

Sharp Identification for Regressions with Interval-Observed and Missing Covariates*

Martin Fankhauser[†], Gil Peled[‡]

October 14, 2026

Abstract

Researchers frequently estimate regressions when key variables are missing or observed only through intervals. Existing approaches typically resolve this incompleteness through imputation or auxiliary assumptions on the incomplete-data process. We instead ask what the maintained regression model and incomplete data alone imply. For a broad class of conditional moment models, we derive an operational characterization of the sharp identified set, develop a Hausdorff-consistent estimator and uniformly valid confidence regions, and provide tractable computational methods for common regression models. An application to a multiple-price-list design shows that incomplete-data-robust inference can remain informative, while also demonstrating that sharpness matters: non-sharp outer sets can substantially overstate the uncertainty implied by the observed data.

Keywords: Partial identification, set identification, conditional moments, missing covariates, interval data, random sets, support function

JEL Codes: C14, C18, C21, C51

*We thank Francesca Molinari, Arie Beresteanu, Federico Bugni, Toru Kitagawa, Ivan Canay, Charles Manski, Filip Obradovic, Pamela Giustinelli, Massimiliano Marcellino, Víctor Sancibrián, Joachim Freyberger, David Pacini, Chris Muris, Ismael Mourifié, Marc Henry, Daniel Wilhelm, Joachim Winter, Raffaella Giacomini and participants of the IAAE 2026 in Lisbon, as well as the 31st Panel Data Conference in Exeter for helpful discussions.

[†]Bocconi University

[‡]Cornell University

1 Introduction

Regression models are ubiquitous in applied economics, yet the variables entering them are often not observed exactly. Outcomes and covariates may be missing because of nonresponse or attrition, reported only in brackets in household surveys, top-coded in administrative data, or observed through intervals generated by experimental and survey designs. In each case, the observed data leave ambiguity about the latent variables entering the regression, so that the parameters of interest are only partially identified.

Empirical work typically addresses incomplete data through sample restrictions, imputation, or assumptions about the data generating process. These choices reflect the trade-off emphasized in C. Manski (2003) in his *law of decreasing credibility*: stronger assumptions can deliver more informative conclusions at the cost of credibility. This paper therefore asks: what does the maintained regression model identify from the incomplete data alone, without additional assumptions about how missing or interval observations are generated?

Formally introduced in Section 2, we study conditional moment models of the form

$$\mathbf{E}[m(Y, X, \theta_0) \mid X] = 0,$$

where some components of the non-conditioning variables Y , the conditioning variables X , or both are not observed exactly. This class includes for instance, linear, probit or logit regressions as well as more general econometric models characterized by conditional moment restrictions. Let $W \equiv (Y, X)$ denote the latent variables entering the model and let $W^O \equiv (Y_l, Y_u, X_l, X_u)$ denote the observed intervals. The observed data imply the containment restriction $Y \in [Y_l, Y_u]$ and $X \in [X_l, X_u]$ almost surely. Our setting nests the missing data case if the observed intervals span the full support of the respective variables.

Our central identifying restriction is deliberately weak. Beyond this containment requirement we leave the process generating incomplete observations unrestricted: whether a variable is missing, as well as the location and width of an observed interval, may depend on the latent

variables themselves. In particular, we do not impose missingness-at-random or other restrictions on the relationship between the observation process and the latent variables. This approach gives a benchmark for assessing what additional assumptions contribute. Whether such weak restrictions yield informative conclusions is an empirical question, which our application addresses.

Our object of interest is the identified set for θ_0 , i.e. the smallest set of parameter values consistent with the model and the incomplete data. A candidate parameter value is retained if there exists a single completion of the latent data that is consistent with the observed data generating process, respects the reported intervals, and satisfies the conditional moment restriction. This identified set, denoted by Θ_I^* , contains θ_0 whenever the containment restriction and the maintained model hold. If the containment restriction holds and $\Theta_I^* = \emptyset$, the conditional moment model is incompatible with the population distribution of the observed data. In contrast, $\Theta_I^* \neq \emptyset$ establishes compatibility with at least one admissible latent distribution; it does not establish that the model holds under the true latent-data distribution.

We begin by examining unconditional moment restrictions generated by successively larger finite collections of instrument functions. Each collection gives an outer approximation to the identified set by retaining parameter values for which one admissible completion satisfies all moments in the collection jointly. As the instrument collection becomes rich enough to recover the conditional restrictions, these approximations converge to the identified set. Building on the random-set approach of Beresteanu and Molinari (2008) and Beresteanu, I. Molchanov, and Molinari (2011), we characterize feasibility through support-function inequalities that depend only on the observed data. As a consequence, elements of the identified set can be verified by solving a mini-max optimization problem. The main technical step extends this approach to infinitely many moments and allows discontinuous instruments, such as indicators partitioning the covariate space.

Estimation and inference require additional arguments as the instrument collection expands.

For set estimation, we establish a triangular-array law of large numbers for random sets of growing dimension and use it to verify the consistency conditions of Chernozhukov, Hong, and Tamer (2007). For confidence regions for θ_0 , we verify the manageability conditions required by Andrews and Shi (2017) for the support-function process indexed by an *uncountable* class of directions and the instrument class.

The results yield procedures for empirical work based on finite outer approximations which converge to the identified set in the limiting case. We estimate these outer approximations and obtain Hausdorff consistency for the identified set as the sample size and number of moments increase. We also construct confidence regions with asymptotically uniform coverage for the true parameter. For important specifications, including linear, probit, and logit regressions, a partition structure of the instruments allows the mini-max optimization problem for a finite truncation to be formulated as a linear program under the conditions in Section 4.

Our empirical application examines how conclusions change as we progressively account for different forms of incomplete observation. We revisit Castillo, Jordan, and Petrie (2018), who use a multiple price list (MPL) to elicit discount rates and study their relationship with on-time high-school graduation. Respondents make a sequence of binary choices between an earlier payment and a later payment, with each choice corresponding to an implied discount rate. Because the offered rates lie on a discrete grid, even a single switching point reveals only an interval containing the latent discount rate. A conventional probit based on midpoint imputation yields a marginal effect of the discount rate on the probability of on-time graduation, evaluated at the mean, of -0.086 with a 90% confidence interval of $[-0.161, -0.011]$. Allowing each respondent's discount rate to take any value within its elicited interval yields a 90% confidence set of $[-0.196, -0.008]$, so the qualitative conclusion is unchanged. Accounting for inconsistent response patterns matters substantially more. Roughly 8% of respondents switch multiple times. Allowing their discount rates to lie anywhere between their first and last switches widens the confidence set to $[-0.387, 0.159]$. Including respondents with

missing graduation outcomes widens the confidence set further. Estimating the identified set directly rather than an outer approximation matters empirically: a non-sharp outer-set procedure based on Aucejo, Bugni, and Hotz (2017), which retains additional parameter values not ruled out by its characterization, yields a 90% confidence set of $[-0.366, 0.176]$ and therefore includes zero, whereas our sharp confidence set of $[-0.196, -0.008]$ does not. Our work is most closely related to the partial-identification literature on regressions with interval or missing covariates. C. Manski and Tamer (2002) study regression models with interval-observed variables and obtain sharp identification results under an auxiliary mean-independence restriction linking the latent variable to its observed interval. Asher, Novosad, and Rafkin (2018) built on their work and impose further assumptions. In contrast, we retain only the containment restriction and otherwise leave the observation process unrestricted. Aucejo, Bugni, and Hotz (2017) study conditional moment models with missing covariates and develop estimation and inference procedures for outer sets, i.e. super sets of the sharp identified set. In comparison, our work imposes very similar assumptions but delivers results for the sharp identified set and can handle interval-valued data. Schennach (2014) develops a more general entropy-based approach to latent-variable models and extends it to conditional moment restrictions through a growing collection of unconditional moments.¹ Our more specialized support-function approach instead exploits the structure of interval and missing observations directly, yielding sharp identification together with estimation, inference, and computational procedures tailored to regression models. In particular, we show that, for important regression specifications, the nested optimization underlying the support-function characterization can be reduced to a linear program.

A related literature obtains sharp identification when the relevant restrictions are finite-dimensional. J. Horowitz and C. Manski (1998) derive sharp worst-case bounds for conditional expectations of the form $\mathbf{E}[g(Y) \mid X \in A]$ when outcomes, covariates, or both

¹Applying this framework to our setting would still require developing estimation, inference, and computational procedures for the resulting growing-moment problem. Some of the tools developed in this paper for handling an expanding collection of moments could potentially also be used for such an implementation.

may be missing, while J. Horowitz and C. Manski (2000) specialize this approach to randomized experiments and bound conditional response probabilities and treatment contrasts. In treatment-effect settings, Lee (2009) obtains sharp trimming bounds under random assignment and monotonicity of treatment-induced selection, with subsequent work extending this approach under related restrictions on selection. A distinct methodological strand uses random sets and support functions to characterize sharp identification from finitely many moment restrictions. Our approach builds directly on Beresteanu and Molinari (2008) and Beresteanu, I. Molchanov, and Molinari (2011), who characterize feasible moment predictions through Aumann expectations and support-function inequalities; Beresteanu, I. Molchanov, and Molinari (2011) apply this framework to best linear prediction with interval data, improving on the computationally demanding characterization in J. L. Horowitz and C. F. Manski (2006). More recently, Franguridi and Moon (2026) extend the same logic to GMM models with unrestricted, potentially endogenous missingness. In contrast, our conditional moment model requires an infinite collection of restrictions and allows the conditioning variables themselves to be incompletely observed. We extend this framework to the infinite-dimensional and potentially discontinuous moment structure needed to recover conditional restrictions, yielding a sharp and operational characterization while leaving the process generating missing and interval observations unrestricted. More broadly, the paper follows the partial-identification approach pioneered by C. F. Manski (2005) and C. F. Manski (2008); see Molinari (2020), Canay and Shaikh (2017), and Kline and Tamer (2023) for reviews. A separate literature restores point identification by placing stronger restrictions on how incomplete observations are generated or by modeling the missing values themselves. Missing-data methods commonly invoke ignorability or missing-at-random assumptions and use likelihood, multiple imputation, or inverse-probability weighting (D. Rubin 1976; D. B. Rubin 1987; Robins, Rotnitzky, and Zhao 1995; Wooldridge 2007); in longitudinal settings, related weighting approaches are routinely used to address attrition (Fitzgerald, Gottschalk, and Moffitt 1998). Selection models instead specify how observation depends on latent out-

comes or unobservables, ranging from the classical parametric correction of Heckman (1979) to semiparametric approaches such as Newey (2009). For more general forms of partial observation, Heitjan and D. Rubin (1991) formulate conditions under which the coarsening mechanism can be ignored. Other approaches exploit auxiliary information, validation samples, restrictions on reporting behavior, or explicit models for the latent values within observed intervals (Hsiao 1983; Rigobon and Stoker 2009; Pacini 2018; Duan and Pelger 2025; Freyberger et al. 2025; Bretschneider and Kaddoura 2026). These methods can be considerably more informative when their assumptions are credible. Our identified set instead provides an assumption-minimal benchmark: each additional restriction can be viewed as eliminating latent completions from the larger set Γ , of all completions feasible under the containment restriction; see Section 2.

The remainder of the paper is organized as follows. Section 2 derives the sharp identification result. Section 3 develops estimation and inference, and Section 4 presents computational implementation. Section 5 studies finite-sample performance, Section 6 presents the empirical application, and Section 7 concludes.

Notation. Throughout the paper, ℓ_1 and ℓ_∞ denote the usual sequence spaces equipped with norms $|\cdot|_1$ and $|\cdot|_\infty$, respectively, and $\langle u, a \rangle$ denotes the dual pairing between $u \in \ell_\infty = (\ell_1)^*$ and $a \in \ell_1$. We write $B^p \equiv \{u \in \ell_\infty : \|u\|_p \leq 1\}$ for the closed unit ball under the ℓ_p norm, where $p \in [1, \infty]$. $h(A, u) \equiv \sup_{a \in A} \langle u, a \rangle$ denotes the support function of a set $A \subseteq \ell_1$. For a random set Q , $\mathbb{E}[Q] = \text{cl} \{\mathbf{E}[q] : q \text{ is an integrable selection of } Q\}$ denotes its Aumann expectation, whereas $\mathbf{E}_F[Z]$ denotes the ordinary expectation of a random element Z under the probability law F , with the subscript suppressed when clear from context. We write $\rho_H(A, B)$ for the Hausdorff distance and $\vec{\rho}_H(A, B)$ for the directed Hausdorff distance. The symbols $A \oplus B$ and $\text{co}(A)$ denote the Minkowski sum and convex hull, respectively. Finally, for a metric space $\mathcal{S}$, $\Delta(\mathcal{S})$ denotes the set of Borel probability measures on $\mathcal{S}$, and $\mathcal{L}_F(Z)$ denotes the law of Z under F .

2 Identification

Formally, we study conditional moment models of the form

$$\mathbf{E}_{F_0} [m(Y, X, \theta_0) \mid X] = 0 \quad F_X\text{-a.s.} \quad (1)$$

Let $(\Omega, \mathcal{A}, F_0)$ denote the underlying probability space. The random variables Y and X take values in $\mathcal{Y} \subseteq \mathbb{R}^{d_y}$ and $\mathcal{X} \subseteq \mathbb{R}^{d_x}$, respectively. The known moment function $m : \mathbb{R}^{d_y} \times \mathbb{R}^{d_x} \times \Theta \rightarrow \mathbb{R}^{d_m}$ satisfies (1) at the true parameter, should it exist, $\theta_0 \in \Theta$. We assume $d_y, d_x, d_m < \infty$; we allow Θ to be arbitrary for the purpose of identification, however, in Section 3 we restrict attention to $\Theta \subset \mathbb{R}^{d_\theta}$. The source of partial identification is not an incomplete model, but incomplete observation of the variables entering it.

Assumption 1. *Let $W \equiv (Y, X)$ take values in $\mathcal{W} \equiv \mathcal{Y} \times \mathcal{X}$, where $\mathcal{Y} \subset \mathbb{R}^{d_y}$ and $\mathcal{X} \subset \mathbb{R}^{d_x}$ are compact. Let $W^O = (Y_l, Y_u, X_l, X_u)$ be observed on the same probability space as W and take values in $\mathcal{W}^O \equiv \mathcal{Y}^2 \times \mathcal{X}^2$, with nonempty componentwise intervals almost surely. Assume*

$$P_{F_0} [Y \in [Y_l, Y_u], X \in [X_l, X_u]] = 1.$$

We find it convenient at times to write this last condition as $P_{F_0}(W \in D(W^O)) = 1$, for $D(W^O) \equiv [Y_L, Y_U] \times [X_L, X_U]$.

Assumption 1 contains the central identifying stance of the paper. Apart from requiring the latent realization to lie inside its observed interval, we place no further restrictions on how incompleteness is generated. A fully observed value is represented by a degenerate interval, while a missing value is represented by the variable's full support. The purpose of the analysis is precisely to determine what follows before adding any stronger assumptions. Let us first define the identified set for our setting. Assumption 1 implicitly defines the set

of admissible completions

$$\Gamma \equiv \{\eta \in \Delta(\mathcal{W} \times \mathcal{W}^O) : \eta_{\mathcal{W}^O} = P_{\mathcal{W}^O}, \quad \eta(\{(w, w^o) : w \in D(w^o)\}) = 1\}$$

An element of Γ is a candidate joint law for the latent and observed data that has the observed marginal $P_{\mathcal{W}^O}$ and assigns probability one to latent values compatible with the reported intervals. Such a law cannot be ruled out from the observed data alone. Assumption 1 also guarantees nonemptiness, because $\eta_0 \equiv \mathcal{L}_{F_0}(W, W^O) \in \Gamma$.

The identified set being discussed is defined directly from the maintained model (1) and assumption 1 as

$$\Theta_I^* \equiv \{\theta \in \Theta : \exists \eta \in \Gamma \text{ such that } \mathbf{E}_\eta[m(Y, X, \theta) \mid X] = 0 \quad \eta_X\text{-a.s.}\}. \quad (2)$$

The set Θ_I^* may be empty. This natural definition of the identified set may sometimes be redefined in convenient ways. In particular, Schennach (2014) considers a distance-based formulation for model 1 expressed with the condition $\|E_\eta[m(Y, X, \theta) \mid X]\| = 0$, F_X -almost surely. The precise relationship between the original identified set and that based on a distance-based formulation has not been previously examined to our knowledge. In Appendix Section A.3 we establish that in our setting the two identified sets coincide, the relationship between them can be expressed well through random sets; and provide a sufficient condition of equivalence. We do not discuss this distance-based formulation any further in the main text of this paper.

The characterization of Θ_I^* in terms of conditional expectations is inconvenient as the conditioning event is not observed. We begin by replacing the conditional equality by a (countably) infinite collection of unconditional equalities using instrument functions, as is standard in the conditional-moment literature; see, among others, Andrews and Shi (2013), Andrews and Shi (2017) and Aucejo, Bugni, and Hotz (2017). In Lemma 1 we establish that this can be performed without changing the identified set. The following assumption gives general

conditions on the instrument class used, whereas in the Appendix Section A.2 we verify them for the class of weighted cubes used in our applications.

Assumption 2.

- (i) Let $\mathcal{H} = \{h_k : k \geq 1\}$ be a countable class of bounded Borel functions on $\mathcal{X}$ such that, for every $f \in C(\mathcal{X})$ and every $\varepsilon > 0$, there exists $v_\varepsilon \in \text{span}(\mathcal{H})$ satisfying

$$\|f - v_\varepsilon\|_\infty < \varepsilon.$$

- (ii) Let $\mathcal{G} = \{g_k : k \geq 1\}$ be the weighted instrument class defined by $g_k = \omega_k h_k$, where the deterministic weights $\omega_k > 0$ are chosen so that

$$\lim_{K \rightarrow \infty} \sup_{x \in \mathcal{X}} \sum_{k > K} |g_k(x)| = 0.$$

In particular having weights decay on the order of $O(k^{-\beta})$ for $\beta > 1$, will achieve this result.

Part (i) ensures that the instruments reveal any nonzero conditional moment, note $\text{span}(\mathcal{H}) = \text{span}(\mathcal{G})$. This mimics Assumption MFA in Andrews and Shi (2017) and Assumption CI in Andrews and Shi (2013) which both require the instrument class to be rich enough that the set of x violating the conditional moments has positive measure if and only if some instrument reveals that fact. Part (ii) ensures that the stacked moment vector lives in ℓ_1 , which handles the technical difficulties of the space and ensures finite truncations are uniformly accurate.

Define the unconditional-moment set by

$$\Theta_I \equiv \left\{ \theta \in \Theta : \begin{array}{l} \exists \eta \in \Gamma \text{ such that} \\ \int_{\mathcal{W} \times \mathcal{W}^o} m_j(y, x, \theta) g_k(x) \eta(dy, dx, dw^o) = 0, \\ j = 1, \dots, d_m, \quad k \geq 1, \end{array} \right\}, \quad (3)$$

With this definition, we state the first auxiliary Lemma.

Lemma 1. *Let $\mathcal{X}$ be compact and let Assumption 2(i) hold. Suppose that*

$$\mathbf{E}_\eta[\|m(Y, X, \theta)\|_\infty] < \infty \quad \text{for all } (\theta, \eta) \in \Theta \times \Gamma.$$

Then, for every $(\theta, \eta) \in \Theta \times \Gamma$,

$$\mathbf{E}_\eta[m(Y, X, \theta) \mid X] = 0 \quad \eta_X\text{-a.s.} \quad \Longleftrightarrow \quad \int m_j(y, x, \theta) g_k(x) d\eta = 0 \quad \text{for every } (j, k).$$

Consequently, $\Theta_I = \Theta_I^*$.

Furthermore given our weights $\omega_k > 0$, one could equivalently define Θ_I and this lemma using the unweighted class $\mathcal{H}$ with elements $h_k(x)$ instead of $g_k(x)$.

It is noted that the compactness and integrability conditions in Lemma 1 are ensured by Assumptions 1 and 3, respectively. Lemma 1 is the well-known property that a conditional moment equality is equivalent to an infinite family of unconditional moment equalities, and it isolates the central feasibility requirement: one admissible completion must satisfy every unconditional restriction simultaneously. Aucejo, Bugni, and Hotz (2017) derive a similar result in their Lemma 1 and note, as we do, that verifying the existence of such an η is extremely difficult and hard to make progress on in this direct formulation. Andrews and Shi (2017) and Chernozhukov, Chetverikov, and Kato (2017) likewise work with infinite families of moment equalities, but typically assume that the variables entering those moments are observable, whereas here η is not identified. The proof is in Appendix A. An immediate implication is instrument-class invariance: every class satisfying Assumption 2 generates the same population set Θ_I^* .

Lemma 1 removes searching over the unobserved conditioning event in the original model and thus makes a random-set formulation possible. For each candidate θ , making use of our weighted instruments $\mathcal{G}$, we collect all moments in an ℓ_1 -valued random set and ask whether

its expectation contains zero. This logic follows the seminal paper Beresteanu, I. Molchanov, and Molinari (2011). We extend their framework to accommodate an infinite-dimensional moment vector and a discontinuous instrument map. In general, random-set methods were introduced to econometrics by Beresteanu and Molinari (2008); see Molchanov (2017) and I. Molchanov and Molinari (2018) for textbook treatments.

For $\theta \in \Theta$, define

$$M_\theta(y, x) \equiv (m_j(y, x, \theta)g_k(x))_{j=1, \dots, d_m, k \geq 1} \in \ell_1.$$

Let $\eta_0 = \mathcal{L}_{F_0}(W, W^O)$ denote the true joint law of the latent and observed variables. By Lemma 1, the model in (1) can equivalently be written as

$$\int_{\mathcal{W} \times \mathcal{W}^O} M_{\theta_0}(y, x) \eta_0(dy, dx, dw^O) = 0.$$

Assumption 1 implies $\eta_0 \in \Gamma$, but the observed data generally do not determine which element of Γ is the true joint law. For each candidate θ , we therefore consider the set of attainable mean-moment vectors

$$I_\theta \equiv \left\{ \int_{\mathcal{W} \times \mathcal{W}^O} M_\theta(y, x) \eta(dy, dx, dw^O) : \eta \in \Gamma \right\}. \quad (4)$$

Thus,

$$\theta \in \Theta_I \iff 0 \in I_\theta.$$

Checking this condition directly requires finding an admissible joint law that satisfies all moment restrictions simultaneously. To obtain a more tractable characterization, we first

collect the moment vectors compatible with each observed interval:

$$Q_\theta(W^O) \equiv \text{cl}_{\ell_1} \{M_\theta(y, x) : (y, x) \in D(W^O)\}.$$

The set $Q_\theta(W^O)$ contains the moment vectors generated by feasible latent values, together with their limits in ℓ_1 . We next show how its Aumann expectation characterizes whether I_θ contains zero, leading to a support-function criterion evaluated under the observed-data distribution. We note $Q_\theta(W^O)$ is directly observable by the data generating process.

Assumption 3.

- (i) For every $\theta \in \Theta$, the map $(y, x) \mapsto m(y, x, \theta)$ is continuous on $\mathcal{W}$.
- (ii) There exists a constant $C_1 < \infty$ such that

$$\sup_{(y, x, \theta) \in \mathcal{W} \times \Theta} \|m(y, x, \theta)\|_\infty \leq C_1.$$

Assumption 3 is satisfied by common linear-mean, logit, and probit specifications on bounded domains. Because Assumption 1 makes $\mathcal{W}$ compact, continuity already gives boundedness for each fixed θ ; Assumption 3 supplies a common bound over Θ and therefore keeps the random sets integrably bounded under the conditions of Lemma 1. The Aumann expectations below are understood as the collection of expectations of ordinary random variables (Aumann 1965).

Theorem 1. *Let Assumptions 1–3 hold. Then $Q_\theta(W^O)$ is almost surely a nonempty compact random set in ℓ_1 , and $\mathbb{E}[Q_\theta(W^O)]$ is a nonempty closed convex subset of ℓ_1 . Moreover,*

$$\theta \in \Theta_I^* \iff 0 \in \mathbb{E}[Q_\theta(W^O)]. \tag{5}$$

Therefore,

$$\begin{aligned}\Theta_I^* &= \{\theta \in \Theta : 0 \in \mathbb{E}[Q_\theta(W^O)]\} \\ &= \{\theta \in \Theta : \mathbf{E}_{P_{W^O}}[h(Q_\theta(W^O), u)] \geq 0 \quad \forall u \in B^p\}.\end{aligned}\tag{6}$$

$$= \left\{ \theta \in \Theta : \inf_{u \in B^p} \mathbf{E}_{P_{W^O}}[h(Q_\theta(W^O), u)] \geq 0 \right\}.\tag{7}$$

$$= \left\{ \theta \in \Theta : \inf_{u \in B^p} \mathbf{E}_{P_{W^O}} \left[\sup_{(x,y) \in D(W^O)} \langle M_\theta(y, x), u \rangle \right] \geq 0 \right\}\tag{8}$$

Where $B^p := \{u \in \ell_p : \|u\|_p \leq 1\}$, for every $p \in [1, \infty]$

The proof is in Appendix B. Here we only sketch the arguments. First, we establish that the random-set construction is well defined and we relate it to the original feasibility problem. The appendix verifies the required random-set machinery that Q_θ is measurable and integrably bounded. The Aumann expectation $\mathbb{E}[Q_\theta(W^O)]$ is therefore well defined and nonempty; it is closed by definition and convex.²

We next connect the random set to the original identification problem. By Lemma 1 and the definition of I_θ :

$$\theta \in \Theta_I^* \iff \theta \in \Theta_I \iff 0 \in I_\theta.$$

The law-selection lemma in the appendix provides the formal bridge between the random set formulation and Θ_I^* .

