theta_I))$;*
2. *(ℓ-face) If $S(p, \Theta_I)$ is not a singleton and is the intersection of more than one hyperplanes, MFCQ holds at any $\theta^* \in ri(S(p, \Theta_I))$; and*
3. *(Vertex) If $S(p, \Theta_I)$ is a singleton and $\#\mathcal{J}^*(\theta^*) \le d$, MFCQ holds at θ^*.*

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

Remark 5.1. The lemma shows that, even if one is not sure about whether θ^* is on a facet or any other lower dimensional elements of the polytope, MFCQ holds as long as $\#\mathcal{J}^*(\theta^*) \leq d$ (and other conditions of the lemma hold). Also, things simplify if $k = d$, in which case conditions (iii) and (iv) imply $\#\mathcal{J}^*(\theta^*) \leq d$ at any support point. Condition (iii) then ensures LICQ, and therefore MFCQ, at θ^*.

Proof. Under our assumptions, Θ_I is nonempty and is in the interior of Θ by Proposition F.1 in Kaido and Santos (2014).

Case 1: Suppose $S(p, \Theta_I)$ is a flat face and $\theta^* \in \text{ri}(S(p, \Theta_I))$. This occurs if and only if $p = cz^r$ (or $p = -cz^r$) for some $r \in \{1, \ldots, k\}$ and $c > 0$, and $p'\theta^* = E(W_1 1\{Z = z^r\})/\pi^r$ (or $E(W_0 1\{Z = z^r\})/\pi^r$). By (iii), such r is unique. Without loss of generality, suppose $p'\theta^* = E(W_1 1\{Z = z^r\})/\pi^r$. The assumption $E[W_1 - W_0|Z = z^r] > 0$ ensures

$$E(W_1 1\{Z = z^r\})/\pi^r - E(W_0 1\{Z = z^r\})/\pi^r = E[E[W_1 - W_0|Z = z^r]1\{Z = z^r\}]/\pi^r > 0. \tag{5.4}$$

Hence, $p'\theta^* > E(W_0 1\{Z = z^r\})/\pi^r$, implying the lower bound is slack. This and $\theta^* \in \text{ri}(S(p, \Theta_I))$ imply $\mathcal{J}^*(\theta^*) = \{j^*\}$ with $j^* = r + k$.

The gradient of the normalized moment is

$$D_{j^*} = \frac{z^{r\prime}}{\sigma_{j^*}} = \frac{z^{r\prime}}{\text{Var}(W_1 1\{Z = z^r\})^{1/2}/\pi^r}.$$

Let $t^* = -z^r$, then

$$D_{j^*} t^* = -\frac{z^{r\prime} z^r}{\text{Var}(W_1 1\{Z = z^r\})^{1/2}/\pi^r} < 0,$$

which establishes MFCQ.

Case 2: We first claim that $\#\mathcal{J}^*(\theta^*) < d$. Suppose otherwise. Then, there are at least d active inequalities at θ^*. Select a subset $\tilde{\mathcal{J}} \subset \mathcal{J}(\theta^*)$ such that $\#\tilde{\mathcal{J}} = d$. Then, by condition (iii), $\{D_j, j \in \tilde{J}\}$ are linearly independent, implying θ^* is a unique solution to the system of linear equations

$$D_j \theta = b_j, \ \forall j \in \tilde{\mathcal{J}}.$$

Furthermore, by the necessary condition of the maximization problem, there is $\lambda \in \mathbb{R}^{2k}_+$ such that

$$p + \sum_{j \in \tilde{\mathcal{J}}} \lambda_j D_j' = 0, \tag{5.5}$$

where $\lambda_j > 0$ for all $j \in \tilde{\mathcal{J}}$.

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

Let $\tilde{\theta} \in S(p, \Theta_I)$ and $\tilde{\theta} \neq \theta^*$. By construction, $p'(\tilde{\theta} - \theta^*) = 0$. This and (5.5) imply

$$\sum_{j \in \tilde{\mathcal{J}}} \lambda_j D_j (\tilde{\theta} - \theta^*) = 0. \tag{5.6}$$