The remaining step is to relate the attainable set I_θ to the Aumann expectation of $Q_\theta(W^O)$. This is also where the main infinite-dimensional subtlety enters. If M_θ were continuous, compactness of $D(W^O)$ would imply that $M_\theta(D(W^O))$ is already closed, and the means of selections of Q_θ could be matched directly with admissible completion laws. Here, how-

²Aumann expectations on all appropriate non-atomic probability spaces are convex. Non atomicity is often a trivial requirement. Without loss of generality the Aumann expectation is taken on the uniform enrichment of the original probability space described in Appendix B. A similar construction is used in the proof of the Artstein Selection Theorem in Molchanov (2017).

ever, the instruments g_k need only be Borel measurable. Thus M_θ may be discontinuous and $M_\theta(D(W^O))$ need not be closed. Consequently, one cannot in general assert that $\mathbb{E}[Q_\theta(W^O)] = I_\theta$. Instead, the Appendix proves $\mathbb{E}[Q_\theta(W^O)] = \text{cl}_{\ell_1} I_\theta$.

In general, $\text{cl}_{\ell_1} I_\theta$ may contain vectors that are not themselves generated by any admissible completion. That is $\text{cl}_{\ell_1} I_\theta \neq I_\theta$; identification, however, only asks whether the origin is attainable. The key additional lemma establishes that $0 \in \text{cl}_{\ell_1} I_\theta \iff 0 \in I_\theta$, i.e. closure is harmless precisely at this point.

Combining the preceding steps gives

$$\theta \in \Theta_I^* \iff 0 \in I_\theta \iff 0 \in \text{cl}_{\ell_1} I_\theta \iff 0 \in \mathbb{E}[Q_\theta(W^O)].$$

Having completed the difficult work, as in Beresteanu and Molinari (2008) or Beresteanu, I. Molchanov, and Molinari (2011) the Hahn–Banach separation theorem implies that the Aumann expectation contains zero if and only if its support function is nonnegative in every $u \in B_1^p$. The support-function identity for Aumann expectations gives

$$h(\mathbb{E}[Q_\theta(W^O)], u) = \mathbf{E}_{P_{W^O}} [h(Q_\theta(W^O), u)],$$

which yields the characterization in Theorem 1.

Theorem 1 transformed the original problem into a geometric problem which involves a nested minimax optimization problem,

$$\Theta_I^* = \{\theta \in \Theta : \inf_{u \in B^p} \mathbf{E}_{P_{W^O}} [\sup_{(y,x) \in D(W^O)} \langle M_\theta(y, x), u \rangle] \geq 0\}$$

as the support function searches over feasible pairs (y, x) , which is still a complex characterization from a computational point of view. However, two properties of the characterization make it appealing.

First, Theorem 1 defines the identified set as a collection of infinite inequalities only de-

pending on observables. Thus, existing inference procedures, e.g. Andrews and Shi (2017) or Chernozhukov, Chetverikov, and Kato (2019), can be used for consistency and inference arguments. We discuss the details in the following section.

Second, although the general problem is computationally demanding, several observations simplify the problem substantially. First, Theorem 1 implies that we can exclude parameter candidates by finding counterexample-directions. In practice, this means we can search a large parameter space effectively by first screening out candidates based on a collection of fixed directions. Second, as we will show in the respective implementation section, the nested optimization problem reduces to a linear program for the important subclass of monotone index models.

While the dual space of ℓ_1 is ℓ_∞ , the compactness of $\mathbb{E}[Q_\theta(W^O)]$ permits the use of unit balls B^p other than B^∞ . These choices are equivalent for identification, although some may be more convenient computationally. As the related literature discussed in Section 1 makes clear, our formulation also has computational benefits and drawbacks. In particular, the scale of $\langle M_\theta(y, x), u \rangle$ has no intrinsic interpretation beyond determining whether zero belongs to $\mathbb{E}[Q_\theta(W^O)]$; we do not assign substantive or finite-sample meaning to a positive distance from zero. Nor do we claim that our choice of instrument class $\mathcal{G}$, norm $\|\cdot\|_p$, or support-function representation is optimal for estimation or the construction of critical values. We favor the formulation for its geometric clarity:

$$\theta \in \Theta_I^* \iff 0 \in \mathbb{E}[Q_\theta(W^O)].$$

Section 3 develops its computational and inferential implementation.

3 Estimation and Computation

3.1 Estimator and Confidence Sets

Theorem 1 is pointwise in θ and requires neither compactness of Θ^3 nor continuity of $m(\cdot)$ in θ . The following assumptions enter the sample results. Compactness of Θ in Assumption 4 and the uniform Lipschitz bound in Assumption 5 control the sample criterion uniformly over θ , as required for the Hausdorff-consistency result in Theorem 2. The coverage proof for Theorem 3 uses a common moment envelope and the decay of the weighted instruments to obtain relative compactness of the covariance kernels over $(\theta, P) \in \mathcal{F}$ for which Assumptions 1–3 are sufficient. Its sequential argument takes limits of covariance kernels and requires no continuity of m in θ .

Assumption 4. *The parameter space Θ is a nonempty compact subset of $\mathbb{R}^{d_\theta}$.*

Assumption 5. *$m(y, x, \theta)$ is uniformly Lipschitz continuous. That is there exists a constant $L < \infty$ such that, for every $\theta, \theta' \in \Theta$,*

$$\sup_{(y,x) \in \mathcal{W}} \|m(y, x, \theta) - m(y, x, \theta')\|_\infty \leq L \|\theta - \theta'\|.$$

The norm over Θ is inconsequential.

Theorem 1 characterizes the sharp identified set through the population criterion

$$T(\theta) \equiv - \inf_{u \in B^p} \mathbf{E} [h(Q_\theta(W^O), u)] .$$

Since $0 \in B^p$, $T(\theta) \geq 0$, and Theorem 1 implies $\Theta_I^* = \{\theta \in \Theta : T(\theta) = 0\}$.

The criterion $T(\theta)$ is not directly computable because it involves countably many moment coordinates and an infinite-dimensional set of directions. Also, we emphasize that the “scale” of $T(\theta)$ is not particularly favorable compared with those of other methods in the related

³In fact, no restriction is put on Θ whatsoever for the purposes of identification.

literature. For estimation, we approximate the conditional moment restriction by retaining only a finite number of instruments. Let $\pi_K : \ell_1 \rightarrow \mathbb{R}^{d_m \cdot K}$, denote the truncation map that selects the coordinates indexed by (j, k) , $j = 1, \dots, d_m$, $k = 1, \dots, K$. Let $\lambda_K : \mathbb{R}^{d_m \cdot K} \rightarrow \ell_1$, denote the zero-padding embedding that maps a finite vector into ℓ_1 by keeping its first $d_m K$ coordinates and setting all remaining coordinates equal to zero. Define the various truncated moment mapping (random vectors) and truncated random sets as⁴

$$\begin{aligned} M_{\theta, K}(y, x) &\equiv \pi_K M_{\theta}(y, x) \\ Q_{\theta, K}(w^o) &\equiv \text{cl}_{\mathbb{R}^{d_m K}} M_{\theta, K}(D(w^o)) = \pi_K Q_{\theta}(w^o) \end{aligned}$$

Let $B_K^p \equiv \{u \in \mathbb{R}^{d_m K} : \|u\|_p \leq 1\}$. The finite-dimensional population criterion is defined as

$$T_K(\theta) \equiv \vec{\rho}_H(0, \mathbb{E}[Q_{\theta, K}(W^O)]) = - \inf_{u \in B_K^p} h(\mathbb{E}[Q_{\theta, K}(W^O)], u) = - \inf_{u \in B_K^p} \mathbf{E}[h(Q_{\theta, K}(W^O), u)].$$

The corresponding finite-dimensional outer approximation to the identified set is $\Theta_{I, K} \equiv \{\theta \in \Theta : T_K(\theta) = 0\}$.

Since every admissible completion satisfying the conditional moment restriction also satisfies any finite collection of the unconditional moment restrictions, $\Theta_I^* \subseteq \Theta_{I, K}$ for every K .

Given an i.i.d. sample $\{W_i^O\}_{i=1}^n$, define the sample Aumann average⁵

$$\widehat{\mathbb{E}}_n[Q_{\theta, K}] \equiv \text{co} \left(\frac{1}{n} \bigoplus_{i=1}^n Q_{\theta, K}(W_i^O) \right)$$

⁴For comparisons with the infinite-dimensional set $Q_{\theta}(w^o) \subset \ell_1$, and to ensure that set distances, such as the Hausdorff distance, are well defined, we require the embedding

$$\tilde{Q}_{\theta, K}(w^o) = \lambda_K Q_{\theta, K}(w^o) = \lambda_K \pi_K Q_{\theta}(w^o),$$

where π_K and λ_K applied to sets denote their image under the corresponding point maps.

⁵In contrast to the population Aumann expectation, the finite-sample Minkowski average need not be convex in the most general case. We therefore convexify it in defining its empirical analogue. This does not alter any support-function inequality because $h(A, u) = h(\text{co}(A), u)$ for every u .

The sample criterion is

$$\widehat{T}_{n,K}(\theta) \equiv \vec{\rho}_H \left(0, \widehat{\mathbb{E}}_n[Q_{\theta,K}] \right) = - \inf_{u \in B_K^p} h \left(\widehat{\mathbb{E}}_n[Q_{\theta,K}], u \right) = - \inf_{u \in B_K^p} \frac{1}{n} \sum_{i=1}^n h(Q_{\theta,K}(W_i^O), u).$$

For a deterministic sequence $K(n) \rightarrow \infty$, we use $\widehat{T}_n(\theta) \equiv \widehat{T}_{n,K(n)}(\theta)$. Thus an estimator based on the criterion function of distance between 0 and $\mathbb{E}[Q_{\theta,K(N)}(W^O)]$ is the contour set

$$\widehat{\Theta}_n \equiv \left\{ \theta \in \Theta : a_n \widehat{T}_{n,K(n)}(\theta) \leq \tau_n \right\}.$$

With this we can state the result concerning consistency of the proposed estimator.

Theorem 2. *Let Assumptions 1–5 hold. Suppose that $\Theta_I^* \neq \emptyset$. Let $K(n) \rightarrow \infty$, and let $a_n \equiv \sqrt{\frac{n}{d_m K(n) + d_\theta}}$. Let $\frac{d_m K(n)}{n} \rightarrow 0$, $\tau_n \rightarrow \infty$ and $\frac{\tau_n}{a_n} \rightarrow 0$. Then $\rho_H(\widehat{\Theta}_n, \Theta_I^*) \xrightarrow{p} 0$.*

For the proof of Theorem 2 we first establish a law of large numbers for random sets with growing dimension; see Appendix C. The estimator is built on the result of Chernozhukov, Hong, and Tamer (2007) for constructing set estimators using a criterion function approach. The sequence $K(n)$ controls the finite-dimensional approximation of the infinite collection of unconditional moment restrictions. Since each of the $K(n)$ retained instruments is interacted with the d_m components of the moment function, the dimension of the finite random set is $d_m K(n)$. $K(n) \rightarrow \infty$ makes the deterministic truncation error vanish, while $d_m K(n)/n \rightarrow 0$ ensures that the sampling error from estimating this growing-dimensional random set vanishes. Thus $K(n)$ grows to include more moment restrictions, but slowly enough to control sampling error. We note that since d_m and d_θ are fixed, the rate conditions could be equivalently stated as $\frac{K(n)}{n} \rightarrow 0$ and $a_n \asymp \sqrt{\frac{n}{K(n)}}$. The estimator above is obtained by accepting parameter values for which the finite-sieve sample violation is small.

3.2 Inference

We construct confidence sets using the many-moment procedure of Andrews and Shi (2017). The procedure is well suited to the present setting because the support-function inequalities are indexed by the uncountable direction class B^p , however we note that alternative procedures exist, e.g. Chernozhukov, Chetverikov, and Kato (2017) which might be also applied to our results. The Andrews and Shi (2017) implementation requires a method for evaluating the supremum over B^p ; the computational section discusses finite and truncated approximations to this problem. We do not consider here the additional error caused by terminating a global optimization routine at an inexact solution, but such ϵ -minimax solutions may be considered Aradillas Fernández et al. (2025). The managability of the relevant function class is at the moment only verified for the case of $p = \infty$, thus we restrict attention to this case in this section.

For the closed infinite-dimensional unit ball under the $\|\cdot\|_p$ norm, write $u = (u_{jk}) \in B^p$, with coordinates indexed in the same order as M_θ . Define

$$\tilde{m}(w^o, \theta, u) \equiv h(Q_\theta(w^o), u) = \sup_{(y,x) \in D(w^o)} \sum_{j=1}^{d_m} \sum_{k \geq 1} u_{jk} g_k(x) m_j(y, x, \theta).$$

Let

$$\mathcal{U} \equiv B^p = \{u = (u_{jk})_{j=1, \dots, d_m, k \geq 1} : \|u\|_p \leq 1\}.$$

Thus each $u \in \mathcal{U}$ is one direction for the complete stacked moment vector and indexes one scalar support-function inequality. The same feasible (y, x) enters every term of the sum inside the supremum as the latent completion must solve all moments simultaneously. We can slightly rewrite Theorem 1 to bring it into the notation of Andrews and Shi (2017)

$$\Theta_I = \Theta_I^* = \{\theta \in \Theta : \mathbf{E}_{P_{w^o}}[\tilde{m}(W^O, \theta, u)] \geq 0 \quad \forall u \in \mathcal{U}\}.$$

The null hypothesis that $\theta \in \Theta_I^*$ can be stated as

$$H_0(\theta) : \quad \mathbf{E}_{P_{WO}}[\tilde{m}(W^O, \theta, u)] \geq 0 \quad \text{for every } u \in \mathcal{U}.$$

Finite direction classes and their inferential interpretation are discussed below. The auxiliary conditioning-instrument class in Andrews and Shi (2017) is not needed here as the functions g_k already encode the original conditioning restriction of model 1 through Lemma 1, and the support direction u indexes the resulting unconditional inequalities.⁶

Following Andrews and Shi (2017), we use the scalar normalization $\sigma_F(\theta) = \hat{\sigma}_n(\theta) \equiv 1$.

Define the sample mean

$$\bar{m}_n(\theta, u) \equiv \frac{1}{n} \sum_{i=1}^n \tilde{m}(W_i^O, \theta, u),$$

the corresponding sample variance

$$\hat{\Sigma}_n(\theta, u) \equiv \frac{1}{n} \sum_{i=1}^n (\tilde{m}(W_i^O, \theta, u) - \bar{m}_n(\theta, u))^2,$$

and its regularized version

$$\bar{\Sigma}_n(\theta, u) \equiv \hat{\Sigma}_n(\theta, u) + \epsilon, \quad \epsilon > 0.$$

The fixed ridge is the scalar specialization of the Andrews and Shi (2017) regularization.

Since $\sigma_F(\theta) = \hat{\sigma}_n(\theta) \equiv 1$, we use their benchmark choice $\epsilon = 1/20$.⁷

For the single lower-bound moment inequality indexed by $u \in \mathcal{U}$, the S_1 function of Andrews

⁶See the first paragraph of page 276 in Andrews and Shi (2017).

⁷A direction-adaptive alternative for the regularization would be an alternative. This addresses the concern that, for directions with very small support-function variance, a fixed ridge may dominate the empirical variance. The alternative is an active question, however, because the ridge can approach zero over the direction class. Its validity requires an additional argument.

and Shi (2017) reduces to

$$S_1(s, \Sigma) \equiv \left[\frac{s}{\sqrt{\Sigma}} \right]_-^2, \quad [x]_- \equiv \max\{-x, 0\}.$$

The negative part appears as our null hypothesis H_0 is that each moment should be weakly greater than zero. The corresponding test statistic is

$$T_n^{AS}(\theta) \equiv \sup_{u \in \mathcal{U}} S_1(\sqrt{n} \bar{m}_n(\theta, u), \bar{\Sigma}_n(\theta, u)).$$

We use the generalized moment-selection critical value of Andrews and Soares (2010), as implemented by Andrews and Shi (2017).

Step 1. Set

$$\kappa_n \equiv (0.3 \log n)^{1/2}, \quad B_n \equiv \left(\frac{0.4 \log n}{\log \log n} \right)^{1/2},$$

and define the scalar standardized slackness measure

$$\xi_n(\theta, u) \equiv \kappa_n^{-1} \frac{\sqrt{n} \bar{m}_n(\theta, u)}{\sqrt{\bar{\Sigma}_n(\theta, u)}}$$

The corresponding GMS adjustment is

$$\bar{\varphi}_n(\theta, u) \equiv \sqrt{\bar{\Sigma}_n(\theta, u) B_n} \mathbf{1}\{\xi_n(\theta, u) > 1\}.$$

Step 2. Generate B_{boot} nonparametric bootstrap samples $\{W_{i,b}^{O,*} : i = 1, \dots, n\}, b = 1, \dots, B_{\text{boot}},$

by sampling with replacement from the empirical distribution. For every b , calculate

$$\begin{aligned}\bar{m}_{n,b}^*(\theta, u) &\equiv \frac{1}{n} \sum_{i=1}^n \tilde{m}(W_{i,b}^{O,*}, \theta, u), \\ \hat{\Sigma}_{n,b}^*(\theta, u) &\equiv \frac{1}{n} \sum_{i=1}^n \left\{ \tilde{m}(W_{i,b}^{O,*}, \theta, u) - \bar{m}_{n,b}^*(\theta, u) \right\}^2, \\ \bar{\Sigma}_{n,b}^*(\theta, u) &\equiv \hat{\Sigma}_{n,b}^*(\theta, u) + \epsilon.\end{aligned}$$

Step 3. Form the recentered bootstrap statistic

$$T_{n,b}^{AS,*}(\theta) \equiv \sup_{u \in \mathcal{U}} \left[\frac{\sqrt{n} \{ \bar{m}_{n,b}^*(\theta, u) - \bar{m}_n(\theta, u) \} + \bar{\varphi}_n(\theta, u)}{\sqrt{\bar{\Sigma}_{n,b}^*(\theta, u)}} \right]^2.$$

Step 4. Let $q_{n,1-\alpha+\eta}^{AS,*}(\theta)$ be the empirical $(1 - \alpha + \eta)$ -quantile of $\{T_{n,b}^{AS,*}(\theta) : b = 1, \dots, B_{\text{boot}}\}$ and define

$$c_{n,1-\alpha}^{AS}(\theta) \equiv q_{n,1-\alpha+\eta}^{AS,*}(\theta) + \eta, \quad 0 < \eta < \alpha.$$

The parameter η is part of the infinitesimal uniformity adjustment in Andrews and Shi (2017). The theoretical results retain a fixed $0 < \eta < \alpha$. Following their implementation convention, η can be set equal to zero in computation.

Remark 1. For the asymptotic results, $q_{n,1-\alpha+\eta}^{AS,*}(\theta)$ denotes the conditional $(1 - \alpha + \eta)$ -quantile of the bootstrap statistic given the original sample. The empirical quantile computed from B_{boot} bootstrap replications is its Monte Carlo approximation; simulation error is not included in the theoretical coverage statement.

The confidence set obtained by test inversion is

$$CS_n^{AS}(1 - \alpha) \equiv \{ \theta \in \Theta : T_n^{AS}(\theta) \leq c_{n,1-\alpha}^{AS}(\theta) \}.$$

The inference result requires a strengthening of the tail condition in Assumption 2 (ii). Assumption 2 (ii) is sufficient to ensure that the stacked moment map is ℓ_1 -valued and that finite truncations approximate the full moment vector uniformly. To control the complexity of the infinite-dimensional class of support directions, we additionally require a polynomial decay rate for the individual instrument coordinates. Define

$$a_k \equiv \sup_{x \in \mathcal{X}} |g_k(x)|, \quad k \geq 1.$$

Because the weighting sequence is already incorporated into the instruments g_k , a_k is the uniform envelope of the k -th weighted instrument. The polynomial decay condition below is stronger than Assumption 2 (i) in general and provides the entropy bound required for the manageability argument.

For the uniform coverage statement, we introduce additional notation. Let $\mathcal{P}$ denote the class of observed-data laws over which inference is required to be valid. For each $P \in \mathcal{P}$, let $\Theta_I^*(P)$ denote the identified set under P . By Theorem 1,

$$\Theta_I^*(P) = \{\theta \in \Theta : \mathbf{E}_P [\tilde{m}(W^O, \theta, u)] \geq 0 \quad \forall u \in \mathcal{U}\}.$$

Define $\mathcal{F}^+ \equiv \Theta \times \mathcal{P}$ and $\mathcal{F} \equiv \{(\theta, P) \in \mathcal{F}^+ : \theta \in \Theta_I^*(P)\}$.

Theorem 3. *Let $\mathcal{P}$ be a class of laws on the $\mathcal{W}^o$. Let Assumptions 1–3 hold. Suppose, in addition, that for some $\beta > 3/2$, $C_a \equiv \sup_{k \geq 1} k^\beta a_k < \infty$. Then, for every $\alpha \in (0, 1)$,*

$$\liminf_{n \rightarrow \infty} \inf_{(\theta, P) \in \mathcal{F}} P^n \{\theta \in CS_n^{AS}(1 - \alpha)\} \geq 1 - \alpha.$$

Moreover, for every fixed $P \in \mathcal{P}$ and every fixed $\theta \notin \Theta_I^*(P)$, $P^n \{T_n^{AS}(\theta) > c_{n, 1-\alpha}^{AS}(\theta)\} \rightarrow 1$.

The restriction $\beta > 3/2$ ensures that the entropy integral for the infinite-dimensional direction class is finite. The proof is given in Appendix D. Theorem 3 gives uniformly valid marginal coverage over all null pairs $(\theta, P) \in \mathcal{F}$. It does not imply simultaneous coverage of

the entire identified set $\Theta_I^*(P)$ or Hausdorff convergence of a fixed-level confidence set.

4 Implementation

4.1 Exact LP Representation for Monotone-Index Models

For a subclass of monotone-index models, such as linear, probit, or logit regressions, and for a finite truncation of a specifically chosen instrument class of weighted cubes, defined below, the inner optimization problem from Theorem 1 can be written as a linear program (LP). This substantially reduces computational time. Throughout this subsection we assume $d_m = 1$ for the LP argument below. The joint-support formulation extends to vector-valued moments, but the exact LP representation requires a separate argument. We focus on the instrument class of weighted cubes introduced in Appendix A.2 for simplicity, the arguments generalizes to any instrument class defined by a partition of the covariate space.

Following Andrews and Shi (2017), transform the conditioning variables so that $\mathcal{X} \subseteq [0, 1]^{d_x}$. For each resolution level r , let the cube instruments be $g_{a,r}(x) = w(r)\mathbf{1}\{x \in C_{a,r}\}$, where

$$C_{a,r} = \prod_{j=1}^{d_x} ((a_j - 1)/(2r), a_j/(2r)], \quad a_j \in \{1, \dots, 2r\},$$

and where $w(r) > 0$ denotes the resolution weight, for example $w(r) = (c_0 + N_r^\beta)^{-1}$ with $N_r = \sum_{s=1}^r (2s)^{d_x}$, $c_0 > 0$ and $\beta > 3/2$. The choice $c_0 = 100$ and $\beta = 2$ retains the tuning constants in Andrews and Shi (2013), with the cumulative index N_r used here to satisfy the enumerated decay condition. The cubes form a partition of $[0, 1]^{d_x}$, with zero included in the first interval of each coordinate.

We first describe the construction for $d_x = 1$. Let the finite truncation use resolution levels $r = 1, \dots, R$. For each resolution r , write

$$\mathcal{B}_{rj} \equiv (b_{r,j-1}, b_{rj}], \quad j = 1, \dots, K_r, \quad 0 = b_{r0} < b_{r1} < \dots < b_{r,K_r} = 1,$$

and define

$$g_j^{(r)}(x) \equiv w_{rj} \mathbf{1}\{x \in \mathcal{B}_{rj}\}, \quad g(x) \equiv (g^{(1)}(x), \dots, g^{(R)}(x)) \in \mathbb{R}^J, \quad J \equiv \sum_{r=1}^R K_r.$$

For the weighted cube class above, $K_r = 2r$ and $w_{rj} = w(r)$ for all $j = 1, \dots, K_r$.

It is useful to collapse the multi-resolution partition into one common refined partition. Let

$$\mathcal{C} \equiv \{0, 1\} \cup \bigcup_{r=1}^R \{b_{r0}, b_{r1}, \dots, b_{r, K_r}\},$$

and let its sorted distinct elements be $0 = \xi_0 < \xi_1 < \dots < \xi_L = 1$. Define the refined bins

$$\tilde{\mathcal{B}}_\ell \equiv (\xi_{\ell-1}, \xi_\ell], \quad \ell = 1, \dots, L,$$

with the same boundary convention at zero, and let

$$\tilde{g}_\ell(x) \equiv \mathbf{1}\{x \in \tilde{\mathcal{B}}_\ell\}, \quad \tilde{g}(x) \equiv (\tilde{g}_1(x), \dots, \tilde{g}_L(x)).$$

For each original bin (r, j) define $I_{rj} \equiv \{\ell : \tilde{\mathcal{B}}_\ell \subseteq \mathcal{B}_{rj}\}$. Then

$$\mathbf{1}\{x \in \mathcal{B}_{rj}\} = \sum_{\ell \in I_{rj}} \tilde{g}_\ell(x), \quad g_j^{(r)}(x) = \sum_{\ell \in I_{rj}} w_{rj} \tilde{g}_\ell(x).$$

Consequently, for any $u \in [-1, 1]^J$, define the induced refined-grid direction $\tilde{u} \in \mathbb{R}^L$ by

$$\tilde{u}_\ell \equiv \sum_{r=1}^R \sum_{j=1}^{K_r} u_{rj} w_{rj} \mathbf{1}\{\ell \in I_{rj}\}, \quad \ell = 1, \dots, L.$$

Then

$$u^\top g(x) = \tilde{u}^\top \tilde{g}(x) \quad \text{for all } x \in [0, 1],$$

under the boundary convention above. Hence every linear combination of the multi-resolution weighted indicators is exactly a step function on the refined grid.

Let $Q_{\theta,i}^{[R]}$ denote the finite-dimensional truncation of $Q_{\theta,i}$ to the instruments included up to resolution R . The observed interval for observation i is

$$D_i \equiv [y_{iL}, y_{iU}] \times [x_{iL}, x_{iU}].$$

For each refined bin $\tilde{\mathcal{B}}_\ell = (c_{\ell-1}, c_\ell]$, define the truncated endpoints

$$x_{i\ell}^L \equiv \max\{x_{iL}, c_{\ell-1}\}, \quad x_{i\ell}^U \equiv \min\{x_{iU}, c_\ell\}.$$

Let

$$S_i \equiv \left\{ \ell \in \{1, \dots, L\} : [x_{iL}, x_{iU}] \cap \tilde{\mathcal{B}}_\ell \neq \emptyset \right\}$$

denote the refined bins intersecting the observed x -interval.

For fixed θ , suppose that $m(\cdot, \cdot, \theta)$ is monotone on each rectangle

$$[y_{iL}, y_{iU}] \times [x_{i\ell}^L, x_{i\ell}^U], \quad \ell \in S_i.$$

Then it suffices to evaluate m at the four corners of this rectangle. Define

$$M_{i\ell}(\theta) \equiv \left\{ m(y_{iL}, x_{i\ell}^L, \theta), m(y_{iL}, x_{i\ell}^U, \theta), m(y_{iU}, x_{i\ell}^L, \theta), m(y_{iU}, x_{i\ell}^U, \theta) \right\},$$

and set

$$m_{i\ell}^{\max}(\theta) \equiv \max M_{i\ell}(\theta), \quad m_{i\ell}^{\min}(\theta) \equiv \min M_{i\ell}(\theta).$$

The corner selected by the support function depends on the sign of $\tilde{u}_\ell$. Hence, for $\ell \in S_i$,

$$\sup_{(y,x) \in [y_{iL}, y_{iU}] \times [x_{i\ell}^L, x_{i\ell}^U]} \tilde{u}_\ell m(y, x, \theta) = \max \{ \tilde{u}_\ell m_{i\ell}^{\max}(\theta), \tilde{u}_\ell m_{i\ell}^{\min}(\theta) \}.$$

Since the refined bins form a partition, the observation-level support function is

$$h_i(Q_{\theta,i}^{[R]}, u) = \max_{\ell \in S_i} \max \{ \tilde{u}_\ell m_{i\ell}^{\max}(\theta), \tilde{u}_\ell m_{i\ell}^{\min}(\theta) \}.$$

This is the maximum of finitely many linear functions in u , because $\tilde{u}_\ell$ is linear in u .

Define the finite-truncation support minimum

$$\hat{s}_n^{[R]}(\theta) \equiv \inf_{u \in [-1, 1]^J} \frac{1}{n} \sum_{i=1}^n h_i(Q_{\theta,i}^{[R]}, u).$$

The criterion used in the inference section is the sign-normalized statistic $\hat{T}_n^{[R]}(\theta) \equiv -\hat{s}_n^{[R]}(\theta)$.

Thus $\hat{s}_n^{[R]}(\theta)$ is the object computed by the following linear program, while $\hat{T}_n^{[R]}(\theta)$ is its negative. In particular, a parameter value is retained by the finite-truncation estimator whenever

$$\hat{s}_n^{[R]}(\theta) \geq -\frac{\tau_n}{a_n}, \quad \text{equivalently} \quad a_n \hat{T}_n^{[R]}(\theta) \leq \tau_n.$$

Using an epigraph reformulation, $\hat{s}_n^{[R]}(\theta)$ is the value of the linear program

$$\begin{aligned} \min_{u, \tilde{u}, t} \quad & \frac{1}{n} \sum_{i=1}^n t_i \\ \text{s.t.} \quad & -1 \leq u_{rj} \leq 1 \quad \forall r = 1, \dots, R, \quad \forall j = 1, \dots, K_r, \\ & \tilde{u}_\ell = \sum_{r=1}^R \sum_{j=1}^{K_r} u_{rj} w_{rj} \mathbf{1}\{\ell \in I_{rj}\} \quad \forall \ell = 1, \dots, L, \\ & t_i \geq \tilde{u}_\ell m_{i\ell}^{\max}(\theta), \quad t_i \geq \tilde{u}_\ell m_{i\ell}^{\min}(\theta) \quad \forall i = 1, \dots, n, \quad \forall \ell \in S_i. \end{aligned}$$

The LP has J direction variables, L refined-grid direction variables, and n epigraph variables. The number of inequality constraints is proportional to $\sum_{i=1}^n |S_i|$, up to the factor two coming from the upper and lower corner values.

Finally, if $d_x > 1$, the same construction applies after replacing refined bins by refined rectangles. The common refinement is obtained by taking the product grid induced by all coordinate-wise cube boundaries across all included resolutions. For scalar Y and d_x -dimensional X , monotonicity implies that the inner maximization over each refined rectangle is attained at one of 2^{1+d_x} corners of

$$[y_{iL}, y_{iU}] \times \prod_{k=1}^{d_x} [x_{ik\ell}^L, x_{ik\ell}^U].$$

The epigraph formulation is unchanged: one introduces one variable t_i per observation and imposes the corresponding linear inequalities for each relevant refined cell and each relevant corner value.

4.2 Construction of Outer Sets

Theorem 1 characterizes Θ_I^* through the inequalities $\mathbf{E}_{P_{w^O}}[\tilde{m}(W^O, \theta, u)] \geq 0$ for all directions $u \in \mathcal{U} = B_1^\infty$, where $\tilde{m}(w^o, \theta, u) = h(Q_\theta(w^o), u)$. Any subcollection of these inequalities defines an *outer set*, that is, a set containing Θ_I^* . Outer sets are cheaper to compute and inherit valid inference (Corollary 2 below). They can be tightened gradually by adding directions, at a purely computational cost. For $\mathcal{V} \subseteq \mathcal{U}$, define

$$\Theta_O(\mathcal{V}) \equiv \{\theta \in \Theta : \mathbf{E}_{P_{w^O}}[\tilde{m}(W^O, \theta, u)] \geq 0 \quad \forall u \in \mathcal{V}\}.$$

Truncation of the instrument class is a special case. For $u \in \mathcal{U}$, let $\Pi_K u \equiv \lambda_K \pi_K u$ retain the first K instrument coordinates in each moment block and set the remaining coordinates to zero, and let $B_K \equiv \Pi_K \mathcal{U}$. Since $h(Q_\theta(w^o), \Pi_K u) = h(Q_{\theta, K}(w^o), \pi_K u)$, inequalities indexed by $\mathcal{V} \subseteq B_K$ involve only the finite-dimensional sets $Q_{\theta, K}$. In particular, $\Theta_O(B_K)$ coincides

with the outer set $\Theta_{I,K}$ of Section 3.1. We write $\Theta_O^K(\mathcal{V}) \equiv \Theta_O(\Pi_K \mathcal{V})$.

Corollary 1. *For all $\mathcal{V} \subseteq \mathcal{V}' \subseteq \mathcal{U}$ and every K ,*

$$\Theta_I^* \subseteq \Theta_O(\mathcal{V}') \subseteq \Theta_O(\mathcal{V}) \quad \text{and} \quad \Theta_I^* \subseteq \Theta_O^K(\mathcal{V}).$$

If, in addition, $\Pi_K \mathcal{V} \subseteq \mathcal{V}$, then $\Theta_O(\mathcal{V}) \subseteq \Theta_O^K(\mathcal{V})$.

Proof. By Theorem 1, $\theta \in \Theta_I^*$ if and only if the inequality holds for every $u \in \mathcal{U}$. Imposing it only on a subset can only enlarge the set, which gives the first chain. The second claim is a special case, since $\Pi_K \mathcal{V} \subseteq \mathcal{U}$. If $\Pi_K \mathcal{V} \subseteq \mathcal{V}$, monotonicity gives the last inclusion. $\square$

Without $\Pi_K \mathcal{V} \subseteq \mathcal{V}$, the sets $\Theta_O(\mathcal{V})$ and $\Theta_O^K(\mathcal{V})$ need not be nested. Truncating a direction changes the inequality it tests, so a parameter value with $\mathbf{E}[\tilde{m}(W^O, \theta, u)] \geq 0$ may violate the inequality at $\Pi_K u$. Both sets are nonetheless outer sets of Θ_I^* . Examples of sets closed under truncation are $\mathcal{U}$, B_K , and the coordinate directions $\{\pm e_j\}$ used in Appendix E; for the latter, $\Pi_K e_j \in \{e_j, 0\}$, and the zero direction imposes no restriction.

Finite direction sets approximate the sharp criterion uniformly. Under the decay condition of Theorem 3, $(\mathcal{U}, \rho_a)$ is totally bounded. For $\xi \in (0, 1]$, the proof of Lemma 13 constructs a finite set $\mathcal{U}_\xi \subseteq \mathcal{U}$ with $|\mathcal{U}_\xi| \leq \lambda(\xi)$ and $\min_{v \in \mathcal{U}_\xi} \rho_a(u, v) \leq \xi/2$ for every $u \in \mathcal{U}$. Combined with the Lipschitz bound $|\tilde{m}(w^o, \theta, u) - \tilde{m}(w^o, \theta, v)| \leq M_0 \rho_a(u, v)$ of Appendix D, this gives, uniformly in θ and P_{W^O} ,

$$\inf_{u \in \mathcal{U}} \mathbf{E}[\tilde{m}(W^O, \theta, u)] \geq \min_{v \in \mathcal{U}_\xi} \mathbf{E}[\tilde{m}(W^O, \theta, v)] - \frac{M_0 \xi}{2}.$$

Hence $\Theta_I^* \subseteq \Theta_O(\mathcal{U}_\xi) \subseteq \{\theta : \inf_{u \in \mathcal{U}} \mathbf{E}[\tilde{m}(W^O, \theta, u)] \geq -M_0 \xi/2\}$. The relevant metric is the weighted semimetric ρ_a , under which $\mathcal{U}$ is totally bounded; under the sup norm it is not.

For inference with a direction set $\mathcal{V} \subseteq \mathcal{U}$, let $CS_n^{AS}(1 - \alpha; \mathcal{V})$ be defined as $CS_n^{AS}(1 - \alpha)$ in Section 3.2, with every supremum over $\mathcal{U}$, in the test statistic and in the bootstrap statistics, replaced by a supremum over $\mathcal{V}$. For finite $\mathcal{V}$ these suprema are maxima. Let $\Theta_O(\mathcal{V})(P)$

denote $\Theta_O(\mathcal{V})$ under the observed-data law P , and let $\mathcal{F}_{\mathcal{V}} \equiv \{(\theta, P) \in \mathcal{F}^+ : \theta \in \Theta_O(\mathcal{V})(P)\}$. By Corollary 1, $\mathcal{F} \subseteq \mathcal{F}_{\mathcal{V}}$.

Corollary 2. *Let the conditions of Theorem 3 hold.*

(a) *Coverage holds uniformly over direction sets and null pairs:*

$$\liminf_{n \rightarrow \infty} \inf_{\mathcal{V} \subseteq \mathcal{U}} \inf_{(\theta, P) \in \mathcal{F}_{\mathcal{V}}} P^n \{\theta \in CS_n^{AS}(1 - \alpha; \mathcal{V})\} \geq 1 - \alpha.$$

In particular, since $\mathcal{F} \subseteq \mathcal{F}_{\mathcal{V}}$, the bound holds uniformly over $(\theta, P) \in \mathcal{F}$ and $\mathcal{V} \subseteq \mathcal{U}$.

(b) *For fixed $\mathcal{V} \subseteq \mathcal{U}$, $P \in \mathcal{P}$ and $\theta \notin \Theta_O(\mathcal{V})(P)$, $P^n \{\theta \notin CS_n^{AS}(1 - \alpha; \mathcal{V})\} \rightarrow 1$.*

(c) *If $K(n) \rightarrow \infty$, then $CS_n^{AS}(1 - \alpha; B_{K(n)})$ has coverage at least $1 - \alpha$ uniformly over $\mathcal{F}$, by (a). It also excludes each fixed $\theta \notin \Theta_I^*(P)$ with probability approaching one. No rate restriction on $K(n)$ is required.*

For a single, fixed, finite $\mathcal{V}$ the decay condition is not needed, since the constant manageability function $\lambda \equiv |\mathcal{V}|$ suffices. The decay condition makes the bound in (a) uniform over $\mathcal{V}$, which the data-adaptive procedure below requires. In (c), the confidence level does not improve as $K(n)$ grows; it remains at least $1 - \alpha$. What tightens is the collection of tested inequalities, so that every fixed $\theta \notin \Theta_I^*$ is eventually excluded. The proof is in Appendix D.

For the empirical implementation, we use Corollary 2(a) to select informative directions in a data-adaptive way while preserving validity. The procedure is as follows.

1. Randomly split the sample into two subsamples A and B of approximately equal size.
2. Using A , run the numerical direction search over the parameter grid. Collect directions that yield strong violations for candidate parameter values, giving a finite direction bank $\hat{\mathcal{V}}_A$. Augment it by a small set of deterministic benchmark directions.
3. Compute the confidence set on B with direction set $\hat{\mathcal{V}}_A$ at level $1 - \alpha/2$.

4. Reverse the roles of A and B , and intersect the two confidence sets.

Because $\widehat{\mathcal{V}}_A$ is a function of A alone, it is independent of B . Writing n_B for the size of B , conditioning on A gives, for every $(\theta, P) \in \mathcal{F}$,

$$\begin{aligned} P\{\theta \in CS_{n_B}^{AS}(1 - \alpha/2; \widehat{\mathcal{V}}_A)\} &= \mathbf{E}\left[P\{\theta \in CS_{n_B}^{AS}(1 - \alpha/2; \widehat{\mathcal{V}}_A) \mid A\}\right] \\ &\geq \inf_{\mathcal{V} \subseteq \mathcal{U}} P^{n_B}\{\theta \in CS_{n_B}^{AS}(1 - \alpha/2; \mathcal{V})\}. \end{aligned}$$

By Corollary 2(a), the limit inferior of the right-hand side is at least $1 - \alpha/2$, uniformly over $\mathcal{F}$. This is where uniformity over $\mathcal{V}$ matters: a guarantee for each fixed direction set separately would not cover a bank that changes with the sample. The direction bank may also depend on the candidate θ , since coverage is assessed one value of θ at a time. By the Bonferroni inequality, the intersection of the two confidence sets has coverage at least $1 - \alpha$. This procedure avoids treating directions selected from the same observations as fixed, while retaining most of the computational advantage of concentrating the bootstrap on a small set of informative directions.

5 Monte Carlo Simulation

We study the finite-sample performance of the proposed confidence sets in a linear regression model,

$$Y_i = \theta_0 + \theta_1 X_i + \varepsilon_i, \quad X_i \sim U[0, 1], \quad (\theta_0, \theta_1) = (0, 1),$$

where ε_i is a centered Gaussian draw clipped to a compact support. The outcome is always observed exactly, while a subset of the covariates is observed only through intervals. For an interval-observed observation we set

$$X_{i,l} = \max\{0, X_i - ws_i\}, \quad X_{i,u} = X_i,$$

where $s_i \sim U[0.5, 1.5]$. Hence, the true covariate is located at the upper endpoint of the reported interval, while the interval length varies randomly across observations. The two simulation exercises below use this same underlying regression model and interval construction. We first study coverage and confidence-set geometry in a genuinely set-identified design, and then consider a point-identified design that allows us to evaluate coverage at a known boundary point and power against nearby alternatives.

We first consider two mechanisms generating interval observation throughout the support of X_i . In the first design, interval observation is independent of (X_i, Y_i) , with probability $p = 10\%$ and $w = 0.2$. Thus, the design corresponds to a relatively mild censoring problem, for which conventional methods based on complete observations are expected to perform well. In the second, more demanding design, interval observation depends on the realized outcome. Specifically, letting $p = 30\%$, observations with $Y_i \leq 0.5$ are interval-observed with probability $1.6p$, whereas those with $Y_i > 0.5$ are interval-observed with probability $0.4p$. Since the distribution of $Y_i - 0.5$ is symmetric around zero, this preserves an unconditional censoring probability of p . We additionally increase the interval width to $w = 0.5$. Thus, the second exercise deliberately combines outcome-dependent interval observation with substantially greater ambiguity about the latent covariate.

For each design we conduct 500 Monte Carlo replications for $n \in \{500, 1000, 5000\}$ and construct 90 percent confidence sets using 499 bootstrap draws in each replication. We implement our split-sample procedure as described in the preceding section and compare to an outer set construction which restricts attention to the coordinatewise extreme directions, i.e. an extension of the Aucejo, Bugni, and Hotz (2017) outer sets to the interval case. As conventional benchmarks, we also report joint 90 percent OLS confidence ellipses obtained either by replacing each interval-valued covariate by its midpoint or by discarding all interval-observed observations.

We evaluate the procedures primarily by empirical coverage of the true parameter and by the size and shape of the resulting confidence regions. To summarize their geometry across

Monte Carlo replications, we report Minkowski averages of the set-valued confidence regions, alongside the average OLS confidence ellipses.

Figure 1 shows the average geometry of the 90 percent confidence sets for each method, with the random censoring design in the first row and the outcome-dependent and more severe censoring design in the second row. The blue-shaded region is the confidence set obtained from our method, whereas the orange confidence set corresponds to a refinement of the procedure of Aucejo, Bugni, and Hotz (2017). The blue dotted and black dashed lines correspond to the implied ordinary least squares ellipses resulting from fitting a model with imputed interval midpoints or ignoring the interval-observed observations, respectively.

Focusing on the first row, all sets shrink as the sample size increases. The censoring problem in this case is relatively mild compared with the sampling uncertainty, so even our missing- and interval-data robust confidence set contracts relatively quickly. However, because the procedure does not impose a parametric selection model for the interval observations, its confidence region is necessarily wider than procedures that rely on additional assumptions. Moreover, the ambiguity induced by interval observation implies that increasing the sample size removes sampling uncertainty but does not remove the underlying identification problem. Comparing the Aucejo, Bugni, and Hotz (2017) outer confidence set with our method, the additional joint restrictions implied by sharpness produce tighter sets. In particular, for $n = 500$, the outer procedure fails to reject the null hypothesis of a zero slope, whereas the confidence set corresponding to the sharp characterization excludes it. Although the average sizes of the confidence sets from midpoint imputation and complete-case least squares appear reasonable in this relatively benign design, the empirical coverage results in Table 1 reveal that midpoint imputation undercovers as the sample size increases. The imputation procedure is silent about the identification problem and therefore becomes increasingly confident around the parameter induced by the particular imputation rule.

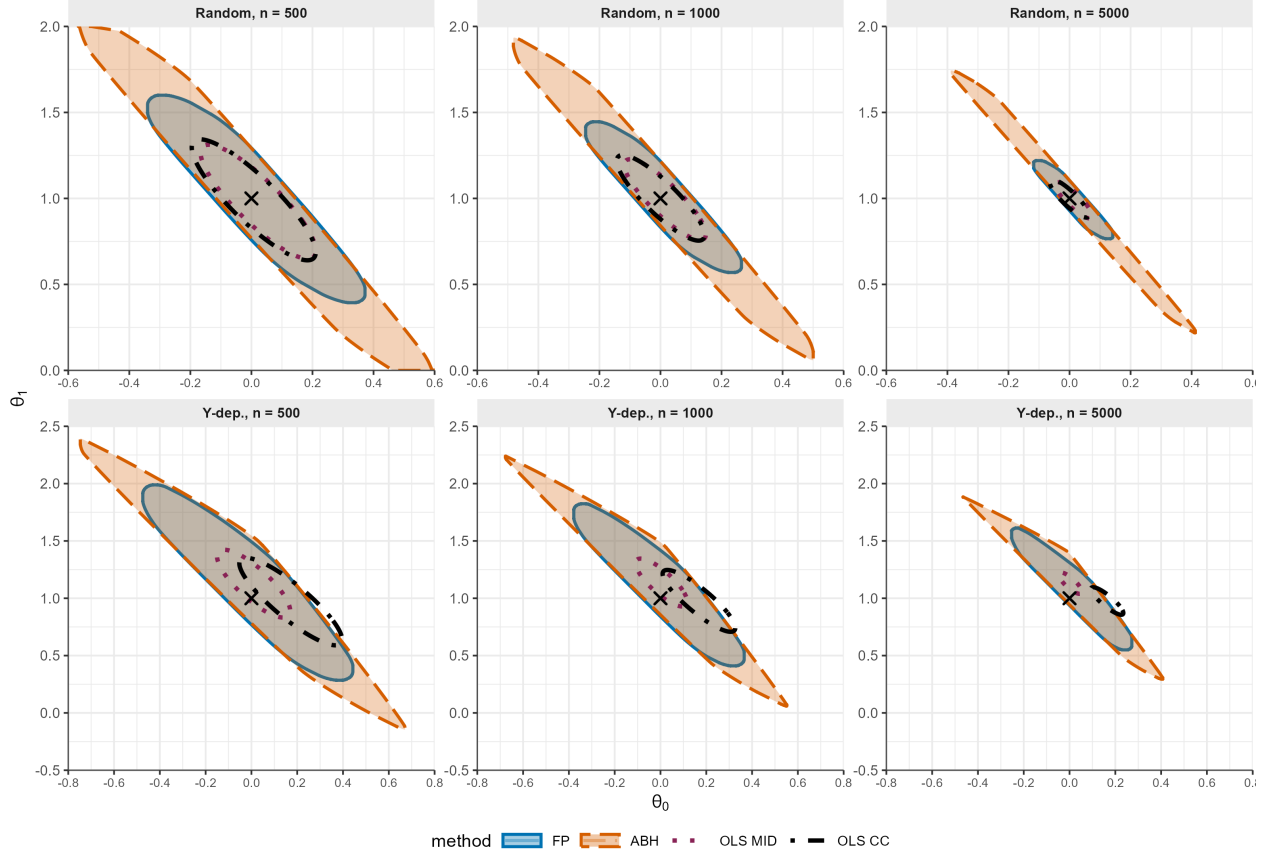


Figure 1: Minkowski-Averaged Confidence Sets across Simulations

The second row illustrates the more demanding case in which interval observation is outcome dependent and substantially more frequent ($p = 30\%$). The resulting confidence sets are considerably larger. In contrast to the random-censoring design, the two ordinary least squares approaches are no longer visually centered around the true value. Instead, both become increasingly concentrated away from the true parameter as the sample size increases. Complete-case OLS is distorted because selection into the complete-case sample depends on the outcome, while midpoint imputation imposes a particular latent completion that is not justified by the model. The proposed robust confidence set, by contrast, continues to account for the ambiguity generated by the interval observations. As a result, it shrinks considerably less than in the first row: in this design the identification uncertainty generated by the large interval-observed share remains quantitatively important even when sampling uncertainty has become small. The empirical coverage results in Table 1 confirm this contrast. Cov-

erage of the conventional procedures deteriorates sharply, whereas the proposed procedure continues to conservatively attain the nominal 90 percent level.

Table 1: Empirical Coverage by DGP and Method

DGP	FP	ABH	OLS MID	OLS CC
Random, n = 500	0.958	0.952	0.894	0.890
Random, n = 1000	0.960	0.966	0.896	0.902
Random, n = 5000	0.974	0.964	0.838	0.890
Y-dep., n = 500	0.984	0.976	0.548	0.152
Y-dep., n = 1000	0.980	0.954	0.294	0.004
Y-dep., n = 5000	0.970	0.940	0.000	0.000

The preceding simulations are meant to illustrate the failure of common approaches and the robustness of our method. However, as we do not know the shape of the identified set for this example, whether the true value is on the boundary of the set or not is not known. If it is an interior point of the set, the simulation evidence above is weaker, as it is easier to cover the true value in this cases. Simultaneously, we want to understand power properties against fixed local alternatives. We therefore consider a closely related design in which the regression parameter is point identified despite interval observation of some covariates.

We retain the same regression model and interval construction but restrict censoring to the lower half of the support of X_i . In particular, censoring can occur only when $X_i \leq 1/2$, while X_i is always observed exactly for $X_i > 1/2$. For a censored observation we use $w = 0.1$. We consider two censoring intensities. In the baseline design,

$$\Pr(C_i = 1 \mid X_i \leq 1/2) = 0.10,$$

whereas in the more-censoring design this probability is 0.50. Censoring is zero for $X_i > 1/2$, so the corresponding unconditional censoring rates are 0.05 and 0.25.