Suppose that $D_j(\tilde{\theta} - \theta^*) > 0$ for some $j \in \tilde{\mathcal{J}}$. This implies

$$D_j \tilde{\theta} > b_j,$$

hence the jth constraint is violated, hence $\tilde{\theta}$ cannot be in $S(p, \Theta_I)$, a contradiction. Similarly, suppose that $D_j(\tilde{\theta} - \theta^*) < 0$ for some $j \in \tilde{\mathcal{J}}$, then by (5.6) and the positivity of the Lagrange multipliers, there must exist $j' \in \tilde{\mathcal{J}}$ such that $D_{j'}(\tilde{\theta} - \theta^*) > 0$, and the same argument applies. The only remaining possibility is $D_j(\tilde{\theta} - \theta^*) = 0$ for all $j \in \tilde{\mathcal{J}}$, in which case

$$D_j \tilde{\theta} = b_j, \ \forall j \in \tilde{\mathcal{J}},$$

which contradicts θ^* being the unique solution to (5.5). Therefore, $\#\mathcal{J}^*(\theta^*) < d$ must hold. Now, by condition (iii) and $\#\mathcal{J}^*(\theta^*) < d$, LICQ holds at θ^*, which implies the claim.

Case 3: By $\#\mathcal{J}^*(\theta^*) \leq d$ and condition (iii), LICQ holds at θ^*, which implies the claim. ■

We conclude by mentioning other conditions. The identified set Θ_I is a finite polytope, hence any projection is the value of a linear program. This immediately clarifies that ACQ obtains, which furthermore means that Assumptions 5 and 7 coincide. Both restrict $\mathcal{L}(\theta^*)$, which locally just coincides with Θ_I, to not intersect the supporting hyperplane other than at θ^*. This will hold iff the support set is a singleton, i.e., a vertex. Finally, the solution to a linear program is necessarily well-separated, so that Assumption 4 holds.

5.2. A Simple Entry Game

Two-player entry games are a workhorse example of partial identification since Tamer (2003). We analyze the game specified in BCS but without covariates. The equilibrium concept is pure strategy Nash equilibrium (PSNE). Firm 1 and, respectively, 2 enters if

$$\varepsilon_1 - \theta_1 A_2 \geq 0,$$
$$\varepsilon_2 - \theta_2 A_1 \geq 0,$$

where $(A_1, A_2) \in \{0, 1\}^2$ are the firms' actions, $(\theta_1, \theta_2) \in [0, 1]^2$ are the interaction parameters, and $(\varepsilon_1, \varepsilon_2)$ are observed by the players but not by the researcher. This system is incomplete: For certain realization of $(\varepsilon_1, \varepsilon_2)$, both $(1, 0)$ and $(0, 1)$ are PSNE and the model is silent on which is played. Hence, different (stochastic) selection mechanisms picking an equilibrium in the region of multiplicity can be

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

coupled with different values of θ to yield the same distribution of observables (A_1, A_2). Nonetheless, inference can be carried out by bounding the likelihood of different outcomes from below by the respective probabilities of them being unique PSNE. In this simple example, this exhausts all the information in the model and data, leading to a sharp identification region. See Beresteanu, Molchanov, and Molinari (2011) for characterizations of sharp identification regions in more complex models.

In this model,

- $(1,1)$ is the unique PSNE if $\varepsilon_1 - \theta_1 A_2 \geq 0$ and $\varepsilon_2 - \theta_2 A_1 \geq 0$,
- $(1,0)$ is the unique PSNE if $\varepsilon_1 - \theta_1 A_2 \geq 0$ and $\varepsilon_2 - \theta_2 A_1 \leq 0$, and
- $(0,1)$ is the unique PSNE if $\varepsilon_1 - \theta_1 A_2 \leq 0$ and $\varepsilon_2 - \theta_2 A_1 \geq 0$.

Letting $\pi_{jk} = \Pr(A_1 = j, A_2 = k)$ and assuming (as in BCS) that $(\varepsilon_1, \varepsilon_2)$ are distributed independently uniformly on $[0,1]$, we have

$$\pi_{11} = (1-\theta_1)(1-\theta_2), \tag{5.7}$$

$$\pi_{10} \geq (1-\theta_1)\theta_2, \tag{5.8}$$

$$\pi_{01} \geq \theta_1(1-\theta_2), \tag{5.9}$$

$$\pi_{00} = 0. \tag{5.10}$$

Geometrically, the identified set Θ_I is an arc segment. The right panel of Figure 4 depicts its true shape if $\pi_{01} = \pi_{10} = \pi_{11} = 1/3$.