As above, we consider both random and outcome-dependent censoring. For the outcome-dependent counterpart, conditional on $X_i \leq 1/2$, we set

$$\begin{aligned}\Pr(C_i = 1 \mid Y_i \leq 0.25) &= 1.6q, \\ \Pr(C_i = 1 \mid Y_i > 0.25) &= 0.4q,\end{aligned}$$

where $q \in \{0.10, 0.50\}$. Since $X_i - 0.25$ is symmetric around zero conditional on $X_i \leq 1/2$, and the regression disturbance is symmetric,

$$\Pr(Y_i \leq 0.25 \mid X_i \leq 1/2) = \frac{1}{2}.$$

Consequently, the average censoring probability in the lower half remains q , so that the random and outcome-dependent designs have the same censoring intensity and differ only in the selection mechanism.

Since X_i is observed exactly on the nondegenerate interval $(1/2, 1]$, the conditional (global) regression restriction implies point identification as the conditional mean function can be learned unambiguously on $(1/2, 1]$. Thus, the sharp identified set collapses to the singleton $\Theta_I^* = \{(0, 1)\}$. The true parameter is therefore a known boundary point of the identified set. This allows us to evaluate finite-sample coverage at the boundary and power against nearby fixed alternatives without introducing numerical error from estimating the population boundary itself.

Figure 2 reports empirical inclusion frequencies over 200 Monte Carlo replications along the slope path $(0, \theta_1)$, with $\theta_1 \in [0.8, 1]$, for $n \in \{1,000, 5,000, 20,000\}$. We compare the proposed sharp procedure with complete-case OLS. The columns compare the baseline and more-censoring designs, while the rows distinguish random from outcome-dependent censoring.

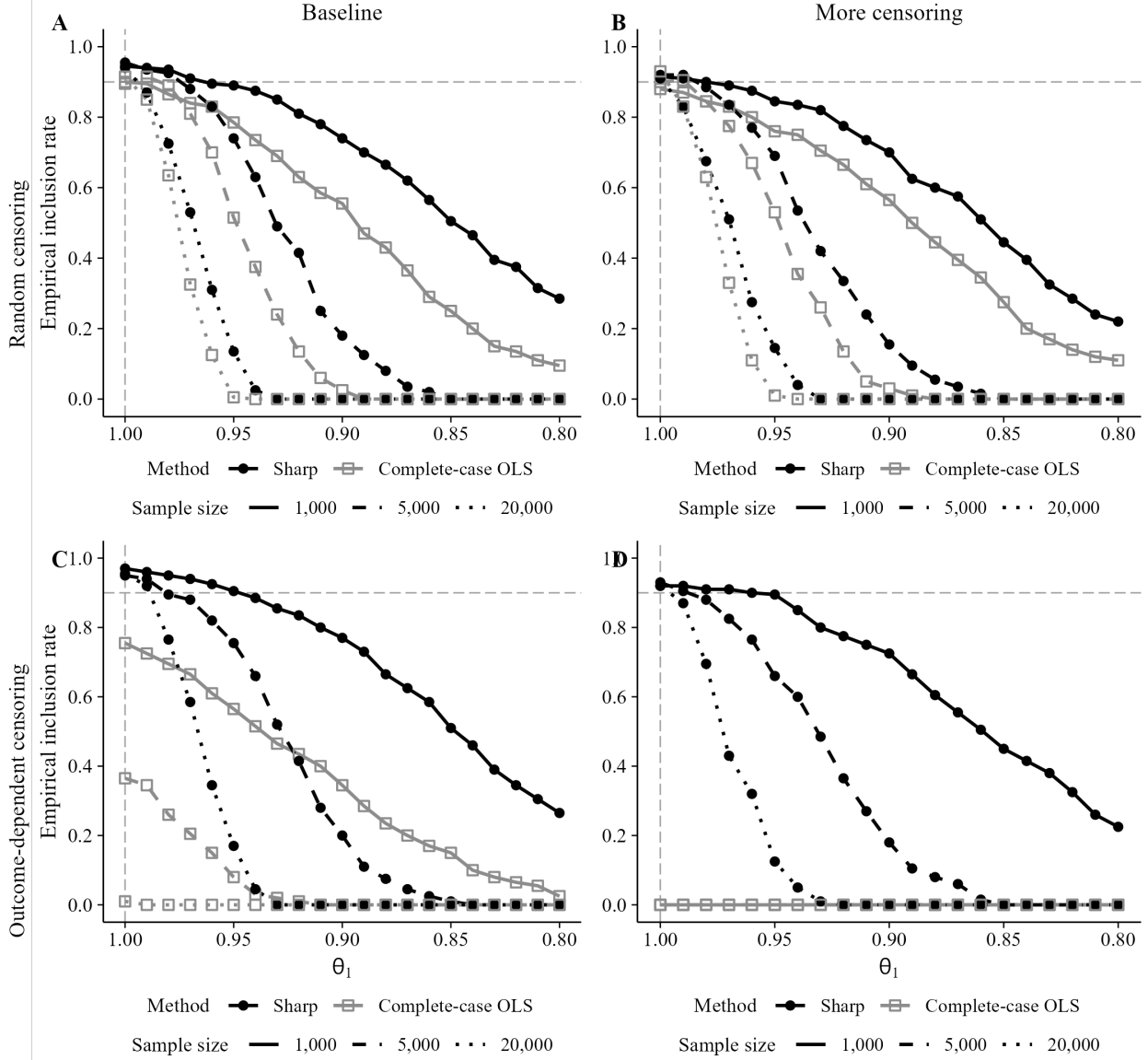


Figure 2: Empirical Inclusion Frequencies along the Slope Path

Figure 2 illustrates the trade-off between robustness and power. Under random censoring, complete-case OLS is valid and generally more powerful than the proposed procedure, since in this case discarding the interval-observed observations does not induce selection bias. The lower power of the sharp procedure in this benign design is therefore the price paid for remaining agnostic about the censoring mechanism. Importantly, however, the proposed procedure continues to deliver approximately nominal coverage at the true parameter under both random and outcome-dependent censoring, while its power against fixed alter-

natives increases substantially with the sample size. Increasing the censoring probability reduces power, as expected from the associated loss of information. At the same time, when censoring is outcome dependent, the same increase in censoring severity makes complete-case OLS increasingly unreliable: its inclusion frequency at the true parameter deteriorates sharply, whereas the robust procedure retains coverage and achieves substantial power once the sample is sufficiently large. Thus, the simulations highlight the central trade-off: robustness entails a finite-sample power cost when complete-case analysis happens to be valid, but protects inference when the censoring mechanism is non-random.

6 Application

As an empirical illustration, we revisit Castillo, Jordan, and Petrie (2018), who elicit children’s time preferences using a multiple price list (20 binary choices between \$49 in one month and a larger amount in seven months) and study how the resulting discount-rate measure predicts on-time high school graduation. In a multiple price list, respondents face a sequence of binary intertemporal choices in which one option is held fixed while the alternative varies monotonically across rows (here, the later payment increases). Under standard monotonicity, choices should exhibit a single cutoff: the respondent selects the sooner option for lower later amounts and then switches once to the later option when the offered premium becomes sufficiently large. This structure maps each switching point into a bracket for the respondent’s discount rate, because all later amounts below the switch are rejected while all later amounts above are accepted.

The setting is attractive for our purposes for three reasons. First, multiple price lists are a workhorse design in experimental and applied micro work, so the example speaks to a broad class of studies. The design intrinsically delivers interval-valued covariates: even with a single switch, the discount rate is only pinned down to a fixed-width bracket determined by two adjacent rows. Second, one common problem for multiple price list elicitation is that

some respondents answer inconsistently, i.e. they switch multiple times between accepting and rejecting the later payment. For instance, such a respondent might state to accept \$49 in a month over \$60 in seven months but prefers \$50 in seven months over \$49 in one month, which is inconsistent with standard preference relations. These respondents are reported to differ systematically in observable characteristics (see e.g. the discussion and potential solutions targeting the study design in Yu, Zhang, and Zuo (2021)). Treating such responses as random error or collapsing them to the interval between the first and last switch generates heterogeneous interval lengths that can be correlated with unobservables, raising concerns for identification strategies that rely on mean independence in the presence of interval measurement (C. Manski and Tamer 2002). Third, the main outcome suffers from substantial selective missingness: roughly 30% of participants have unknown graduation status because they leave the district or otherwise attrit, motivating inverse-probability weighting in the original analysis and providing a natural testbed for methods designed to remain valid under missing outcomes alongside interval measurement.

In the main specification of the original study the authors use an algorithm which finds for all inconsistent decision makers an associated single switch pattern which minimizes the cost to a consistent decision. As the algorithm used produces potential multiplicity, Castillo, Jordan, and Petrie (2018) use either the average of these artificial single switching patterns, the first switch, the last switch and a cost minimizing third alternative. Thus, for all observations a single switch pattern is either imputed or observed and the midpoint of the associated discount rate interval is used as a regressor.

Our estimator allows us to derive confidence sets for the parameters of interest robust to the interval and missing data problems. First, taking the single switch algorithm as given, we can rerun the probit model only accounting for the ambiguity within one interval of width 0.0735 which is inherent to the survey design (*Interval*). Second, we present results additionally accounting for the fact that for inconsistent individual responses, the discount rate is known only to lie between the first and last switches (*Interval + Inconsistent*). An

argument could be made that in these cases the discount rate should be treated as missing and should not be bounded by the observed switching pattern. Our results below already show that the original conclusions of the paper do not survive treating the responses as somewhat informative, treating them as missing would only increase the ambiguity. Finally, the last specification includes the individuals for which high school graduation is not observed but for which discount rates have been elicited (*Interval + Missing*). We use a probit model regressing the outcome on an intercept and the potentially interval-measured discount rate. Table 2 reports the marginal effect at the same reference value used in Castillo, Jordan, and Petrie (2018), i.e. the mean of the imputed midpoints. The point estimate of approximately -0.09 from the univariate probit model is essentially identical to the estimate obtained with the full set of controls in Castillo, Jordan, and Petrie (2018). We report the projection of the original 90% confidence set alongside three confidence sets from our method that progressively account for the different forms of data incompleteness. Accounting only for the interval width of 0.0735 inherent in the MPL design has a relatively modest effect: the confidence set widens from $[-0.161, -0.011]$ to $[-0.196, -0.008]$. Thus, even after allowing the discount rate to take any value within its elicited bracket, zero remains excluded and the negative association with on-time graduation remains statistically significant. In this specification, sampling uncertainty therefore dominates the ambiguity generated by interval measurement. The conclusion changes once we additionally account for inconsistent response patterns. Allowing the discount rate of these respondents to lie anywhere between their first and last switching points widens the confidence set to $[-0.387, 0.159]$, so that a zero marginal effect can no longer be rejected. Finally, including the approximately 30% of students whose graduation outcome is unobserved further enlarges the confidence set, as shown in the final row of Table 2.

Sharpness is important for the substantive interpretation of these results. To illustrate this point, we also implement the Aucejo, Bugni, and Hotz (2017) style outer set approach used in the simulation section. When accounting only for the interval measurement mechanically

induced by the MPL design, their outer-set construction yields a 90% confidence set of $[-0.366, 0.176]$, which already contains zero. Taken at face value, this would suggest that allowing for the elicited intervals alone is sufficient to eliminate the statistical significance of the estimated marginal effect. Our sharp confidence set, by contrast, is $[-0.196, -0.008]$ and continues to exclude zero. Hence, the loss of significance under the outer-set procedure does not reflect ambiguity that is actually implied by the interval-observed data and the maintained model, but rather the additional conservativeness introduced by a non-sharp characterization.

Data treatment	Conventional probit	Our method	ABH-style outer set
Midpoint imputation	$-0.086 [-0.161, -0.011]$	—	—
MPL intervals	—	$[-0.196, -0.008]$	$[-0.366, 0.176]$
MPL intervals + inconsistent choices	—	$[-0.387, 0.159]$	$[-0.523, 0.398]$
MPL intervals + missing outcomes	—	$[-0.489, 0.451]$	$[-0.990, 0.661]$

Table 2: Marginal Effects at a Common Reference Discount Rate

The application illustrates the distinction between sampling uncertainty and uncertainty generated by incomplete data. Our method does not mechanically render empirical conclusions uninformative: when only the interval measurement inherent in the MPL design is taken into account, the original qualitative conclusion survives. Instead, the loss of statistical significance arises primarily once the ambiguity associated with inconsistent responses is incorporated, and becomes still greater when missing graduation outcomes are admitted. By contrast, the example has shown that outer sets might display considerably more conservative results.

More generally, the approach replaces a collection of alternative imputations, sample restrictions, and data-completion rules with a single confidence set that considers all data completions consistent with the maintained model and the observed information. In this sense, the confidence set provides a compact and transparent robustness summary: it shows

directly which empirical conclusions are supported by the model and the incomplete data themselves, and which depend on additional assumptions about how the missing or interval-observed values are resolved.

7 Conclusion

Conditional moment equality models are central to empirical economics, but the variables on which they condition are often not fully observed. This paper studies this problem without replacing the latent variables by imputations, conditioning on the observed intervals, or imposing missing-at-random or mean-independence assumptions. We characterize the sharp identified set: the set of all parameter values consistent with the maintained conditional moment model and the incomplete data alone. The resulting object provides a worst-case benchmark against which stronger identifying assumptions and imputation-based procedures can be assessed.

The main identification result rewrites the conditional moment restriction as a countable collection of unconditional moment restrictions and embeds these restrictions in an infinite-dimensional random set. A support-function characterization then converts the existence of an admissible completion of the latent data into a collection of moment inequalities. This extends the random-set approach to a setting in which incomplete continuous covariates generate infinitely many restrictions. We also show how the characterization can be operationalized through finite-dimensional sieve approximations, yielding a Hausdorff-consistent estimator and uniformly valid confidence regions. For monotone-index models, including linear, probit, and logit regressions, the finite approximations are computationally tractable. The empirical application illustrates how the proposed confidence sets can distinguish uncertainty due to sampling variation from uncertainty generated by incomplete data. In the multiple price list application, accounting only for the interval measurement inherent in the survey design leaves the original qualitative conclusion intact: the marginal effect of the

discount rate on on-time graduation remains significantly negative. The conclusion changes, however, once the ambiguity associated with inconsistent switching behavior is incorporated, and allowing additionally for missing graduation outcomes widens the confidence set further. More broadly, the exercise shows that empirical conclusions based on a particular imputation or data-cleaning rule may depend importantly on how incomplete observations are resolved. Our approach provides a unified way to assess this sensitivity by reporting the range of conclusions compatible with the maintained regression model and the information contained in the observed data.

References

- Andrews, D. and X. Shi (2013). “Inference based on Conditional Moment Inequalities”. In: *Econometrica* 81.2, pp. 609–666.
- (2017). “Inference based on many conditional moment inequalities”. In: *Journal of Econometrics* 196, pp. 275–287.
- Andrews, D. and G. Soares (2010). “Inference for Parameters defined by Moment Inequalities using Generalized Moment Selection”. In: *Econometrica* 78.1, pp. 119–157.
- Aradillas Fernández, A. et al. (2025). “Epsilon-Minimax Solutions of Statistical Decision Problems”. In: *Working paper*. URL: <https://arxiv.org/abs/2509.08107v1>.
- Asher, S., P. Novosad, and C. Rafkin (2018). “Partial identification of expectations with interval data”. In: *Permanent Working Paper*.
- Aucejo, E., F. Bugni, and J. Hotz (2017). “Identification and Inference on Regressions with Missing Covariate Data”. In: *Econometric Theory* 33, pp. 196–241.
- Aumann, R. (1965). “Integrals of Set-Valued Functions”. In: *Journal of Mathematical Analysis and Applications* 12.1, pp. 1–12.
- Beer, Gerald (1993). *Topologies on Closed and Closed Convex Sets*. Vol. 268. Mathematics and Its Applications. Dordrecht: Kluwer Academic Publishers.

- Beresteanu, A., I. Molchanov, and F. Molinari (2011). “Sharp Identification Regions in Models with convex Moment Predictions”. In: *Econometrica* 79.6.
- Beresteanu, A. and F. Molinari (2008). “Asymptotic Properties for a Class of Partially Identified Models”. In: *Econometrica* 76.4, pp. 763–814.
- Bogachev, Vladimir I. (2007). *Measure Theory, Volumes I and II*. Berlin: Springer-Verlag. DOI: 10.1007/978-3-540-34514-5.
- Bretschneider, T. and Y. Kaddoura (2026). “Factor-Based Imputation of Missing Values Using Cross-Sectional Averages”. In.
- Canay, I. and A. Shaikh (2017). “Practical and theoretical advances in inference for partially identified models”. In: *Advances in Economics and Econometrics 2*, pp. 271–306.
- Castillo, M., J. Jordan, and R. Petrie (2018). “Discount Rates of Children and High School Graduation”. In: *The Economic Journal* 129, pp. 1153–1181.
- Chernozhukov, V., D. Chetverikov, and K. Kato (2017). “Central Limit Theorems and Bootstrap in High Dimensions”. In: *The Annals of Probability* 45, pp. 2309–2352.
- (2019). “Inference on Causal and Structural Parameters using Many Moment Inequalities”. In: *Review of Economic Studies* 86, pp. 1867–1900.
- Chernozhukov, V., H. Hong, and E. Tamer (2007). “Parameter set inference in a class of econometric models”. In: *Econometrica* 75.5, pp. 1243–1284.
- Duan, J. and M. Pelger (2025). “Imputation-powered inference for missing covariates”. In: *NBER Working Paper*.
- Dudley, Richard M. (1967). “The sizes of compact subsets of Hilbert space and continuity of Gaussian processes”. In: *Journal of Functional Analysis* 1.3, pp. 290–330.
- Fitzgerald, J., P. Gottschalk, and R. Moffitt (1998). “An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics”. In: *Journal of Human Resources* 33.2, pp. 251–299. DOI: 10.2307/146433.
- Franguridi, G. and H. Moon (2026). “Generalized Method of Moments with Partially Missing Data”. In: *arXiv preprint arXiv:2511.21988*.

- Freyberger, J. et al. (2025). “Missing Data in Asset Pricing Panels”. In: *Review of Financial Studies* 38.2.
- Heckman, J. (1979). “Sample Selection Bias as a Specification Error”. In: *Econometrica* 47.1, pp. 153–161. DOI: 10.2307/1912352.
- Heitjan, D. and D. Rubin (1991). “Ignorability and Coarse Data”. In: *The Annals of Statistics* 19.4, pp. 2244–2253. DOI: 10.1214/aos/1176348396.
- Horowitz, J. and C. Manski (1998). “Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations”. In: *Journal of Econometrics* 84.1, pp. 37–58.
- (2000). “Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data”. In: *Journal of the American Statistical Association*, pp. 77/84.
- Horowitz, J. L. and C. F. Manski (2006). “Identification and Estimation of Statistical Functionals Using Incomplete Data”. In: *Journal of Econometrics* 132.2, pp. 445–459. DOI: 10.1016/j.jeconom.2005.02.007.
- Hsiao, C. (1983). “Regression Analysis with a Categorized Explanatory Variable”. In: *Studies in Econometrics, Time Series and Multivariate Statistics*, pp. 93–129.
- Kechris, Alexander S. (1995). *Classical Descriptive Set Theory*. Vol. 156. Graduate Texts in Mathematics. New York: Springer-Verlag. DOI: 10.1007/978-1-4612-4190-4.
- Kline, B. and E. Tamer (2023). “Recent Developments in Partial Identification”. In: *Annual Review of Economics* 15.
- Lee, David S. (2009). “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects”. In: *Review of Economic Studies* 76.3, pp. 1071–1102.
- Manski, C. (2003). *Partial Identification of Probability Distributions*. New York: Springer.
- Manski, C. and E. Tamer (2002). “Inference on Regressions with Interval Data on a Regressor or Outcome”. In: *Econometrica* 70, pp. 519–546.
- Manski, C. F. (2005). “Partial identification with missing data: Concepts and findings”. In: *International Journal of Approximate Reasoning* 39.2-3, pp. 151–165.

- Manski, C. F. (2008). “Partial Identification in Econometrics”. In: *The New Palgrave Dictionary of Economics*. Ed. by S. N. Durlauf and L. E. Blume. Palgrave Macmillan.
- Molchanov, I (2017). *Theory of Random Sets*.
- Molchanov, I. and F. Molinari (2018). *Random Sets in Econometrics*.
- Molinari, F. (2020). “Microeconometrics with partial identification”. In: *Handbook of Econometrics* 7A.
- Newey, W. (2009). “Two-Step Series Estimation of Sample Selection Models”. In: *The Econometrics Journal* 12.S1, S217–S229. DOI: 10.1111/j.1368-423X.2008.00263.x.
- Pacini, David (2018). “The two-sample linear regression model with interval-censored covariates”. In: *Journal of Applied Econometrics* 33, pp. 1–16.
- Pisier, G. (1989). *The Volume of Convex Bodies and Banach Space Geometry*. Vol. 94. Cambridge Tracts in Mathematics. Cambridge University Press.
- Pollard, David (1990). *Empirical Processes: Theory and Applications*. Vol. 2. NSF-CBMS Regional Conference Series in Probability and Statistics. Hayward, CA: Institute of Mathematical Statistics.
- Rigobon, R. and T. Stoker (2009). “Bias from censored regressors”. In: *Journal of Business and Economic Statistics* 27, pp. 340–353.
- Robins, J., A. Rotnitzky, and L. Zhao (1995). “Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data”. In: *Journal of the American Statistical Association* 90.429, pp. 106–121. DOI: 10.1080/01621459.1995.10476493.
- Rubin, D. (1976). “Inference and Missing Data”. In: *Biometrika* 63.3, pp. 581–592. DOI: 10.1093/biomet/63.3.581.
- Rubin, Donald B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons. DOI: 10.1002/9780470316696.
- Schennach, S. (2014). “Entropic Latent Variable Integration via Simulation”. In: *Econometrica* 82.1, pp. 345–385.

- Wooldridge, J. (2007). “Inverse Probability Weighted Estimation for General Missing Data Problems”. In: *Journal of Econometrics* 141.2, pp. 1281–1301. DOI: 10.1016/j.jeconom.2007.02.002.
- Yu, C., J. Zhang, and S. Zuo (2021). “Multiple Switching and Data Quality in the Multiple Price List”. In: *The Review of Economics and Statistics* 103.1, pp. 136–150.

A Proofs of Section 2

A.1 Proof of Lemma 1

For compact notation, let $\mathbf{C}(\theta, \eta)$ denote the statement that

$$\mathbf{E}_\eta[m(Y, X, \theta) \mid X] = 0 \quad \eta_X\text{-a.s.},$$

and let $\mathbf{U}(\theta, \eta)$ denote the statement that

$$\int m_j(y, x, \theta) g_k(x) \eta(dy, dx, dw^o) = 0 \quad \forall j, k.$$

Proof. Fix $(\theta, \eta) \in \Theta \times \Gamma$.

Suppose $\mathbf{C}(\theta, \eta)$ holds and set

$$\mu_j(x) \equiv \mathbf{E}_\eta[m_j(Y, X, \theta) \mid X = x].$$

Iterated expectations give

$$\int m_j(y, x, \theta) g_k(x) d\eta = \int \mu_j(x) g_k(x) \eta_X(dx) = 0$$

for every (j, k) . Thus $\mathbf{U}(\theta, \eta)$ holds.

Suppose $\mathbf{U}(\theta, \eta)$ holds and define $\mu_j(X) = \mathbf{E}_\eta[m_j(Y, X, \theta) \mid X]$. By linearity, the equality holds for every $v \in \text{span}(\mathcal{G})$. Fix $f \in C(\mathcal{X})$. Assumption 2(i) supplies $v_r \in \text{span}(\mathcal{G})$ with $\|f - v_r\|_\infty \rightarrow 0$. For all sufficiently large r ,

$$|\mu_j v_r| \leq (\|f\|_\infty + 1) |\mu_j|.$$

Dominated convergence therefore gives

$$\int \mu_j(x) f(x) \eta_X(dx) = \lim_{r \rightarrow \infty} \int \mu_j(x) v_r(x) \eta_X(dx) = 0 \quad \forall f \in C(\mathcal{X}).$$

Let ν_j be the finite signed measure with density μ_j relative to η_X . Since $\mathcal{X}$ is compact, uniqueness in the Riesz representation theorem implies from the preceding display that $\nu_j = 0$. Thus $\mu_j = 0$ η_X -almost surely for every j , and $\mathbf{C}(\theta, \eta)$ holds.

We have proved the fixed-pair equivalence

$$\mathbf{C}(\theta, \eta) \iff \mathbf{U}(\theta, \eta). \quad (\text{FP})$$

Both identified sets quantify over the same $\eta \in \Gamma$, with the appropriate integrability included in $\mathbf{C}$ and $\mathbf{U}$. Therefore $\Theta_I = \Theta_I^*$. $\square$

A.2 Class of Weighted Cubes

The following proposition introduces the instrument class used and establishes that the class satisfies Assumption 2. The choice of weighted cubes, allows the exact linear program epigraph formulation described in the main text. For this reason it is the selected class in our implementations, however, we note that many other classes would also fulfill the conditions needed for our theoretical results.

Proposition 1. *For the cube construction, let $\mathcal{X} = [0, 1]^{d_x}$, following Andrews and Shi (2013). Let $r_0 \geq 1$ be an integer. For each $r \geq r_0$ let $N_r \equiv \sum_{s=r_0}^r (2s)^{d_x}$, with $N_{r_0-1} = 0$, and for $a = (a_1, \dots, a_{d_x}) \in \{1, \dots, 2r\}^{d_x}$, define*

$$I_{a_j, r} \equiv \begin{cases} [0, (2r)^{-1}], & a_j = 1, \\ ((a_j - 1)/(2r), a_j/(2r)], & a_j = 2, \dots, 2r, \end{cases} \quad C_{a, r} \equiv \prod_{j=1}^{d_x} I_{a_j, r},$$

$$g_{a, r}(x) \equiv \mathbf{1}\{x \in C_{a, r}\}, \quad \omega_{a, r} \equiv w(r) = (c_0 + N_r^\beta)^{-1}.$$

If $c_0 > 0$, $\beta > 1$, and the pairs (a, r) are enumerated by increasing resolution, the resulting instruments and weights satisfy Assumption 2.

Proof. At every resolution r , the cubes $\{C_{a,r} : a \in \{1, \dots, 2r\}^{d_x}\}$ form a partition of $[0, 1]^{d_x}$ under the stated boundary convention. Exactly one indicator is active at each x , so

$$\sup_{x \in [0, 1]^{d_x}} \sum_{r=r_0}^{\infty} \sum_a \omega_{a,r} |g_{a,r}(x)| = \sum_{r=r_0}^{\infty} (c_0 + N_r^\beta)^{-1} < \infty.$$

The series converges because $N_r \geq r - r_0 + 1$ and $\beta > 1$. Let $R(K)$ be the largest resolution completed among the first K instruments, with $R(K) = r_0 - 1$ if none is complete. Enumeration by increasing resolution implies $R(K) \rightarrow \infty$. At any x , at most one of the remaining indicators is active at each unresolved resolution; hence the tail is bounded uniformly by

$$\sum_{r > R(K)} (c_0 + N_r^\beta)^{-1} \rightarrow 0.$$

This proves Assumption 2(ii).

For Part (i), fix $f \in C([0, 1]^{d_x})$ and $\varepsilon > 0$. Uniform continuity supplies a $\delta > 0$ such that $|f(x) - f(x')| < \varepsilon$ whenever $\|x - x'\|_2 < \delta$. Choose r with $\sqrt{d_x}/(2r) < \delta$; this quantity bounds the diameter of every $C_{a,r}$. Select $x_{a,r} \in C_{a,r}$ and define

$$v_r(x) \equiv \sum_a f(x_{a,r}) g_{a,r}(x).$$

The sum is finite, so $v_r \in \text{span}(\mathcal{G})$, and the diameter bound gives $\|f - v_r\|_\infty < \varepsilon$. $\square$

The positive weights do not alter this finite-span approximation because the weights enter the stacked moments separately from the instruments. Their role is to place the infinite stack in ℓ_1 and control its tail.

For inference, the same weights also satisfy the decay condition in Theorem 3 when $\beta > 3/2$. To see this, let $(a(k), r(k))$ be the cell and resolution of instrument k in this

enumeration, so that $(g_k, \omega_k) = (g_{a(k), r(k)}, \omega_{a(k), r(k)})$ and $N_{r(k)-1} < k \leq N_{r(k)}$. Recall the scalar effective weight $a_k \equiv \omega_k \sup_{x \in \mathcal{X}} |g_k(x)|$ from Theorem 3. Since each cube indicator has supremum one,

$$a_k = \omega_k = (c_0 + N_{r(k)}^\beta)^{-1} \leq N_{r(k)}^{-\beta} \leq k^{-\beta}.$$

For truncation through resolution R , let $K = N_R$ and let $\Pi_K v$ retain the first K instruments in each primitive-moment block and zero-pad the rest. Uniformly over w^o , θ , and $v \in B$,

$$\begin{aligned} |\phi(w^o, \theta, v) - \phi(w^o, \theta, \Pi_K v)| &\leq d_m C_1 \sup_{x \in \mathcal{X}} \sum_{k > K} \omega_k |g_k(x)| \\ &= d_m C_1 \sum_{r > R} (c_0 + N_r^\beta)^{-1}. \end{aligned}$$

Exactly one cube indicator is active at each resolution. Since $N_r \asymp r^{d_x+1}$, the last expression is $O(R^{1-\beta(d_x+1)})$.

A.3 Exact feasibility and closure-based identification

This subsection compares the exact set Θ_I with a distance-based formulation and shows that they coincide under the paper's assumptions. The point is not that I_θ must be closed; it need not be. Rather, any sequence of admissible moment vectors converging to zero yields an admissible law that attains zero.

In the notation of this paper, the distance-based formulation associated with Schennach (2014, Equation (3)) is

$$\Theta_I^{\text{Sch}} \equiv \left\{ \theta \in \Theta : \inf_{\eta \in \Gamma} \left\| \int M_\theta(y, x) \eta(dy, dx, dw^o) \right\|_1 = 0 \right\}. \quad (9)$$

By (4),

$$\inf_{\eta \in \Gamma} \left\| \int M_\theta d\eta \right\|_1 = \text{dist}_{\ell_1}(0, I_\theta).$$

Distance to a set equals distance to its closure, so

$$\Theta_I^{\text{Sch}} = \{\theta \in \Theta : 0 \in \text{cl}_{\ell_1} I_\theta\}. \quad (10)$$

Because Γ is nonempty,

$$\Theta_I \subseteq \Theta_I^{\text{Sch}},$$

so that

$$\Theta_I^{\text{Sch}} \setminus \Theta_I = \{\theta \in \Theta : 0 \in \text{cl}_{\ell_1} I_\theta \text{ and } 0 \notin I_\theta\}.$$

Proposition 2 (Equivalence in the present model). *Under Assumptions 1–3,*

$$\Theta_I = \Theta_I^{\text{Sch}}.$$

By (10), $\Theta_I^{\text{Sch}} = \{\theta \in \Theta : 0 \in \text{cl}_{\ell_1} I_\theta\}$, and Appendix Lemma 6 gives $0 \in I_\theta$ if and only if $0 \in \text{cl}_{\ell_1} I_\theta$. The two sets therefore coincide. Lemma 6 is proved in Appendix B.

For comparison, the following general observation gives a familiar sufficient condition under which exact and distance-based definitions agree.

Proposition 3 (A general compactness condition). *Let Γ' be a nonempty compact topological*

space and suppose that $b_\theta : \Gamma' \rightarrow \ell_1$, or $\mathbb{R}^L$ for finite L , is well defined for every $\theta \in \Theta$. Define

$$\Theta^{\text{exact}} \equiv \{\theta : b_\theta(\gamma) = \mathbf{0} \text{ for some } \gamma \in \Gamma'\},$$

$$\Theta^{\text{dist}} \equiv \{\theta : \inf_{\gamma \in \Gamma'} \|b_\theta(\gamma)\|_1 = 0\}.$$

If $\gamma \mapsto \|b_\theta(\gamma)\|_1$ is lower semicontinuous on Γ' for every θ , then $\Theta^{\text{exact}} = \Theta^{\text{dist}}$.

Proof. The inclusion $\Theta^{\text{exact}} \subseteq \Theta^{\text{dist}}$ is immediate. Fix $\theta \in \Theta^{\text{dist}}$ and for each $n \geq 1$ define

$$L_n(\theta) \equiv \{\gamma \in \Gamma' : \|b_\theta(\gamma)\|_1 \leq n^{-1}\}.$$

Because $\inf_{\gamma \in \Gamma'} \|b_\theta(\gamma)\|_1 = 0$, every $L_n(\theta)$ is nonempty. Lower semicontinuity makes each $L_n(\theta)$ closed in Γ' (Beer 1993), and $L_{n+1}(\theta) \subseteq L_n(\theta)$. Compactness and the finite-intersection property therefore give $\bigcap_{n \geq 1} L_n(\theta) \neq \emptyset$. For any γ^* in this intersection, $0 \leq \|b_\theta(\gamma^*)\|_1 \leq n^{-1}$ for every n , so $b_\theta(\gamma^*) = \mathbf{0}$ and $\theta \in \Theta^{\text{exact}}$. $\square$

In the present model $b_\theta(\eta) = \int M_\theta d\eta$. If every instrument g_k is continuous, then M_θ is a uniform ℓ_1 -limit of continuous truncations and is therefore continuous, so $\eta \mapsto \|\int M_\theta d\eta\|_1$ is continuous and satisfies the lower-semicontinuity condition of Proposition 3. This is a sufficient condition, not the route used for Theorem 1. Continuity of the instruments would substantially simplify some parts of the following identification proof, see the remarks in the respective section. For the general theory developed in this paper we do not rely on continuity.

The multiresolution cube geometry follows Andrews and Shi (2013); the ℓ_1 coordinate weights are imposed here for Assumption 2(ii) and Theorem 3. The offset c_0 is a positive normalization or tuning constant, not an identifying restriction. Setting $c_0 = 100$ and $\beta = 2$ retains the tuning constants in Andrews and Shi (2013); here they enter $w(r) = (c_0 + N_r^\beta)^{-1}$ to satisfy the enumerated decay condition. No restriction is imposed on the probability that X lies on a cube boundary.

B Proof of Theorem 1

B.1 Roadmap and probability-space convention

The proof has five steps. Lemma 2 establishes that the moment image is a measurable integrably bounded random compact set. Lemma 3 identifies admissible laws with laws of feasible latent selections. Lemma 4 records compactness and convexity of the admissible-law set. Lemma 5 connects the Aumann expectation to the attainable mean-moment set I_θ . Lemma 6 shows that taking the closure does not alter membership at zero. The main proof then applies Molchanov (2017) convexity and support-function results, similar to Beresteanu, I. Molchanov, and Molinari (2011).

The probability space carrying the true latent variables $(\Omega, \mathcal{A}, F_0)$ is only a representation of their joint law. An element $\eta \in \Gamma$ should not be understood as a perturbation of the true law $\eta_0 = \mathcal{L}_{F_0}(W, W^o)$ on that particular space. Rather, it represents an alternative complete data-generating process having the observed marginal P_{W^o} and assigning probability one to latent values compatible with the observed set. Because the maintained restrictions and the definition of Γ depend only on joint laws, we may, without changing the model or its identified set, work throughout this appendix on the completed enrichment of the original probability space obtained by adjoining an independent $V \sim \text{Unif}[0, 1]$. All original random variables are identified with their lifts to the enriched space, which preserves their joint laws and therefore also preserves the conditional moment restriction in (1). Selections and Aumann expectations below are understood on this enriched space, while the original notation is retained for simplicity. The uniform variable allows every conditional probability kernel represented by an element of Γ to be realized jointly with W^o , as established in Lemma 3.⁸ The enriched probability space is nonatomic, so the Aumann expectations used below are convex by Molchanov (2017, Theorem 2.1.30). Thus their population convexity does not rely on nonatomicity of the original probability space or on the i.i.d. sampling assumption; the

⁸This enrichment follows the construction in the proof of Molchanov (2017, Theorem 1.4.8, pp. 82–83), see also Molchanov (2017, Definition 1.4.7 and Example 1.4.9, pp. 82–83).

latter remains in force for the sampling and asymptotic results.

Recall that π_K is the truncation map and λ_K is the corresponding zero-padding embedding defined in Section 3.

B.2 Auxiliary lemmas

Lemma 2. *The map $M_\theta : \mathcal{W} \rightarrow \ell_1$ is bounded and Borel measurable. Moreover,*

$$Q_\theta(W^O) = \text{cl}_{\ell_1} M_\theta(D(W^O))$$

is an integrably bounded, almost surely nonempty random compact set in ℓ_1 . For every finite K , $Q_{\theta,K} \equiv \pi_K Q_\theta$ is a random compact set in $\mathbb{R}^{d_m K}$.

Proof. Index the coordinates by (j, k) and write

$$M_\theta(y, x)_{j,k} = m_j(y, x, \theta) g_k(x).$$

Assumptions 2 and 3 give

$$\begin{aligned} \|M_\theta(y, x)\|_1 &\leq \sum_{j=1}^{d_m} \sum_{k \geq 1} |m_j(y, x, \theta)| |g_k(x)| \\ &\leq d_m C_1 \sup_{x' \in \mathcal{X}} \sum_{k \geq 1} |g_k(x')| < \infty, \end{aligned} \tag{11}$$

uniformly in (y, x) . Thus M_θ is bounded and ℓ_1 -valued.

Let $M_{\theta,K}^0 = \lambda_K \pi_K M_\theta$. The finite-coordinate map $\pi_K M_\theta : \mathcal{W} \rightarrow \mathbb{R}^{d_m K}$ is Borel because its coordinates are products of continuous and Borel functions; composing it with the continuous embedding λ_K shows that $M_{\theta,K}^0$ is Borel. Moreover,

$$\sup_{(y,x) \in \mathcal{W}} \|M_\theta(y, x) - M_{\theta,K}^0(y, x)\|_1 \leq d_m C_1 \sup_{x \in \mathcal{X}} \sum_{k > K} |g_k(x)| \longrightarrow 0. \tag{12}$$

A pointwise limit of Borel maps into a metric space is Borel, so M_θ is Borel.

The same tail bound makes the range totally bounded. Given $\varepsilon > 0$, choose K so that the bound in (12) is less than $\varepsilon/2$. The bounded set $M_{\theta,K}^0(\mathcal{W})$ lies in a finite-dimensional subspace and has a finite $\varepsilon/2$ -net. That net is an ε -net for $M_\theta(\mathcal{W})$. Since ℓ_1 is complete, $\text{cl}_{\ell_1} M_\theta(\mathcal{W})$ is compact. Each $Q_\theta(w^o)$ is a closed subset of this common compact set and is nonempty because $D(w^o)$ is almost surely nonempty under Assumption 1. Thus its values are almost surely compact and nonempty.

It remains to verify random-set measurability on Ω . Let $O \subseteq \ell_1$ be open. Then $B = M_\theta^{-1}(O)$ is Borel, and

$$\begin{aligned} H_B &\equiv \{w^o : D(w^o) \cap B \neq \emptyset\} \\ &= \text{proj}_{\mathcal{W}^o} \{(w, w^o) \in \text{Gr}(D) : w \in B\}. \end{aligned}$$

The set inside the projection is Borel in a product of Polish spaces, so H_B is analytic and therefore $P_{\mathcal{W}^o}$ -universally measurable. Closure does not change whether a set meets an open set, and hence

$$\{\omega : Q_\theta(W^O(\omega)) \cap O \neq \emptyset\} = \{\omega : W^O(\omega) \in H_B\}.$$

The right-hand side belongs to $\mathcal{A}$ because the probability space is complete. This proves Effros measurability of $\omega \mapsto Q_\theta(W^O(\omega))$. The envelope (11) gives integrable boundedness. Finally, the continuous image of a compact random set under π_K is a compact random set, which establishes the finite-dimensional claim. $\square$

Lemma 3. *Let $\mathcal{S} \equiv \mathcal{Y} \times \mathcal{X}$ and let $\mathcal{W}_o$ denote the state space of W_o . Suppose $\mathcal{S} \times \mathcal{W}_o$ is Polish and $D : \mathcal{W}_o \rightrightarrows \mathcal{S}$ is nonempty compact-valued with closed graph. Let Γ be as defined*

in Section 2. Equivalently, for $\text{Gr}(D) \equiv \{(s, w) \in \mathcal{S} \times \mathcal{W}_o : s \in D(w)\}$, we have

$$\Gamma = \{\eta \in \Delta(\mathcal{S} \times \mathcal{W}_o) : \eta_{W_o} = P_{W_o}, \eta(\text{Gr}(D)) = 1\}.$$

Under the enriched probability-space convention described above, any measurable selection $S' = (Y', X')$ of $D(W_o)$ induces an element $\eta \equiv \mathcal{L}(S', W_o) \in \Gamma$. Conversely, for every $\eta \in \Gamma$, there exists a measurable selection $S^\eta = (Y^\eta, X^\eta)$ of $D(W_o)$ such that $\mathcal{L}(S^\eta, W_o) = \eta$. In other words, $\Gamma = \{\mathcal{L}(S, W_o) : S(\omega) \in D(W_o(\omega)) \text{ a.s.}\}$.

Proof. First let $S' = (Y', X')$ be a measurable selection of $D(W_o)$ and define $\eta \equiv \mathcal{L}(S', W_o)$. Then $\eta_{W_o} = P_{W_o}$ by construction. Moreover, $F_0(S' \in D(W_o)) = 1$, so $\eta(\text{Gr}(D)) = 1$. Thus $\eta \in \Gamma$.

Conversely, fix $\eta \in \Gamma$. Define

$$\mathbb{D}(\omega) := D(W_o(\omega)) \times \{W_o(\omega)\} \subseteq \mathcal{S} \times \mathcal{W}_o.$$

By Assumption 1, $\mathbb{D}$ is a nonempty random compact set. We verify Artstein's inequalities. Let $F \subseteq \mathcal{S} \times \mathcal{W}_o$ be closed. Since $\eta(\text{Gr}(D)) = 1$, $\eta(F) = \eta(F \cap \text{Gr}(D))$. If $(s, w_o) \in F \cap \text{Gr}(D)$, then $s \in D(w_o)$ and $(s, w_o) \in F$. Hence $F \cap (D(w_o) \times \{w_o\}) \neq \emptyset$. Therefore,

$$\mathbb{I}\{(s, w_o) \in F \cap \text{Gr}(D)\} \leq \mathbb{I}\{F \cap (D(w_o) \times \{w_o\}) \neq \emptyset\}.$$

Integrating with respect to η gives

$$\eta(F) \leq \int \mathbb{I}\{F \cap (D(w) \times \{w\}) \neq \emptyset\} \eta_{W_o}(dw).$$

Because $\eta_{W_o} = P_{W_o}$,

$$\eta(F) \leq F_0(F \cap (D(W_o) \times \{W_o\}) \neq \emptyset) = F_0(\mathbb{D} \cap F \neq \emptyset).$$

Thus η satisfies Artstein's inequalities for $\mathbb{D}$. By Artstein's selectionability theorem (Molchanov 2017, Theorem 1.4.8, pp. 82–83), η is selectionable with respect to the law of $\mathbb{D}$. More specifically, the construction in the proof of that theorem yields a measurable map f_η such that, using the independent $V \sim \text{Unif}[0, 1]$ supplied by the enriched probability space, the random element $Z^\eta \equiv f_\eta(V, \mathbb{D})$ satisfies $Z^\eta \in \mathbb{D}$ almost surely and $\mathcal{L}(Z^\eta) = \eta$. Since $\mathbb{D} = D(W_o) \times \{W_o\}$, the last component of Z^η equals W_o almost surely. Writing $Z^\eta = (Y^\eta, X^\eta, W_o)$ therefore gives $(Y^\eta, X^\eta) \in D(W_o)$ F_0 -a.s. and $\mathcal{L}(Y^\eta, X^\eta, W_o) = \eta$. $\square$

This law-selection bridge is the formal random-set foundation for the sharpness statement in the main text. It plays a role analogous to that of Assumption 2.4 in Beresteanu, I. Molchanov, and Molinari (2011), but here the existence of the relevant selection is derived from the definition of Γ and the enriched probability-space convention rather than imposed as an additional model assumption.

Lemma 4. *The set Γ is nonempty, convex, compact, and metrizable under weak convergence.*

Proof. Nonemptiness follows from $\eta_0 = \mathcal{L}_{F_0}(W, W^O) \in \Gamma$. The ambient space $\mathcal{W} \times \mathcal{W}^O$ is compact metric, so its space of probability measures is compact and metrizable under weak convergence (Bogachev 2007, Section 8.6). It remains to show that Γ is closed.

Let $\eta_n \in \Gamma$ and $\eta_n \Rightarrow \eta$. Continuity of the coordinate projection gives $(\eta_n)_{W^O} \Rightarrow \eta_{W^O}$, so $\eta_{W^O} = P_{W^O}$. Since $\text{Gr}(D)$ is closed, the Portmanteau theorem gives

$$1 = \limsup_{n \rightarrow \infty} \eta_n(\text{Gr}(D)) \leq \eta(\text{Gr}(D)) \leq 1.$$

Thus $\eta \in \Gamma$, proving closedness and hence compactness. Finally, the fixed-marginal and probability-one support restrictions defining Γ are affine. Therefore every mixture of two elements of Γ also belongs to Γ , so Γ is convex. $\square$

The preceding lemmata establish that Q_θ is well defined and connect the admissible completion laws in Γ to feasible latent selections. It remains to relate the attainable mean-

moment set I_θ generated by Γ to the Aumann expectation of $Q_\theta(W^O)$. Two closure operations must be handled carefully. First, because M_θ need only be Borel measurable, $M_\theta(D(w^o))$ need not be closed, so $Q_\theta(w^o)$ is defined as its closure. Second, the Aumann expectation is the closure of the set of selection means. Lemma 5 shows that the combined effect of these operations is $\mathbb{E}[Q_\theta(W^O)] = \text{cl}_{\ell_1} I_\theta$, while Lemma 6 shows that this closure does not affect identification because it does not alter membership at zero.

Let

$$\mathcal{A}_\theta \equiv \{\mathbf{E}[q] : q \text{ is an integrable selection of } Q_\theta(W^O)\}$$

denote the unclosed set of selection means. Recall that I_θ was defined in (4) and that $\mathbb{E}[Q_\theta(W^O)] = \text{cl}_{\ell_1} \mathcal{A}_\theta$.

Lemma 5. *For every $\theta \in \Theta$, $I_\theta \subseteq \mathcal{A}_\theta \subseteq \text{cl}_{\ell_1} I_\theta$, and consequently $\mathbb{E}[Q_\theta(W^O)] = \text{cl}_{\ell_1} \mathcal{A}_\theta = \text{cl}_{\ell_1} I_\theta$.*

Proof. For the first inclusion, let $z \in I_\theta$. Then there exists $\eta \in \Gamma$ such that

$$z = \int M_\theta(w) \eta(dw, dw^o).$$

By Lemma 3, there exists a measurable $S^\eta \in D(W^O)$ whose joint law with W^O is η . Hence $q^\eta \equiv M_\theta(S^\eta)$ is an integrable selection of $Q_\theta(W^O)$ and

$$\mathbf{E}[q^\eta] = \int M_\theta(w) \eta(dw, dw^o) = z.$$

Thus $I_\theta \subseteq \mathcal{A}_\theta$.

For $\mathcal{A}_\theta \subseteq \text{cl}_{\ell_1} I_\theta$, let q be an arbitrary integrable selection of $Q_\theta(W^O)$. By definition of Q_θ ,

$$q(\omega) \in \text{cl}_{\ell_1} M_\theta(D(W^O(\omega))) \quad F_0\text{-a.s.}$$

The closure is relevant because M_θ is only Borel. Hence $q(\omega)$ need not admit an exact preimage in $D(W^O(\omega))$, although it admits admissible preimages whose moment vectors are arbitrarily close. Fix $r \geq 1$ and define the relation

$$\mathcal{R}_r \equiv \{((w^o, z), w) \in (\mathcal{W}^O \times \ell_1) \times \mathcal{W} : w \in D(w^o), \|M_\theta(w) - z\|_1 < r^{-1}\}.$$

For $e = (w^o, z)$, let $\mathcal{R}_r^\circ(e) \equiv w : (e, w) \in \mathcal{R}_r$. Thus $\mathcal{R}_r^\circ(w^o, z)$ is the set of admissible latent values $w \in D(w^o)$ whose moment vectors approximate the target z within r^{-1} .

The relation $\mathcal{R}_r$ is Borel, because $(w^o, w) : w \in D(w^o) = \text{Gr}(D)$ is closed, while $(w, z) \rightarrow \|M_\theta(w) - z\|_1$ is Borel because M_θ is Borel by Lemma 2 and the norm is continuous.

Next define

$$\mathbf{E} \equiv \mathcal{W}^O \times \ell_1, \quad C_r \equiv \text{proj}_{\mathbf{E}} \mathcal{R}_r, \quad Z_q(\omega) \equiv (W^O(\omega), q(\omega)), \quad \mu \equiv \mathcal{L}_{F_0}(Z_q).$$

Since q is an integrable ℓ_1 -valued selection, it is measurable, and hence Z_q is an $\mathbf{E}$ -valued measurable random element. Moreover, $\mathbf{E}$ and $\mathcal{W}$ are standard Borel spaces.

Since $\mathcal{R}_r$ is Borel in $\mathbf{E} \times \mathcal{W}$, its projection C_r is analytic. Hence C_r is universally measurable and therefore belongs to the μ -completion $\overline{\mathcal{B}(\mathbf{E})}^\mu$.

By $q(\omega) \in \text{cl}_{\ell_1} M_\theta(D(W^O(\omega)))$ F_0 -a.s., for every r there exists an admissible $w \in D(W^O(\omega))$ such that $\|M_\theta(w) - q(\omega)\|_1 < r^{-1}$. Hence $\mathcal{R}_r^\circ(W^O(\omega), q(\omega)) \neq \emptyset$ F_0 -a.s. Equivalently, $Z_q(\omega) \in C_r$ F_0 -a.s. Thus the random pair $Z_q = (W^O, q)$ lies almost surely in the part of the relation on which an approximate inverse exists, and therefore $\mu(C_r) = 1$.