Consider now the problem of bounding θ_2 from above; all other bounds on individual parameters are similar. Guessing that (5.7) and (5.8) will bind at the solution, one can easily solve for

$$\theta^* = \begin{bmatrix} \theta_1^* \\ \theta_2^* \end{bmatrix} = \begin{bmatrix} \pi_{01} \\ \pi_{10}/(\pi_{10}+\pi_{11}) \end{bmatrix}.$$

Furthermore, the linear and tangent cone are tangent to (5.7) and are spanned by $(\pi_{10}+\pi_{11}, -\pi_{11}/(\pi_{10}+\pi_{11}))$. This vector has strictly negative inner product with $p = (0,1)$. We conclude:

- Assumption 2 and any equivalent assumption cannot hold because Θ_I has no interior.
- All of Assumptions 3, 4, 5, 7, LICQ, and ACQ hold.

However, fulfillment of some of these assumptions is delicate in that it depends on θ^* being an exposed point of Θ_I. Other directions of optimization (e.g., $p = (1,-1)$, corresponding to testing the null that $\theta_1 = \theta_2$) have θ^* in the relative interior of Θ_I, i.e., at a point where only (5.7) is active. Assumptions 4, 5, and 7 will then fail.

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

6. CONCLUSION

The literature on partial identification uses constraint qualifications in many ways: To ensure Hausdorff consistency or rates of convergence for simple estimators of identified sets (CHT; Yildiz, 2012), to justify inference for the full parameter vector θ (CHT) or subvectors (PPHI; Cho and Russell, 2019; Gafarov, 2019), or to justify efficiency bounds (Kaido and Santos, 2014). However, some of these uses are implicit, making it difficult even for expert readers to compare assumptions. We provide a guide to how different high-level assumptions relate to each other and to well-known constraint qualifications. A simple, important message is that several high-level assumptions are tightly related to the Mangasarian–Fromowitz constraint qualification and are essentially mutually equivalent. We believe that this provides useful guidance to readers trying to make sense of the large menu of inference methods for partially identified vectors and subvectors (Canay and Shaikh, 2017; Molinari, 2020). For example, it clarifies costs and benefits relative to work that has weak-to-no geometric regularization, mostly at the expense of computational effort.

REFERENCES

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

Andrews, I., J. Roth, & A. Pakes (2019) Inference for Linear Conditional Moment Inequalities. Working Paper 26374, National Bureau of Economic Research.

Bazaraa, M.S., H.D. Sherali, & C. Shetty (2006) *Nonlinear Programming: Theory and Algorithms*. 3rd ed. Wiley-Interscience.

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

Bonnans, J. & A. Shapiro (2000) *Perturbation Analysis of Optimization Problems*. Springer.

Bugni, F. A. (2009) Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. *Econometrica* 78, 735–753.

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

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

Canay, I.A. & A.M. Shaikh (2017) *Practical and Theoretical Advances in Inference for Partially Identified Models*, vol. 2 of Econometric Society Monographs, pp. 271–306. Cambridge University Press.

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

Cho, J. & T.M. Russell (2019) Simple inference on functionals of set-identified parameters defined by convex moments, arXiv:1810.03180.

Eizenberg, A. (2014) Upstream innovation and product variety in the U.S. Home PC Market. *The Review of Economic Studies* 81, 1003–1045.

Gafarov, B. (2019) Inference in high-dimensional set-identified affine models, arXiv:1904.00111.

Ho, K. & A. Pakes (2014) Hospital choices, hospital prices, and financial incentives to physicians. *American Economic Review* 104, 3841–3884.

Holmes, T. J. (2011) The diffusion of Wal-Mart and economies of density. *Econometrica* 79, 253–302.

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press

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

Kaido, H., F. Molinari, J. Stoye, & M. Thirkettle (2017) Calibrated projection in MATLAB, arxiv:1710.09707.

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

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

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

Newey, W.K. & D. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, chap. 36. Elsevier.

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

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

Shapiro, A. (1990) On differential stability in stochastic programming. *Mathematical Programming* 47, 107–116.

Shapiro, A. (1991) Asymptotic analysis of stochastic programs. *Annals of Operations Research* 30, 169–186.

Shapiro, A. (1993) Asymptotic behavior of optimal solutions in stochastic programming. *Mathematics of Operations Research* 18, 829–845.

Tamer, E. (2003) Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies* 70, 147–165.

Wachsmuth, G. (2013) On LICQ and the uniqueness of Lagrange multipliers. *Operations Research Letters* 41, 78–80.

Yildiz, N. (2012) Consistency of plug-in estimators of upper contour and level sets. *Econometric Theory* 28, 309–327.

https://doi.org/10.1017/S0266466621000207 Published online by Cambridge University Press