By the Jankov-von Neumann uniformization theorem (Kechris 1995, Theorem 18.1), there exists a universally measurable map $\tilde{\sigma}_r : C_r \rightarrow \mathcal{W}$ such that

$$\tilde{\sigma}_r(e) \in \mathcal{R}_r^\circ(e) \quad \text{for every } e \in C_r.$$

Fix $w_0 \in \mathcal{W}$ and extend $\tilde{\sigma}_r$ to all of $\mathbf{E}$ by setting $\tilde{\sigma}_r(e) = w_0$ for $e \notin C_r$. Since C_r and

the inverse images of Borel sets under the selector are universally measurable, this extension remains universally measurable. In particular, it is measurable with respect to the μ -completion of $\mathcal{B}(\mathbf{E})$.

Define

$$S_{q,r}(\omega) \equiv \tilde{\sigma}_r(Z_q(\omega)) = \tilde{\sigma}_r(W^O(\omega), q(\omega)).$$

Then $S_{q,r}$ is $\mathcal{A}$ -measurable. Indeed, for every $B \in \mathcal{B}(\mathcal{W})$, $\tilde{\sigma}_r^{-1}(B)$ belongs to the μ -completion of $\mathcal{B}(\mathbf{E})$. Hence it differs from a Borel subset of $\mathbf{E}$ only by a μ -null set. Pulling back through Z_q therefore gives a set that differs from an $\mathcal{A}$ -measurable set only by an F_0 -null set. Completeness of $(\Omega, \mathcal{A}, F_0)$ yields measurability.

Since $Z_q \in C_r$ almost surely, the defining property of the selector gives

$$S_{q,r} \in D(W^O), \quad \|M_\theta(S_{q,r}) - q\|_1 < r^{-1} \quad F_0\text{-a.s.}$$

Thus $S_{q,r}$ is an admissible latent completion whose moment vector approximates the selected vector q pointwise.

Let $\eta_r \equiv \mathcal{L}_{F_0}(S_{q,r}, W^O)$. By Lemma 3, $\eta_r \in \Gamma$. Moreover, the uniform ℓ_1 bound on M_θ implies that $M_\theta(S_{q,r})$ is Bochner integrable. Therefore

$$z_r \equiv \int M_\theta(w), \eta_r(dw, dw^o) = \mathbf{E}[M_\theta(S_{q,r})] \in I_\theta.$$

Using above and the Bochner triangle inequality,

$$|z_r - \mathbf{E}[q]|_1 = |\mathbf{E}[M_\theta(S_{q,r}) - q]|_1 \leq \mathbf{E}[\|M_\theta(S_{q,r}) - q\|_1] < r^{-1}.$$

Hence $z_r \in I_\theta$ and $z_r \rightarrow \mathbf{E}[q]$ in ℓ_1 . Therefore $\mathbf{E}[q] \in \text{cl}_{\ell_1} I_\theta$. Since q was an arbitrary integrable selection of $Q_\theta(W^O)$, $\mathcal{A}_\theta \subseteq \text{cl}_{\ell_1} I_\theta$. Together with the first inclusion, this proves

the first part of the Lemma.

Finally, taking norm closures yields $\text{cl}_{\ell_1} \mathcal{A}_\theta = \text{cl}_{\ell_1} I_\theta$. By definition of the Aumann expectation, $\mathbb{E}[Q_\theta(W^O)] = \text{cl}_{\ell_1} \mathcal{A}_\theta$, which establishes the second Lemma claim. $\square$

Lemma 6. *Under Assumptions 1–3,*

$$0 \in \text{cl}_{\ell_1} I_\theta \quad \Longleftrightarrow \quad 0 \in I_\theta. \quad (13)$$

Proof. Only the forward implication requires proof. Suppose $0 \in \text{cl}_{\ell_1} I_\theta$. There are $\eta_n \in \Gamma$ such that

$$z_n \equiv \int M_\theta(w) \eta_n(dw, dw^o), \quad \|z_n\|_1 \rightarrow 0.$$

By Lemma 4, pass to a subsequence with $\eta_n \Rightarrow \eta \in \Gamma$.

Fix $j \in \{1, \dots, d_m\}$, $f \in C(\mathcal{X})$, and $\varepsilon > 0$. Assumption 2(i) gives

$$v_\varepsilon(x) = \sum_{k \in K_\varepsilon} a_k g_k(x), \quad K_\varepsilon \subset \mathbb{N} \text{ finite}, \quad \|f - v_\varepsilon\|_\infty < \varepsilon.$$

Let $B_\varepsilon = \sum_{k \in K_\varepsilon} |a_k| < \infty$. Because the (j, k) coordinate of z_n is $\int m_j g_k d\eta_n$,

$$\begin{aligned} \left| \int m_j(y, x, \theta) v_\varepsilon(x) d\eta_n \right| &= \left| \sum_{k \in K_\varepsilon} a_k \int m_j(y, x, \theta) g_k(x) d\eta_n \right| \\ &\leq B_\varepsilon \|z_n\|_1 \longrightarrow 0. \end{aligned} \quad (14)$$

Assumption 3 then gives

$$\left| \int m_j(y, x, \theta) f(x) d\eta_n \right| \leq B_\varepsilon \|z_n\|_1 + C_1 \varepsilon. \quad (15)$$

The function $(y, x, w^o) \mapsto m_j(y, x, \theta) f(x)$ is bounded and continuous. Weak convergence

therefore permits passage to the limit in its integral. From (15),

$$\left| \int m_j(y, x, \theta) f(x) d\eta \right| \leq C_1 \varepsilon.$$

Letting $\varepsilon \downarrow 0$ gives

$$\int m_j(y, x, \theta) f(x) d\eta = 0 \quad \forall f \in C(\mathcal{X}), \forall j. \quad (16)$$

Let $\mu_j(X) = \mathbf{E}_\eta[m_j(Y, X, \theta) \mid X]$. By iterated expectations, the finite signed measure with density μ_j relative to η_X integrates every $f \in C(\mathcal{X})$ to zero. The Riesz representation theorem therefore implies $\mu_j = 0$ η_X -almost surely for every j . The fixed-pair equivalence (FP), together with the coordinate identity for Bochner integrals, gives

$$\int M_\theta(w) \eta(dw, dw^o) = 0.$$

Because $\eta \in \Gamma$, this vector belongs to I_θ . Thus $0 \in I_\theta$.

Notice that the proof passes to the weak limit only through continuous test functions. It never requires the limiting law to assign zero probability to the discontinuity boundaries of the cube instruments. $\square$

Remark 2. A different and simpler proof strategy could be employed under a continuous moment class effectively making Lemma 5 and Lemma 6 superfluous. Suppose $M_\theta : \mathcal{W} \rightarrow \ell_1$ is continuous. Then $Q_\theta(W^O) = M_\theta(D(W^O))$ F_0 -a.s. and $\mathcal{A}_\theta = I_\theta = \mathbb{E}[Q_\theta(W^O)]$. Moreover, I_θ is compact in ℓ_1 .

Proof. Because $D(W^O)$ is compact and M_θ is continuous, $M_\theta(D(W^O))$ is compact and already closed. Thus the closure defining Q_θ is redundant.

Let q be an integrable selection of $Q_\theta(W^O)$ and define

$$H_q(\omega) \equiv \{w \in D(W^O(\omega)) : M_\theta(w) = q(\omega)\}.$$

The correspondence is nonempty and random closed by the random-inverse theorem (Molchanov 2017, Theorem 1.3.27). The Fundamental Selection Theorem (Molchanov 2017, Theorem 1.4.1) supplies a measurable $S_q \in H_q$. Lemma 3 implies $\eta_q = \mathcal{L}_{F_0}(S_q, W^O) \in \Gamma$, and hence

$$\mathbf{E}[q] = \int M_\theta(w) \eta_q(dw, dw^o) \in I_\theta.$$

Therefore $\mathcal{A}_\theta \subseteq I_\theta$; the first inclusion in Lemma 5 gives the reverse inclusion.

To show closedness, define

$$T_\theta : \Gamma \rightarrow \ell_1, \quad T_\theta(\eta) = \int M_\theta(w) \eta(dw, dw^o).$$

If $\eta_n \Rightarrow \eta$, the first $d_m K$ coordinates converge for every fixed K because they are bounded and continuous. The tail bound (12) holds uniformly over Γ . Finite-coordinate convergence followed by the uniform tail bound gives $\|T_\theta(\eta_n) - T_\theta(\eta)\|_1 \rightarrow 0$. Thus T_θ is continuous. Lemma 4 makes Γ compact, so $I_\theta = T_\theta(\Gamma)$ is compact and closed. Since $\mathbf{E}[Q_\theta(W^O)] = \text{cl}_{\ell_1} \mathcal{A}_\theta$, all claimed equalities follow. $\square$

B.3 Main Proof of Theorem 1

Proof. Lemma 2 establishes the asserted random-set properties. The Fundamental Selection Theorem (Molchanov 2017, Theorem 1.4.1) supplies a measurable selection of $Q_\theta(W^O)$; the common envelope in (11) makes that selection integrable. Hence $\mathbf{E}[Q_\theta(W^O)]$ is nonempty.

By the coordinate identity for Bochner integrals and the fixed-pair equivalence (FP),

$$\theta \in \Theta_I^* \iff \theta \in \Theta_I \iff 0 \in I_\theta.$$

Lemmas 5 and 6 give

$$0 \in \mathbb{E}[Q_\theta(W^O)] \iff 0 \in \text{cl}_{\ell_1} I_\theta \iff 0 \in I_\theta.$$

This proves (5) and the first line of (6).

The set $\mathbb{E}[Q_\theta(W^O)]$ is closed by Molchanov (2017, Definition 2.1.21). It is convex because the extended probability space is nonatomic and Molchanov (2017, Theorem 2.1.30) applies to the integrably bounded random closed set $Q_\theta(W^O)$ in the separable Banach space ℓ_1 . Equivalently, after Lemma 5, convexity also follows from convexity of Γ and affinity of $\eta \mapsto \int M_\theta d\eta$.

By Molchanov (2017, Theorem 2.1.35),

$$h(\mathbb{E}[Q_\theta(W^O)], u) = \mathbb{E}_{P_{W^O}}[h(Q_\theta(W^O), u)] \quad \forall u \in \ell_\infty.$$

For a nonempty closed convex subset C of ℓ_1 , the Hahn–Banach separation theorem gives

$$0 \in C \iff h(C, u) \geq 0 \quad \forall u \in (\ell_1)^* = \ell_\infty.$$

Positive homogeneity of $u \mapsto h(C, u)$ permits restriction to B_1^∞ . Combining the last two displays proves the final equality in (6). $\square$

C Proof of Theorem 2

This section first proves a triangular law of large numbers for random set. We then proceed by stating and proving several auxiliary Lemmata which are needed to verify the high

level conditions in Chernozhukov, Hong, and Tamer (2007) for our case and thus establish consistency of the proposed estimator.

C.1 Triangular Law of Large Numbers for Random Sets

Before stating the consistency result for the proposed estimator we will introduce an auxiliary Lemma establishing a law of large numbers for random sets growing at rate p_n which might be of independent interest and is a crucial part of the proof.

Lemma 7. *Let p_n denote the ambient dimension of the n -th row of a triangular array, with $p_n \rightarrow \infty$. For each n , let $X_{n,1}, \dots, X_{n,n}$ be i.i.d. nonempty compact random subsets of $\mathbb{R}^{p_n}$, and let X_n denote a generic copy from the n -th row and suppose they are uniformly bounded by a deterministic constant b . Then*

$$\rho_H \left(\text{co} \left[\frac{1}{n} \bigoplus_{i=1}^n X_{n,i} \right], \mathbb{E}[\text{co } X_n] \right) = O_p \left(\sqrt{\frac{p_n}{n}} \right).$$

If, in addition, the Aumann expectation satisfies $\mathbb{E}[X_n] = \mathbb{E}[\text{co } X_n]$, then

$$\rho_H \left(\frac{1}{n} \bigoplus_{i=1}^n X_{n,i}, \mathbb{E}[X_n] \right) = O_p \left(\sqrt{\frac{p_n}{n}} + \frac{p_n}{n} \right).$$

Consequently, if $p_n/n \rightarrow 0$, then

$$\rho_H \left(\frac{1}{n} \bigoplus_{i=1}^n X_{n,i}, \mathbb{E}[X_n] \right) = O_p \left(\sqrt{\frac{p_n}{n}} \right).$$

Proof. Let $\hat{X}_n \equiv \frac{1}{n} \bigoplus_{i=1}^n X_{n,i}$. First, let us control the convexified average. Note that $\text{co}(\hat{X}_n) = \frac{1}{n} \bigoplus_{i=1}^n \text{co}(X_{n,i})$. Using $h(\text{co}(\hat{X}_n), u) = h(\hat{X}_n, u) = \frac{1}{n} \sum_{i=1}^n h(X_{n,i}, u)$ and $h(\mathbb{E}[\text{co}(X_n)], u) = \mathbb{E}[h(\text{co}(X_n), u)] = \mathbb{E}[h(X_n, u)]$, we can rewrite the Hausdorff distance between the convexi-

fied average and the Aumann expectation as the empirical process supremum:

$$\rho_H \left(\text{co}(\hat{X}_n), \mathbb{E}[\text{co}(X_n)] \right) = \sup_{u \in \mathbb{B}_n^*} \left| \frac{1}{n} \sum_{i=1}^n h(X_{n,i}, u) - \mathbf{E}[h(X_n, u)] \right|$$

where $\mathbb{B}_n^*$ denotes the finite-dimensional dual unit ball corresponding to the ambient space $\mathbb{R}^{p_n}$.

Define the support function class as $\mathcal{H}_n \equiv \{f_u : f_u(X) = h(X, u), u \in \mathbb{B}_n^*\}$, thus above is akin to $\sup_{h \in \mathcal{H}_n} |(P_n - P_{W^O})h|$, which is controlled by the entropy of the function class $\mathcal{H}_n$. From Lemma 4.10 in Pisier (1989) we have $N(\mathbb{B}_n^*, \delta, \|\cdot\|_\infty) \leq (1 + \frac{2}{\delta})^{p_n}$ combining with Lipschitz continuity of the support function, i.e. $|f_u(X) - f_v(X)| = |h(X, u) - h(X, v)| \leq \sup_{x \in X} |\langle x, u - v \rangle| \leq \sup_{x \in X} \|x\|_1 \|u - v\|_\infty \leq b \|u - v\|_\infty$, where b is the uniform deterministic bound from the statement of the lemma. Hence $\|f_u - f_v\|_{L_2(P_{W^O})} \leq b \|u - v\|_\infty$ and we have $N(\epsilon, \mathcal{H}_n, L_2(P_{W^O})) \leq N(\frac{\epsilon}{b}, \mathbb{B}_n^*, \|\cdot\|_\infty)$, thus

$$\log N(\epsilon, \mathcal{H}_n, L_2(P_{W^O})) \leq p_n \log \left(1 + \frac{2b}{\epsilon} \right) \leq p_n \log \left(\frac{Cb}{\epsilon} \right)$$

By Dudley's entropy theorem (Dudley (1967)) we have

$$\begin{aligned} \mathbb{E} \left[\sup_{u \in \mathbb{B}_n^*} \left| \frac{1}{n} \sum_{i=1}^n h(X_{n,i}, u) - \mathbf{E}[h(X_n, u)] \right| \right] &\leq \frac{C}{\sqrt{n}} \int_0^b \sqrt{p_n \log \left(\frac{Cb}{\epsilon} \right)} d\epsilon \\ &= C \sqrt{\frac{p_n}{n}} b \int_0^1 \sqrt{\log(C/t)} dt \\ &\leq Cb \sqrt{\frac{p_n}{n}} \end{aligned}$$

where the equality uses a change of variable and the last inequality uses $\int_0^1 \sqrt{\log(C/t)} dt < \infty$.

Using Markov's inequality it follows that

$$\sup_{u \in \mathbb{B}_n^*} \left| \frac{1}{n} \sum_{i=1}^n h(X_{n,i}, u) - \mathbf{E}[h(X_n, u)] \right| = O_p \left(\sqrt{\frac{p_n}{n}} \right).$$

Hence,

$$\rho_H \left(\text{co}(\hat{X}_n), \mathbb{E}[\text{co}(X_n)] \right) = O_p \left(\sqrt{\frac{p_n}{n}} \right).$$

It remains to pass from the convexified average back to the original, potentially non-convex, Minkowski average. For each n , the sets $X_{n,i}$ are compact subsets of the finite-dimensional space $\mathbb{R}^{p_n}$. Therefore, applying the finite-dimensional Shapley-Folkman-Starr theorem row by row gives

$$\rho_H \left(\hat{X}_n, \text{co}(\hat{X}_n) \right) \leq \frac{p_n}{n} \max_{1 \leq i \leq n} \text{diam}_1(X_{n,i}) \leq 2b \frac{p_n}{n}.$$

If $\mathbb{E}[X_n] = \mathbb{E}[\text{co}(X_n)]$, then by the triangle inequality,

$$\begin{aligned} \rho_H \left(\hat{X}_n, \mathbb{E}[X_n] \right) &\leq \rho_H \left(\hat{X}_n, \text{co}(\hat{X}_n) \right) + \rho_H \left(\text{co}(\hat{X}_n), \mathbb{E}[\text{co}(X_n)] \right) \\ &= O \left(\frac{p_n}{n} \right) + O_p \left(\sqrt{\frac{p_n}{n}} \right). \end{aligned}$$

Thus,

$$\rho_H \left(\frac{1}{n} \bigoplus_{i=1}^n X_{n,i}, \mathbb{E}[X_n] \right) = O_p \left(\sqrt{\frac{p_n}{n}} + \frac{p_n}{n} \right).$$

If $p_n/n \rightarrow 0$, then $p_n/n = o(\sqrt{p_n/n})$, and therefore

$$\rho_H \left(\frac{1}{n} \bigoplus_{i=1}^n X_{n,i}, \mathbb{E}[X_n] \right) = O_p \left(\sqrt{\frac{p_n}{n}} \right).$$

□

C.2 Auxiliary Lemmata

The first auxiliary Lemma establishes a uniform version of the triangular law of large numbers for the random sets established above.

Lemma 8. *Let Assumptions A1–A5 hold. Let $K_n \rightarrow \infty$, and let $Q_{\theta, K_n} \subset \mathbb{R}^{d_m K_n}$ denote the truncation of Q_θ to the first K_n instruments $g_1, \dots, g_{K_n}$. For $i = 1, \dots, n$, write $Q_{\theta, K_n, i} \equiv Q_{\theta, K_n}(W_i^O)$. Suppose that $\frac{d_m K_n}{n} \rightarrow 0$. Then*

$$\sup_{\theta \in \Theta} \rho_H \left(\frac{1}{n} \bigoplus_{i=1}^n Q_{\theta, K_n, i}, \mathbb{E}[Q_{\theta, K_n}] \right) = O_p \left(\sqrt{\frac{d_m K_n + d_\theta}{n}} \right).$$

Since d_m and d_θ are fixed, this is equivalently

$$\sup_{\theta \in \Theta} \rho_H \left(\frac{1}{n} \bigoplus_{i=1}^n Q_{\theta, K_n, i}, \mathbb{E}[Q_{\theta, K_n}] \right) = O_p \left(\sqrt{\frac{K_n}{n}} \right).$$

Proof. Let

$$\hat{Q}_{n, \theta, K_n} \equiv \frac{1}{n} \bigoplus_{i=1}^n Q_{\theta, K_n, i}.$$

First, control the convexified average. Since $Q_{\theta, K_n} \subset \mathbb{R}^{d_m K_n}$, the support-function representation of the Hausdorff distance for compact convex sets gives

$$\rho_H \left(\text{co} \left(\hat{Q}_{n, \theta, K_n} \right), \mathbb{E}[\text{co}(Q_{\theta, K_n})] \right) = \sup_{u \in B_{K_n}} \left| \frac{1}{n} \sum_{i=1}^n h(Q_{\theta, K_n, i}, u) - \mathbb{E}[h(Q_{\theta, K_n}, u)] \right|,$$

where

$$B_{K_n} \equiv \{u \in \mathbb{R}^{d_m K_n} : \|u\|_\infty \leq 1\}.$$

Here we used $h(\text{co}(A), u) = h(A, u)$, $h(n^{-1} \oplus_i A_i, u) = n^{-1} \sum_i h(A_i, u)$, and

$$h(\mathbb{E}[\text{co}(Q_{\theta, K_n})], u) = \mathbb{E}[h(Q_{\theta, K_n}, u)].$$

Define the support-function class

$$\mathcal{H}_n^\Theta \equiv \{f_{\theta,u}(W_i^O) = h(Q_{\theta,K_n}(W_i^O), u) : \theta \in \Theta, u \in B_{K_n}\}.$$

By Assumptions A2 and A3, there exists $b < \infty$, independent of n , such that

$$\sup_{\theta \in \Theta} \sup_{q \in Q_{\theta,K_n}} \|q\|_1 \leq b.$$

Indeed, one may take

$$b = d_m C_1 \sup_{x \in \mathcal{X}} \sum_{k=1}^{\infty} |g_k(x)| < \infty.$$

Hence the class has envelope b . Moreover, for $u, v \in B_{K_n}$,

$$|h(Q_{\theta,K_n}, u) - h(Q_{\theta,K_n}, v)| \leq b \|u - v\|_\infty.$$

Assumption A5 and the summability condition in Assumption A2 imply that, for some constant $\bar{L} < \infty$ independent of n ,

$$|h(Q_{\theta,K_n}, u) - h(Q_{\theta',K_n}, u)| \leq \bar{L} \|\theta - \theta'\|.$$

Therefore, covering both B_{K_n} and Θ gives

$$\log N(\epsilon, \mathcal{H}_n^\Theta, L_2(P_{W^O})) \leq d_m K_n \log \left(\frac{Cb}{\epsilon} \right) + d_\theta \log \left(\frac{C\bar{L}}{\epsilon} \right).$$

By Dudley's entropy theorem,

$$\begin{aligned} \mathbb{E} \left[\sup_{\theta \in \Theta} \sup_{u \in B_{K_n}} \left| \frac{1}{n} \sum_{i=1}^n \{h(Q_{\theta,K_n,i}, u) - \mathbb{E}[h(Q_{\theta,K_n}, u)]\} \right| \right] &\leq \frac{C}{\sqrt{n}} \int_0^b \sqrt{d_m K_n \log \left(\frac{Cb}{\epsilon} \right) + d_\theta \log \left(\frac{C\bar{L}}{\epsilon} \right)} d\epsilon \\ &\leq C \sqrt{\frac{d_m K_n + d_\theta}{n}}. \end{aligned}$$

Hence, by Markov's inequality,

$$\sup_{\theta \in \Theta} \rho_H \left(\text{co} \left(\hat{Q}_{n,\theta,K_n} \right), \mathbb{E}[\text{co}(Q_{\theta,K_n})] \right) = O_p \left(\sqrt{\frac{d_m K_n + d_\theta}{n}} \right).$$

It remains to remove convexification. For each n and each θ , $Q_{\theta,K_n,i}$ is a compact subset of the finite-dimensional space $\mathbb{R}^{d_m K_n}$. Thus the finite-dimensional Shapley-Folkman-Starr theorem gives, uniformly in θ ,

$$\rho_H \left(\hat{Q}_{n,\theta,K_n}, \text{co} \left(\hat{Q}_{n,\theta,K_n} \right) \right) \leq \frac{d_m K_n}{n} \max_{1 \leq i \leq n} \text{diam}_1(Q_{\theta,K_n,i}) \leq 2b \frac{d_m K_n}{n}.$$

By the convexification identity for the Aumann expectation,

$$\mathbb{E}[Q_{\theta,K_n}] = \mathbb{E}[\text{co}(Q_{\theta,K_n})].$$

Therefore, by the triangle inequality,

$$\begin{aligned} \sup_{\theta \in \Theta} \rho_H \left(\hat{Q}_{n,\theta,K_n}, \mathbb{E}[Q_{\theta,K_n}] \right) &\leq \sup_{\theta \in \Theta} \rho_H \left(\hat{Q}_{n,\theta,K_n}, \text{co} \left(\hat{Q}_{n,\theta,K_n} \right) \right) \\ &\quad + \sup_{\theta \in \Theta} \rho_H \left(\text{co} \left(\hat{Q}_{n,\theta,K_n} \right), \mathbb{E}[\text{co}(Q_{\theta,K_n})] \right) \\ &= O \left(\frac{d_m K_n}{n} \right) + O_p \left(\sqrt{\frac{d_m K_n + d_\theta}{n}} \right). \end{aligned}$$

Hence

$$\sup_{\theta \in \Theta} \rho_H \left(\frac{1}{n} \bigoplus_{i=1}^n Q_{\theta,K_n,i}, \mathbb{E}[Q_{\theta,K_n}] \right) = O_p \left(\sqrt{\frac{d_m K_n + d_\theta}{n}} + \frac{d_m K_n}{n} \right).$$

If $d_m K_n/n \rightarrow 0$, the second term is absorbed into the first. Since d_m and d_θ are fixed, this

yields

$$\sup_{\theta \in \Theta} \rho_H \left(\frac{1}{n} \bigoplus_{i=1}^n Q_{\theta, K_n, i}, \mathbb{E}[Q_{\theta, K_n}] \right) = O_p \left(\sqrt{\frac{K_n}{n}} \right).$$

□

The following Lemma links the deviation of directed Hausdorff distances to the undirected Hausdorff distances between two sets.

Lemma 9. *Let A and B be nonempty closed bounded subsets of a common normed space. Then*

$$|\vec{\rho}_H(0, A) - \vec{\rho}_H(0, B)| \leq \rho_H(A, B).$$

Proof. By the triangle inequality of hemi-metrics $q \in B$, $\vec{\rho}_H(0, A) \leq \vec{\rho}_H(0, q) + \vec{\rho}_H(q, A)$ (see Lemma 6.2.16 in Goubault-Larrecq, Ch. 6). Therefore,

$$\begin{aligned} \vec{\rho}_H(0, A) &\leq \min_{q \in B} \left(\vec{\rho}_H(0, q) + \vec{\rho}_H(q, A) \right) \leq \min_{q \in B} \left(\rho_H(0, q) + \rho_H(q, A) \right) \\ &\leq \vec{\rho}_H(0, B) + \rho_H(A, B). \end{aligned}$$

Hence $\vec{\rho}_H(0, A) - \vec{\rho}_H(0, B) \leq \rho_H(A, B)$. Reversing the inequality, for all $q \in A$ we similarly obtain $\vec{\rho}_H(0, B) \leq \vec{\rho}_H(0, q) + \vec{\rho}_H(q, B)$ so that

$$\begin{aligned} \vec{\rho}_H(0, B) &\leq \min_{q \in A} \left(\vec{\rho}_H(0, q) + \vec{\rho}_H(q, B) \right) \leq \min_{q \in A} \left(\rho_H(0, q) + \rho_H(q, B) \right) \\ &\leq \vec{\rho}_H(0, A) + \rho_H(A, B). \end{aligned}$$

Hence $\vec{\rho}_H(0, B) - \vec{\rho}_H(0, A) \leq \rho_H(A, B)$ Combining both inequalities gives the desired result.

□

Next we bound the deterministic approximation error from retaining only the first K

instruments $(g_1, \dots, g_K)$ instead of the full infinite instrument class.

Lemma 10. *Let A1-A3 hold. Let $Q_{\theta,K} \subset \mathbb{R}^{d_m K}$ denote the truncation of Q_θ to the first K instruments $g_1, \dots, g_K$, and let $\tilde{Q}_{\theta,K} \subset \ell_1$ denote its zero-padded embedding, $\tilde{Q}_{\theta,K}(w^o) = \lambda_K Q_{\theta,K}(w^o) = \lambda_K \pi_K Q_\theta(w^o)$. Then*

$$\sup_{\theta \in \Theta} \left| \vec{\rho}_H(0, \mathbb{E}[Q_\theta]) - \vec{\rho}_H(0, \mathbb{E}[\tilde{Q}_{\theta,K}]) \right| \rightarrow 0 \quad \text{as } K \rightarrow \infty.$$

Moreover, since $\vec{\rho}_H(0, \mathbb{E}[\tilde{Q}_{\theta,K}]) = \vec{\rho}_H(0, \mathbb{E}[Q_{\theta,K}])$, the finite-dimensional criterion satisfies

$$\sup_{\theta \in \Theta} |T(\theta) - T_K(\theta)| \leq d_m C_1 \sup_{x \in \mathcal{X}} \sum_{k > K} |g_k(x)| \rightarrow 0.$$

Proof. Fix $\theta \in \Theta$. By Lemma 9, $\left| \vec{\rho}_H(0, \mathbb{E}[Q_\theta]) - \vec{\rho}_H(0, \mathbb{E}[\tilde{Q}_{\theta,K}]) \right| \leq \rho_H(\mathbb{E}[Q_\theta], \mathbb{E}[\tilde{Q}_{\theta,K}])$. The right-hand side can be bounded using the support-function representation of the Hausdorff distance for closed convex bounded sets and the support-function representation of the Aumann expectation:

$$\begin{aligned} \rho_H(\mathbb{E}[Q_\theta], \mathbb{E}[\tilde{Q}_{\theta,K}]) &= \sup_{u \in B_1^\infty} \left| h(\mathbb{E}[Q_\theta], u) - h(\mathbb{E}[\tilde{Q}_{\theta,K}], u) \right| \\ &= \sup_{u \in B_1^\infty} \left| \mathbb{E}_{P_{W^O}}[h(Q_\theta, u)] - \mathbb{E}_{P_{W^O}}[h(\tilde{Q}_{\theta,K}, u)] \right| \\ &\leq \mathbf{E} \left[\sup_{u \in B_1^\infty} \left| h(Q_\theta, u) - h(\tilde{Q}_{\theta,K}, u) \right| \right]. \end{aligned}$$

The last inequality follows from $\sup_u |\mathbf{E}[Z_u]| \leq \mathbf{E}[\sup_u |Z_u|]$. By Lipschitz continuity of support functions and $\|u\|_\infty \leq 1$,

$$\sup_{u \in B_1^\infty} \left| h(Q_\theta, u) - h(\tilde{Q}_{\theta,K}, u) \right| \leq \rho_H(Q_\theta, \tilde{Q}_{\theta,K}).$$

Since $Q_\theta(W^O)$ and $\tilde{Q}_{\theta,K}(W^O)$ are generated from the same feasible set $D(W^O)$, we have

pathwise

$$\begin{aligned}
\rho_H(Q_\theta(W^O), \tilde{Q}_{\theta,K}(W^O)) &\leq \sup_{(y,x) \in D(W^O)} \|M_\theta(y, x) - \lambda_K \pi_K M_\theta(y, x)\|_1 \\
&= \sup_{(y,x) \in D(W^O)} \sum_{j=1}^{d_m} \sum_{k>K} |m_j(y, x, \theta)| |g_k(x)| \\
&\leq d_m C_1 \sup_{x \in \mathcal{X}} \sum_{k>K} |g_k(x)|,
\end{aligned}$$

where the last inequality follows from Assumption 3.

Combining the preceding displays yields, uniformly in θ ,

$$\left| \vec{\rho}_H(0, \mathbb{E}[Q_\theta]) - \vec{\rho}_H(0, \mathbb{E}[\tilde{Q}_{\theta,K}]) \right| \leq d_m C_1 \sup_{x \in \mathcal{X}} \sum_{k>K} |g_k(x)|.$$

Taking the supremum over $\theta \in \Theta$ and using the weighted-tail condition in Assumption A2 gives

$$\sup_{\theta \in \Theta} \left| \vec{\rho}_H(0, \mathbb{E}[Q_\theta]) - \vec{\rho}_H(0, \mathbb{E}[\tilde{Q}_{\theta,K}]) \right| \rightarrow 0.$$

Finally, under the zero-padding identification, $\vec{\rho}_H(0, \mathbb{E}[\tilde{Q}_{\theta,K}]) = \vec{\rho}_H(0, \mathbb{E}[Q_{\theta,K}])$ so the displayed bound also gives

$$\sup_{\theta \in \Theta} |T(\theta) - T_K(\theta)| \leq d_m C_1 \sup_{x \in \mathcal{X}} \sum_{k>K} |g_k(x)| \rightarrow 0$$

□

C.3 Main Proof of Theorem 2

We want to show $\rho_H(\hat{\Theta}_n, \Theta_I^*) \xrightarrow{P} 0$ by applying Theorem 3.1 in Chernozhukov, Hong, and Tamer (2007). Their result rests on condition C.1 of the same paper, which we later verify

for our setting. Define

$$a_n = \sqrt{\frac{n}{d_m K(n) + d_\theta}}, \quad \delta_K = \sup_{\theta \in \Theta} |T(\theta) - T_K(\theta)|,$$

and, for the verification of Chernozhukov, Hong, and Tamer (2007) condition C.1(d), define

$$b_n = \left(\delta_{K(n)} + \sqrt{\frac{d_m K(n) + d_\theta}{n}} \right)^{-1}.$$

By Lemma 10 and Assumption 2, $\delta_{K(n)} \rightarrow 0$. Since $d_m K(n)/n \rightarrow 0$, it follows that $b_n \rightarrow \infty$. Recall,

$$\begin{aligned} T(\theta) &\equiv \vec{\rho}_H(0, \mathbb{E}[Q_\theta]), \\ T_{K(n)}(\theta) &\equiv \vec{\rho}_H(0, \mathbb{E}[Q_{\theta, K(n)}]), \\ \hat{T}_n(\theta) &\equiv - \inf_{u \in B_{K(n)}} \frac{1}{n} \sum_{i=1}^n h(Q_{\theta, K(n), i}, u). \end{aligned}$$

Equivalently,

$$\hat{T}_n(\theta) = \vec{\rho}_H \left(0, \text{co} \left[\frac{1}{n} \bigoplus_{i=1}^n Q_{\theta, K(n), i} \right] \right).$$

We restate the estimator defined in the main text:

$$\begin{aligned} \hat{\Theta}_n &= \left\{ \theta \in \Theta : - \inf_{u \in B_{K(n)}} \frac{1}{n} \sum_{i=1}^n h(Q_{\theta, K(n), i}, u) \leq \frac{\tau_n}{a_n} \right\} \\ &= \{\theta \in \Theta : a_n \hat{T}_n(\theta) \leq \tau_n\}. \end{aligned}$$

with $K(n) \rightarrow \infty$, $d_m K(n)/n \rightarrow 0$, $\tau_n \rightarrow \infty$, and $\tau_n/a_n \rightarrow 0$.

With these we are equipped to check conditions (a)-(e) in condition C.1 of Chernozhukov, Hong, and Tamer (2007):

- a) Θ is a compact subset of $\mathbb{R}^d$

Holds by Assumption 4.

b) $T(\theta)$ is continuous and $\inf_{\theta \in \Theta} T(\theta) = 0$

By Lemma 8, $|T(\theta) - T(\theta')| \leq \rho_H(\mathbb{E}[Q_\theta], \mathbb{E}[Q_{\theta'}])$. For any feasible (y, x) , let $q_\theta = M_\theta(y, x)$ and $q_{\theta'} = M_{\theta'}(y, x)$. Then

$$\begin{aligned} \|q_\theta - q_{\theta'}\|_1 &\leq \sum_{j=1}^{d_m} \sum_{k \geq 1} |m_j(y, x, \theta) - m_j(y, x, \theta')| |g_k(x)| \\ &\leq \left[\sup_{x \in \mathcal{X}} \sum_{k \geq 1} |g_k(x)| \right] \sup_{(y, x) \in \mathcal{W}} \sum_{j=1}^{d_m} |m_j(y, x, \theta) - m_j(y, x, \theta')|. \end{aligned}$$

By Assumption 2 the bracketed term is finite, and by Assumption 5 the second factor converges to zero as $\theta' \rightarrow \theta$, uniformly on the compact domain. Thus $\rho_H(\mathbb{E}[Q_\theta], \mathbb{E}[Q_{\theta'}]) \rightarrow 0$, and hence T is continuous.

c) $\hat{T}_n(\theta)$ is jointly measurable.

Not strictly necessary according to Chernozhukov, Hong, and Tamer (2007). Nevertheless, it follows from the measurability of $Q_{\theta, K(n)}$, using the same arguments as in Theorem 1.

d) The sample criterion function is uniformly no smaller than the population function in large samples:

$$\begin{aligned} \sup_{\theta \in \Theta} (T(\theta) - \hat{T}_n(\theta))_+ &\leq \sup_{\theta \in \Theta} |T(\theta) - \hat{T}_n(\theta)| \\ &= \sup_{\theta \in \Theta} |T(\theta) - T_{K(n)}(\theta) + T_{K(n)}(\theta) - \hat{T}_n(\theta)| \\ &\leq \sup_{\theta \in \Theta} |T(\theta) - T_{K(n)}(\theta)| + \sup_{\theta \in \Theta} |T_{K(n)}(\theta) - \hat{T}_n(\theta)|. \end{aligned}$$

The first term is the deterministic truncation error. By Lemma 10,

$$\sup_{\theta \in \Theta} |T(\theta) - T_{K(n)}(\theta)| \leq d_m C_1 \sup_{x \in \mathcal{X}} \sum_{k > K(n)} |g_k(x)| \equiv \delta_{K(n)} \rightarrow 0.$$

For the second term, recall that

$$T_{K(n)}(\theta) = \vec{\rho}_H(0, \mathbb{E}[Q_{\theta, K(n)}]), \quad , \quad \widehat{T}_n(\theta) = \vec{\rho}_H\left(0, \text{co}\left[\frac{1}{n} \bigoplus_{i=1}^n Q_{\theta, K(n), i}\right]\right).$$

Hence, by Lemma 8 and the Lemma 9,

$$\begin{aligned} \sup_{\theta \in \Theta} |T_{K(n)}(\theta) - \widehat{T}_n(\theta)| &\leq \sup_{\theta \in \Theta} \rho_H\left(\mathbb{E}[Q_{\theta, K(n)}], \text{co}\left[\frac{1}{n} \bigoplus_{i=1}^n Q_{\theta, K(n), i}\right]\right) \\ &= O_p\left(\sqrt{\frac{d_m K(n) + d_\theta}{n}}\right). \end{aligned}$$

Therefore,

$$\begin{aligned} \sup_{\theta \in \Theta} (T(\theta) - \widehat{T}_n(\theta))_+ &= O_p\left(\delta_{K(n)} + \sqrt{\frac{d_m K(n) + d_\theta}{n}}\right) \\ &= O_p(1/b_n), \end{aligned}$$

By Assumption 2, Lemma 10, and $d_m K(n)/n \rightarrow 0$, we have $b_n \rightarrow \infty$. The term $\delta_{K(n)}$ captures the deterministic approximation error from using finitely many instruments instead of the full infinite instrument class, while the second term captures the sampling error of the growing-dimensional random set estimator.

- e) The sample criterion converges to the limit criterion function over the identified set Θ_I^*

at the $1/a_n$ rate.

$$\begin{aligned}
\sup_{\theta \in \Theta_I^*} \widehat{T}_n(\theta) &= \sup_{\theta \in \Theta_I^*} \vec{\rho}_H \left(0, \text{co} \left[\frac{1}{n} \bigoplus_{i=1}^n Q_{\theta, K(n), i} \right] \right) \\
&\leq \sup_{\theta \in \Theta_I^*} \vec{\rho}_H(0, \mathbb{E}[Q_{\theta, K(n)}]) \\
&\quad + \sup_{\theta \in \Theta_I^*} \rho_H \left(\mathbb{E}[Q_{\theta, K(n)}], \text{co} \left[\frac{1}{n} \bigoplus_{i=1}^n Q_{\theta, K(n), i} \right] \right) \\
&= \sup_{\theta \in \Theta_I^*} \rho_H \left(\mathbb{E}[Q_{\theta, K(n)}], \text{co} \left[\frac{1}{n} \bigoplus_{i=1}^n Q_{\theta, K(n), i} \right] \right) \\
&= O_p(1/a_n).
\end{aligned}$$

The first line is definitional. The second line follows from the inequality $\vec{\rho}_H(0, B) \leq \vec{\rho}_H(0, A) + \rho_H(A, B)$, applied with $A = \mathbb{E}[Q_{\theta, K(n)}]$ and $B = \text{co} \left[\frac{1}{n} \bigoplus_{i=1}^n Q_{\theta, K(n), i} \right]$. The third line uses the fact that if $\theta \in \Theta_I^*$, then $0 \in \mathbb{E}[Q_{\theta}]$. Lemmas 5 and 6 yield an $\eta \in \Gamma$ with $\int M_{\theta} d\eta = 0$, and Lemma 3 then produces an integrable selection $q_{\theta}(W^O) \in Q_{\theta}(W^O)$ satisfying $\mathbb{E}[q_{\theta}(W^O)] = 0$. Applying the coordinate projection $\pi_{K(n)}$ gives an integrable selection $\pi_{K(n)} q_{\theta}(W^O) \in Q_{\theta, K(n)}(W^O)$ satisfying $\mathbb{E}[\pi_{K(n)} q_{\theta}(W^O)] = \pi_{K(n)} \mathbb{E}[q_{\theta}(W^O)] = 0$. Therefore $0 \in \mathbb{E}[Q_{\theta, K(n)}]$ and $\vec{\rho}_H(0, \mathbb{E}[Q_{\theta, K(n)}]) = 0$.

Finally, the last rate follows from the convexified part of the uniform triangular-array law of large numbers and the definition

$$a_n = \sqrt{\frac{n}{d_m K(n) + d_{\theta}}}.$$

To complete the proof, after verifying condition C.1 in Chernozhukov, Hong, and Tamer (2007), the result follows from Theorem 3.1(i) in Chernozhukov, Hong, and Tamer (2007). Indeed, by part (e), $\mathcal{C}_n \equiv \sup_{\theta \in \Theta_I^*} a_n \widehat{T}_n(\theta) = O_p(1)$. Since $\tau_n \rightarrow \infty$, we have $\tau_n \geq \mathcal{C}_n$ with probability approaching one. Moreover, since $\tau_n/a_n \rightarrow 0$, the unscaled contour level vanishes. Therefore Theorem 3.1(i) of Chernozhukov, Hong, and Tamer (2007) implies $\rho_H(\widehat{\Theta}_n, \Theta_I^*) \xrightarrow{p} 0$. □

D Proof of Theorem 3

This appendix verifies the high-level conditions of Andrews and Shi (2017), applies their Theorem 6.1, and adapts the proof of their Theorem 5.1 to the full class of null pairs. The principal technical step is to verify manageability of the process indexed by $\mathcal{U} = B_1^\infty$.

D.1 Manageability of the support-function process

Recall the uniform coordinate envelopes

$$a_k \equiv \sup_{x \in \mathcal{X}} |g_k(x)|, \quad A \equiv \sum_{k=1}^{\infty} a_k.$$

The polynomial decay condition in Theorem 3 gives $A < \infty$ and an explicit rate for its tail.

Equip B_1^∞ with the normalized weighted semimetric

$$\rho_a(v, v') \equiv \frac{1}{d_m A} \sum_{\ell=1}^{d_m} \sum_{k=1}^{\infty} a_k |v_{\ell k} - v'_{\ell k}|.$$

Assumption 3 and the elementary inequality $|\sup_s f(s) - \sup_s g(s)| \leq \sup_s |f(s) - g(s)|$ imply, uniformly over $w^o \in \mathcal{W}^o$, $\theta \in \Theta$, and $u, u' \in \mathcal{U}$,

$$\begin{aligned} |\tilde{m}(w^o, \theta, u) - \tilde{m}(w^o, \theta, u')| &\leq C_1 \sum_{\ell=1}^{d_m} \sum_{k=1}^{\infty} a_k |u_{\ell k} - u'_{\ell k}| \\ &= M_0 \rho_a(u, u') \end{aligned}$$

with $M_0 \equiv d_m C_1 A$. Moreover, for every $u \in \mathcal{U}$,

$$\begin{aligned} |\tilde{m}(w^o, \theta, u)| &\leq C_1 \sum_{\ell=1}^{d_m} \sum_{k=1}^{\infty} a_k |u_{\ell k}| \\ &\leq d_m C_1 A = M_0. \end{aligned}$$

Thus M_0 is a uniform scalar envelope for the support-function moment class.

Fix an arbitrary sequence $\{(\theta_n, P_n) \in \mathcal{F}^+ : n \geq 1\}$, and let $\{W_{n,i}^O : i \leq n\}$ be a row-wise i.i.d. triangular array with $W_{n,i}^O \sim P_n$. For a realization ω , define

$$\mathcal{M}_{n,\omega}(\theta_n) \equiv \left\{ (\tilde{m}(W_{n,1}^O(\omega), \theta_n, u), \dots, \tilde{m}(W_{n,n}^O(\omega), \theta_n, u))' : u \in \mathcal{U} \right\} \subset \mathbb{R}^n.$$

For a subset V of a metric space, let $N(\xi, V)$ and $D(\xi, V)$ denote its covering and packing numbers. Appendix B of Andrews and Shi (2017) records the standard inequalities

$$N(\xi, V) \leq D(\xi, V) \leq N(\xi/2, V).$$

The same reference restates Definition 7.9 of Pollard (1990): the triangular array is manageable with respect to an envelope array if there is a deterministic $\lambda : (0, 1] \rightarrow \mathbb{R}$ such that

$$\int_0^1 \sqrt{\log \lambda(\xi)} d\xi < \infty$$

and, for every $\xi \in (0, 1]$, every realization ω , every n , and every nonnegative n -vector α ,

$$D(\xi \| \alpha \odot M_n \|_2, \alpha \odot \mathcal{M}_{n,\omega}(\theta_n)) \leq \lambda(\xi), \quad M_n \equiv (M_0, \dots, M_0)'.$$

The first lemma determines the number of instrument coordinates needed to control the coordinate-envelope tail

Lemma 11. *Suppose $C_a \equiv \sup_{k \geq 1} k^\beta a_k < \infty$ for some $\beta > 1$. Equivalently, $a_k \leq C_a k^{-\beta}$ for every k . For $\xi \in (0, 1]$, let*

$$K_\xi \equiv \left\lceil \left(\frac{4C_a}{(\beta-1)A\xi} \right)^{1/(\beta-1)} \right\rceil.$$

Then

$$\sum_{k>K_\xi} a_k \leq \frac{A\xi}{4}$$

and $K_\xi = O(\xi^{-1/(\beta-1)})$ as $\xi \downarrow 0$.

Proof. By the definition of C_a ,

$$a_k \leq C_a k^{-\beta} \quad \text{for every } k \geq 1.$$

Hence, for every integer $K \geq 1$, the integral comparison for the decreasing function $x \mapsto x^{-\beta}$ gives

$$\begin{aligned} \sum_{k>K} a_k &\leq C_a \sum_{k>K} k^{-\beta} \\ &\leq C_a \int_K^\infty x^{-\beta} dx \\ &= \frac{C_a}{\beta-1} K^{1-\beta}, \end{aligned}$$

where the last expression is finite because $\beta > 1$.

By definition,

$$K_\xi \geq \left(\frac{4C_a}{(\beta-1)A\xi} \right)^{1/(\beta-1)}.$$

Since $1-\beta < 0$, it follows that

$$\begin{aligned} K_\xi^{1-\beta} &\leq \left(\frac{4C_a}{(\beta-1)A\xi} \right)^{-1} \\ &= \frac{(\beta-1)A\xi}{4C_a}. \end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{k>K_\xi} a_k &\leq \frac{C_a}{\beta-1} K_\xi^{1-\beta} \\ &\leq \frac{A\xi}{4}.\end{aligned}$$

Finally, using $\lceil x \rceil \leq x + 1$,

$$K_\xi \leq 1 + \left(\frac{4C_a}{(\beta-1)A} \right)^{1/(\beta-1)} \xi^{-1/(\beta-1)}.$$

Hence

$$K_\xi = O\left(\xi^{-1/(\beta-1)}\right) \quad \text{as } \xi \downarrow 0.$$

□

The next lemma verifies the entropy-integrability requirement.

Lemma 12. *Let K_ξ be defined as in Lemma 11. For a sufficiently large constant $C_0 > 1$, define*

$$\lambda(\xi) \equiv \left(\frac{C_0}{\xi} \right)^{d_m K_\xi}, \quad \xi \in (0, 1].$$

If $\beta > 3/2$, then

$$\int_0^1 \sqrt{\log \lambda(\xi)} d\xi < \infty.$$

Proof. Let $q \equiv 1/(\beta-1)$. Lemma 11 implies that, for some finite constant C ,

$$\begin{aligned}\log \lambda(\xi) &= d_m K_\xi \log(C_0/\xi) \\ &\leq C \xi^{-q} \log(C_0/\xi)\end{aligned}$$

for all $\xi \in (0, 1]$. Consequently,

$$\int_0^1 \sqrt{\log \lambda(\xi)} d\xi \leq \sqrt{C} \int_0^1 \xi^{-q/2} \sqrt{\log(C_0/\xi)} d\xi.$$

Set $t = \log(1/\xi)$, so that $\xi = e^{-t}$ and $d\xi = -e^{-t}dt$. The right-hand side becomes

$$\sqrt{C} \int_0^\infty e^{-ct} \sqrt{t + \log C_0} dt, \quad c \equiv 1 - \frac{q}{2} = \frac{2\beta - 3}{2(\beta - 1)}.$$

Because $C_0 > 1$,

$$\begin{aligned} \sqrt{t + \log C_0} &\leq \sqrt{t} + \sqrt{\log C_0} \\ &\leq 1 + t + \sqrt{\log C_0}. \end{aligned}$$

Hence, if $c > 0$,

$$\begin{aligned} \int_0^\infty e^{-ct} \sqrt{t + \log C_0} dt &\leq \int_0^\infty e^{-ct} (1 + t + \sqrt{\log C_0}) dt \\ &= \frac{1 + \sqrt{\log C_0}}{c} + \frac{1}{c^2} < \infty. \end{aligned}$$

It follows that

$$\int_0^1 \sqrt{\log \lambda(\xi)} d\xi < \infty.$$

Finally, $c > 0$ is equivalent to $\beta > 3/2$. □

Lemma 13. *Assumption PS2 of Andrews and Shi (2017) holds in the present setting with*

$$\lambda(\xi) = \left(\frac{C_0}{\xi} \right)^{d_m K_\xi},$$

where K_ξ is the integer chosen in Lemma 11 so that the coordinate-envelope tail is at most $A\xi/4$.

Proof. Fix $\xi \in (0, 1]$. Let $\Pi_{K_\xi} u$ retain the first K_ξ instrument coordinates in each primitive-moment block of u and set the remaining coordinates equal to zero. Cover $[-1, 1]^{d_m K_\xi}$ in the sup norm by a grid $\mathcal{U}_\xi$ of radius $\xi/4$. For a constant $C_0 > 1$ independent of ξ , the grid may be chosen so that

$$|\mathcal{U}_\xi| \leq \left(\frac{C_0}{\xi} \right)^{d_m K_\xi}.$$

For every $u \in \mathcal{U}$, choose $v \in \mathcal{U}_\xi$ within $\xi/4$ of $\Pi_{K_\xi} u$ in the sup norm and extend v by zeros. Then

$$\begin{aligned} \rho_a(u, v) &\leq \frac{1}{d_m A} \sum_{\ell=1}^{d_m} \left\{ \sum_{k \leq K_\xi} a_k |u_{\ell k} - v_{\ell k}| + \sum_{k > K_\xi} a_k |u_{\ell k}| \right\} \\ &\leq \frac{\xi}{4} + \frac{\xi}{4} = \frac{\xi}{2}. \end{aligned}$$

Thus $\mathcal{U}_\xi$ is a $\xi/2$ cover of B_1^∞ under ρ_a .

Let $\alpha = (\alpha_1, \dots, \alpha_n)'$ be nonnegative. The Lipschitz bound above implies

$$\begin{aligned} &\left\| \alpha \odot \left((\tilde{m}(W_{n,i}^O, \theta_n, u))_{i=1}^n - (\tilde{m}(W_{n,i}^O, \theta_n, v))_{i=1}^n \right) \right\|_2 \\ &\leq \rho_a(u, v) \|\alpha \odot M_n\|_2. \end{aligned}$$

The same direction grid therefore covers the sample paths. The packing–covering inequality from Andrews and Shi (2017) then gives

$$D(\xi \|\alpha \odot M_n\|_2, \alpha \odot \mathcal{M}_{n,\omega}(\theta_n)) \leq \left(\frac{C_0}{\xi} \right)^{d_m K_\xi} = \lambda(\xi).$$

Lemma 12 verifies the required entropy integral. Hence the support-function process is manageable in the sense of Definition 7.9 of Pollard (1990). $\square$

Since the preceding bounds are deterministic and do not depend on the sequence $\{P_n\}$,

the same manageability function λ applies to every sequence $\{(\theta_n, P_n) \in \mathcal{F}^+\}$.

D.2 Verification of the Andrews–Shi assumptions

We apply Andrews and Shi (2017) to the scalar process $\{\tilde{m}(\cdot, \theta, u) : u \in \mathcal{U}\}$. In their notation there is one moment inequality, their index τ is $u \in \mathcal{U}$, the conditioning variable is degenerate, the instrument class is the singleton containing the constant function, and the weight function is a point mass on it. Their test statistic, GMS function and bootstrap critical value then reduce to $T_n^{AS}(\theta)$, $\bar{\varphi}_n(\theta, u)$ and $c_{n,1-\alpha}^{AS}(\theta)$. Recall the envelope and Lipschitz bounds established above, which hold uniformly in (w^o, θ) :

$$|\tilde{m}(w^o, \theta, u)| \leq M_0, \quad |\tilde{m}(w^o, \theta, u) - \tilde{m}(w^o, \theta, u')| \leq M_0 \rho_a(u, u'). \quad (17)$$

For $(\theta, P) \in \mathcal{F}^+$ and $u, u' \in \mathcal{U}$ define

$$\begin{aligned} \nu_{n,P}(\theta, u) &\equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\tilde{m}(W_i^O, \theta, u) - \mathbf{E}_P[\tilde{m}(W_i^O, \theta, u)]\}, \\ h_{1,n,P}(\theta, u) &\equiv \sqrt{n} \mathbf{E}_P[\tilde{m}(W_i^O, \theta, u)], \\ h_{2,P}(\theta; u, u') &\equiv \text{Cov}_P(\tilde{m}(W_i^O, \theta, u), \tilde{m}(W_i^O, \theta, u')). \end{aligned}$$

Let $\hat{h}_{2,n}(\theta; u, u')$ denote the sample analogue of $h_{2,P}(\theta; u, u')$. Then $\sqrt{n} \bar{m}_n(\theta, u) = \nu_{n,P}(\theta, u) + h_{1,n,P}(\theta, u)$ and $\bar{\Sigma}_n(\theta, u) = \hat{h}_{2,n}(\theta; u, u) + \epsilon$. Moreover, $h_{1,n,P}(\theta, \cdot) \geq 0$ whenever $(\theta, P) \in \mathcal{F}$. Kernels on $\mathcal{U} \times \mathcal{U}$ are compared in the uniform metric $d(h, h') \equiv \sup_{u, u' \in \mathcal{U}} |h(u, u') - h'(u, u')|$.

(i) **PS1.** Under every $(\theta, P) \in \mathcal{F}^+$ the observations are i.i.d. and $\sigma_P(\theta) = 1$. By (17),

M_0 is a deterministic envelope, so $\sup_{P \in \mathcal{P}} \mathbf{E}_P[M_0^{2+\delta}] < \infty$ for every $\delta > 0$.

(ii) **PS2.** This is Lemma 13, whose bound holds for every sequence in $\mathcal{F}^+$ with row-wise i.i.d. arrays. The map $w^o \mapsto \tilde{m}(w^o, \theta, u) = h(Q_\theta(w^o), u)$ is measurable, being the support function of the random compact set of Lemma 2. By (17), suprema over $\mathcal{U}$

coincide with suprema over a countable ρ_a -dense subset.

- (iii) **PS3.** With a degenerate conditioning variable, the moment inequality reads $\mathbf{E}_P[\tilde{m}(W_i^O, \theta, u)] \geq 0$ for all $u \in \mathcal{U}$. By Theorem 1, this holds if and only if $\theta \in \Theta_I^*(P)$.
- (iv) **SIG1 and SIG2.** Both hold identically, because $\hat{\sigma}_n(\theta) = \sigma_P(\theta) = 1$.
- (v) **MFA.** If $\theta \notin \Theta_I^*(P)$, Theorem 1 gives some $u_{\theta,P} \in \mathcal{U}$ with $\mathbf{E}_P[\tilde{m}(W_i^O, \theta, u_{\theta,P})] < 0$. With a degenerate conditioning variable, this is a violation on a set of probability one.
- (vi) **PS4 and PS5.** The constant instrument satisfies Assumption M of Andrews and Shi (2017), and S_1 satisfies their Assumptions S1 and S2 by Lemma 1 of Andrews and Shi (2013). Their Lemma D.2 derives PS4 and PS5 from these assumptions together with PS1 and PS2. As formulated there, PS4 and PS5 apply to every sequence (n_r, θ_r, P_r) with (n_r) strictly increasing, $(\theta_r, P_r) \in \mathcal{F}^+$ and $d(h_{2,P_r}(\theta_r; \cdot, \cdot), h_{2,0}) \rightarrow 0$, for *some* covariance kernel $h_{2,0}$, whether or not it is attained.

- PS4 gives $\nu_{n_r, P_r}(\theta_r, \cdot) \Rightarrow \nu_{h_{2,0}}$ and $d(\hat{h}_{2,n_r}(\theta_r), h_{2,0}) \xrightarrow{p} 0$.
- PS5 gives the same two conclusions for the bootstrap process $\sqrt{n_r}(\bar{m}_{n_r}^* - \bar{m}_{n_r})(\theta_r, \cdot)$ and the bootstrap kernel $\hat{h}_{2,n_r}^*(\theta_r)$. These hold conditionally on almost every sample path along which $d(\hat{h}_{2,n_r}(\theta_r), h_{2,0}) \rightarrow 0$.

Next let us show relative compactness of the covariance kernels. Let $\mathcal{H}_{2,P} \equiv \{h_{2,P}(\theta; \cdot, \cdot) : (\theta, P) \in \mathcal{F}\}$. Identify directions at zero ρ_a -distance. By (17), such directions generate identical moment functions, so all kernels are defined on the quotient. The quotient is compact. Given $\{u^{(r)}\} \subseteq \mathcal{U}$, a diagonal argument yields a subsequence converging coordinatewise to some $u \in \mathcal{U}$. For every K , this subsequence satisfies $\limsup_r \rho_a(u^{(r)}, u) \leq (2/A) \sum_{k>K} a_k$, which vanishes as $K \rightarrow \infty$ because $A < \infty$. Write $f_u \equiv \tilde{m}(W_i^O, \theta, u)$. Expanding $h_{2,P}(\theta; u, u') - h_{2,P}(\theta; v, v') = \text{Cov}_P(f_u - f_v, f_{u'}) + \text{Cov}_P(f_v, f_{u'} - f_{v'})$ and applying

Cauchy–Schwarz together with (17) gives, uniformly over $(\theta, P) \in \mathcal{F}$,

$$|h_{2,P}(\theta; u, u')| \leq M_0^2, \quad |h_{2,P}(\theta; u, u') - h_{2,P}(\theta; v, v')| \leq M_0^2 \{\rho_a(u, v) + \rho_a(u', v')\}.$$

Hence $\mathcal{H}_{2,\mathcal{P}}$ is uniformly bounded and uniformly equicontinuous on the compact quotient of $\mathcal{U} \times \mathcal{U}$. By the Arzelà–Ascoli theorem, it is relatively compact under d , and its closure $\overline{\mathcal{H}}_{2,\mathcal{P}}$ is compact. Uniform limits preserve symmetry and positive semidefiniteness, so every element of $\overline{\mathcal{H}}_{2,\mathcal{P}}$ is a covariance kernel. Such a limit need not be attained by any $(\theta, P) \in \mathcal{F}$. For a kernel h on $\mathcal{U} \times \mathcal{U}$, define the semimetric $\rho_h(u, v) \equiv \{h(u, u) - 2h(u, v) + h(v, v)\}^{1/2}$.

- *Domination.* For $h \in \mathcal{H}_{2,\mathcal{P}}$, $\rho_h(u, v)$ is the $L_2(P)$ distance between f_u and f_v , so $\rho_h \leq M_0 \rho_a$ by (17). Since ρ_h is d -continuous in h , the bound extends to $\overline{\mathcal{H}}_{2,\mathcal{P}}$.
- *Entropy.* The proof of Lemma 13 shows $N(\varepsilon, \mathcal{U}, \rho_a) \leq \lambda(2\varepsilon)$ for $\varepsilon \in (0, 1/2]$, and $\int_0^1 \sqrt{\log \lambda(\xi)} d\xi < \infty$ by Lemma 12.

Dudley’s entropy bound (Dudley 1967) therefore yields:

- (G1) for every $h \in \overline{\mathcal{H}}_{2,\mathcal{P}}$ there is a mean-zero Gaussian process ν_h on $\mathcal{U}$ with covariance h and bounded, uniformly ρ_a -continuous sample paths, and $\lim_{\delta \downarrow 0} \sup_{h \in \overline{\mathcal{H}}_{2,\mathcal{P}}} \mathbf{E} \sup_{\rho_a(u,v) < \delta} |\nu_h(u) - \nu_h(v)| = 0$;
- (G2) if $h_r, h \in \overline{\mathcal{H}}_{2,\mathcal{P}}$ and $d(h_r, h) \rightarrow 0$, then $\nu_{h_r} \Rightarrow \nu_h$ in $\ell^\infty(\mathcal{U})$, because the finite-dimensional laws converge and (G1) gives asymptotic uniform equicontinuity.

For $h \in \overline{\mathcal{H}}_{2,\mathcal{P}}$ and $h_1 : \mathcal{U} \rightarrow [0, \infty)$, define

$$T(h_1, h) \equiv \sup_{u \in \mathcal{U}} \left[\frac{\nu_h(u) + h_1(u)}{\sqrt{h(u, u) + \epsilon}} \right]_-^2.$$

This is a lower semicontinuous functional of the continuous path ν_h , hence a random variable. Let $c_0(h_1, h, 1 - \alpha)$ denote its $(1 - \alpha)$ -quantile. When h_1 is data dependent, such as $\overline{\varphi}_n(\theta, \cdot)$, c_0 is evaluated with h_1 held fixed and ν_h independent of the data.

Finally, let ν, ν' be functions on $\mathcal{U}$ with sup norms at most B , let $s, s' \geq \sqrt{\epsilon}$, and let $\mu : \mathcal{U} \rightarrow [0, \infty)$ be arbitrary. Because $[x + \mu]_- \leq [x]_-$ and $|[a]_- - [b]_-| \leq |a - b|$,

$$\left| \sup_u \left[\frac{\nu(u) + \mu(u)}{s(u)} \right]_-^2 - \sup_u \left[\frac{\nu'(u) + \mu(u)}{s'(u)} \right]_-^2 \right| \leq \frac{2B}{\epsilon} \left(\|\nu - \nu'\|_\infty + \frac{B}{\sqrt{\epsilon}} \|s - s'\|_\infty \right). \quad (18)$$

The bound does not depend on μ . This is what permits nonnegative slackness functions that need not converge.

Theorem 5.1 of Andrews and Shi (2017) rests on their Lemma D.1, Theorem D.3(a) and Lemma D.4. All of these are *stated* for compact subsets of the attainable kernel class, and $\overline{\mathcal{H}}_{2,\mathcal{P}}$ need not be such a set. In their proofs, and in the proofs of Theorem 1 and Lemmas A2–A4 of Andrews and Shi (2013) to which they refer, compactness is used once: to pass from a near-extremal sequence of null pairs to a subsequence whose kernels converge uniformly to some covariance kernel. The subsequent steps use only PS4, PS5, monotonicity and continuity of S , GMS1, and $\nu_{h_{2,r}} \Rightarrow \nu_{h_{2,0}}$. The following lemma repeats those steps along a sequence that is kernel-convergent by hypothesis. The only properties of the limit it uses are (G1)–(G2), which hold on all of $\overline{\mathcal{H}}_{2,\mathcal{P}}$.

Lemma 14. *Let the conditions of Theorem 3 hold, let (n_r) be strictly increasing, and let $(\theta_r, P_r) \in \mathcal{F}$ satisfy $d(h_{2,P_r}(\theta_r; \cdot, \cdot), h_{2,0}) \rightarrow 0$ for some covariance kernel $h_{2,0}$. Then*

$$\liminf_{r \rightarrow \infty} P_r^{n_r} \{ \theta_r \in CS_{n_r}^{AS}(1 - \alpha) \} \geq 1 - \alpha.$$

Proof. We use the following abbreviations: $h_{1,r} \equiv h_{1,n_r,P_r}(\theta_r, \cdot) \geq 0$, $h_{2,r} \equiv h_{2,P_r}(\theta_r; \cdot, \cdot)$, $\nu_r \equiv \nu_{n_r,P_r}(\theta_r, \cdot)$, $\overline{\varphi}_r \equiv \overline{\varphi}_{n_r}(\theta_r, \cdot)$, $s_r(u) \equiv \{\overline{\Sigma}_{n_r}(\theta_r, u)\}^{1/2} \geq \sqrt{\epsilon}$, $s_{2,r}(u) \equiv \{h_{2,r}(u, u) + \epsilon\}^{1/2}$, $s_0(u) \equiv \{h_{2,0}(u, u) + \epsilon\}^{1/2}$, and $c_r \equiv c_0(h_{1,r}, h_{2,r}, 1 - \alpha)$. Since $h_{2,r} \in \mathcal{H}_{2,\mathcal{P}}$, the limit satisfies $h_{2,0} \in \overline{\mathcal{H}}_{2,\mathcal{P}}$. Hence (G1), (G2) and item (vi) apply, while $h_{1,r}$ need not converge. Fix $\delta \in (0, \eta)$.

(a) *Oracle upper-tail bound:* $\limsup_r P_r^{n_r} \{ T_{n_r}^{AS}(\theta_r) > c_r + \delta \} \leq \alpha$. By PS4, $(\nu_r, s_r) \Rightarrow (\nu_{h_{2,0}}, s_0)$. By (G2), $\nu_{h_{2,r}} \Rightarrow \nu_{h_{2,0}}$, and $s_{2,r} \rightarrow s_0$ uniformly. By the almost sure representation

theorem (Pollard 1990, Theorem 9.4), there exist two pairs of random elements:

- $(\tilde{\nu}_r, \tilde{s}_r)$ with the law of (ν_r, s_r) , and $\tilde{\nu}_0$ with the law of $\nu_{h_{2,0}}$, such that $\|\tilde{\nu}_r - \tilde{\nu}_0\|_\infty + \|\tilde{s}_r - s_0\|_\infty \rightarrow 0$ almost surely;
- $\tilde{\nu}_r^G$ with the law of $\nu_{h_{2,r}}$, and $\tilde{\nu}_0'$ with the law of $\nu_{h_{2,0}}$, such that $\|\tilde{\nu}_r^G - \tilde{\nu}_0'\|_\infty \rightarrow 0$ almost surely.

Let $T_{r,0} \equiv \sup_u [(\nu_{h_{2,0}}(u) + h_{1,r}(u))/s_0(u)]_-^2$. The limits have bounded paths, so we may apply (18) with $\mu = h_{1,r}$ and B equal to the sup norm of the limit path plus one. This shows that $T_{n_r}^{AS}(\theta_r)$ and $T(h_{1,r}, h_{2,r})$ can each be coupled with a copy of $T_{r,0}$ such that the difference tends to zero almost surely. Therefore

$$P_r^{n_r}\{T_{n_r}^{AS}(\theta_r) > c_r + \delta\} \leq P\{T_{r,0} > c_r + \delta/2\} + o(1) \leq P\{T(h_{1,r}, h_{2,r}) > c_r\} + o(1) \leq \alpha + o(1).$$

This is the argument of Theorem D.3(a) of Andrews and Shi (2017) combined with Lemma A2 of Andrews and Shi (2013).

(b) *GMS adjustment versus true slackness.* Let $R_r \equiv \sup_u |\nu_r(u)|$, which is $O_p(1)$ by PS4 and the continuous mapping theorem. Consider a direction u .

- If $\xi_{n_r}(\theta_r, u) > 1$, then $\nu_r(u) + h_{1,r}(u) > \kappa_{n_r} s_r(u)$, so $h_{1,r}(u) > \kappa_{n_r} s_r(u) - R_r$, while $\bar{\varphi}_r(u) = B_{n_r} s_r(u)$.
- If $\xi_{n_r}(\theta_r, u) \leq 1$, then $\bar{\varphi}_r(u) = 0 \leq h_{1,r}(u)$.

Once $\kappa_{n_r} > B_{n_r}$, since $s_r \geq \sqrt{\epsilon}$, the inequality $\bar{\varphi}_r \leq h_{1,r}$ therefore holds simultaneously over $\mathcal{U}$ on the event $E_r \equiv \{R_r \leq (\kappa_{n_r} - B_{n_r})\sqrt{\epsilon}\}$. By GMS1, $(\kappa_{n_r} - B_{n_r})\sqrt{\epsilon} = \kappa_{n_r}(1 - B_{n_r}/\kappa_{n_r})\sqrt{\epsilon} \rightarrow \infty$, so $P_r^{n_r}(E_r) \rightarrow 1$. On E_r , every path of $\nu_{h_{2,r}}$ satisfies $[\nu_{h_{2,r}} + \bar{\varphi}_r]_- \geq [\nu_{h_{2,r}} + h_{1,r}]_-$ pointwise. Hence $T(\bar{\varphi}_r, h_{2,r}) \geq T(h_{1,r}, h_{2,r})$ pathwise for the same Gaussian process, and

$$c_0(\bar{\varphi}_r, h_{2,r}, 1 - \alpha) \geq c_r \quad \text{on } E_r.$$

Replacing the true slackness by the smaller GMS adjustment weakens the inequalities in the Gaussian experiment. This enlarges the statistic and its quantile, which is the sense in which GMS is conservative.

(c) *Bootstrap critical value:* $P_r^{n_r}\{c_{n_r,1-\alpha}^{AS}(\theta_r) < c_0(\bar{\varphi}_r, h_{2,r}, 1 - \alpha) + \delta\} \rightarrow 0$. By PS4, $d(\hat{h}_{2,n_r}(\theta_r), h_{2,0}) \xrightarrow{P} 0$. It therefore suffices to argue along a further subsequence of an arbitrary subsequence on which this convergence holds almost surely. Fix such a sample path. Conditionally on it, $\bar{\varphi}_r$ is a fixed nonnegative function, and by PS5, $(\nu_r^*, s_r^*) \Rightarrow (\nu_{h_{2,0}}, s_0)$, where $\nu_r^* \equiv \sqrt{n_r}(\bar{m}_{n_r}^* - \bar{m}_{n_r})(\theta_r, \cdot)$ and $s_r^*(u) \equiv \{\bar{\Sigma}_{n_r}^*(\theta_r, u)\}^{1/2}$. Repeat the argument of (a) with $(\nu_r^*, s_r^*, \bar{\varphi}_r)$ in place of $(\nu_r, s_r, h_{1,r})$, and let P^* denote bootstrap probability and T_r^* the bootstrap statistic. For every $\xi > 0$ and all constants x_r ,

$$\limsup_r [P^*\{T_r^* \leq x_r\} - P\{T(\bar{\varphi}_r, h_{2,r}) \leq x_r + \xi\}] \leq 0.$$

Now take $x_r \equiv c_0(\bar{\varphi}_r, h_{2,r}, 1 - \alpha) + \delta - \eta$ and $\xi \in (0, \eta - \delta)$.

- Since $x_r + \xi$ lies below the $(1 - \alpha)$ -quantile, $P\{T(\bar{\varphi}_r, h_{2,r}) \leq x_r + \xi\} < 1 - \alpha$.
- Hence $P^*\{T_r^* \leq x_r\} < 1 - \alpha + \eta$ for all large r , so x_r lies below the conditional $(1 - \alpha + \eta)$ -quantile.
- Therefore $c_{n_r,1-\alpha}^{AS}(\theta_r) > x_r + \eta = c_0(\bar{\varphi}_r, h_{2,r}, 1 - \alpha) + \delta$ for all large r .

Bounded convergence yields the claim. This is Lemma D.4 of Andrews and Shi (2017); it uses the fixed $\eta > \delta$.

Combining (a), (b) and (c),

$$P_r^{n_r}\{\theta_r \notin CS_{n_r}^{AS}(1 - \alpha)\} \leq P_r^{n_r}\{T_{n_r}^{AS}(\theta_r) > c_r + \delta\} + P_r^{n_r}(E_r^c) + o(1) \leq \alpha + o(1).$$

□

Finally, let us establish uniformity by contradiction. Suppose the coverage claim of Theorem 3 fails. Then there exist $\delta_0 > 0$, a strictly increasing sequence (n_r) , and pairs

$(\theta_r, P_r) \in \mathcal{F}$ with $P_r^{n_r}\{\theta_r \in CS_{n_r}^{AS}(1 - \alpha)\} \leq 1 - \alpha - \delta_0$ for all r . By relative compactness of $\mathcal{H}_{2,\mathcal{P}}$, a subsequence satisfies $d(h_{2,P_r}(\theta_r; \cdot, \cdot), h_{2,0}) \rightarrow 0$ for some $h_{2,0} \in \overline{\mathcal{H}_{2,\mathcal{P}}}$, which Lemma 14 contradicts. Hence

$$\liminf_{n \rightarrow \infty} \inf_{(\theta, P) \in \mathcal{F}} P^n\{\theta \in CS_n^{AS}(1 - \alpha)\} \geq 1 - \alpha.$$

The kernel convergence used here is extracted from relative compactness, which rests only on (17). Neither θ_r nor $h_{2,0}$ needs to correspond to a limiting pair, so no continuity of $\theta \mapsto h_{2,P}(\theta; \cdot, \cdot)$ is used.

The second claim of Theorem 3 concerns power against fixed alternatives. Fix $P \in \mathcal{P}$ and $\theta \in \Theta \setminus \Theta_I^*(P)$. Theorem 6.1 of Andrews and Shi (2017) is a fixed-alternative result, not restricted to compact kernel sets. It is a corollary of their Theorem E.1, whose conditions all hold here:

- PS4 and PS5, by item (vi) applied to the constant sequence;
- MFA by item (v), and SIG2 by item (iv);
- CI and MQ, because the single constant instrument detects the violation in (v) and the weight function is a point mass;
- S1, S3 and S4, because S_1 is nonincreasing, positive if and only if its argument is negative, and homogeneous of degree two.

Therefore

$$P^n\{T_n^{AS}(\theta) > c_{n,1-\alpha}^{AS}(\theta)\} \longrightarrow 1.$$

This proves Theorem 3. □

D.3 Finite and truncated direction classes

Proof of Corollary 2. Throughout, direction sets are nonempty, and the restricted statistics and conditional bootstrap quantiles are assumed measurable.

(a) Consider a strictly increasing sequence (n_r) and a sequence $(\theta_r, P_r, \mathcal{V}_r)$ with $\mathcal{V}_r \subseteq \mathcal{U}$ and $(\theta_r, P_r) \in \mathcal{F}_{\mathcal{V}_r} \subseteq \mathcal{F}^+$. The envelope bound, the covariance Lipschitz bound and the domination $\rho_h \leq M_0 \rho_a$ used in the proof of Theorem 3 hold uniformly on $\mathcal{F}^+$. Consequently, the covariance-kernel family $\{h_{2,P}(\theta; \cdot, \cdot) : (\theta, P) \in \mathcal{F}^+\}$ is relatively compact by the same Arzelà–Ascoli argument, and (G1)–(G2) hold on its closure. Suppose that the kernels $h_{2,P_r}(\theta_r; \cdot, \cdot)$ converge uniformly.

Apply PS4 and PS5, which are stated on $\mathcal{F}^+$, to the empirical and bootstrap processes on the full index set $\mathcal{U}$. The comparison bound (18) remains valid when its supremum is restricted to $\mathcal{V}_r$. Its right-hand side is bounded by the same full-space process and covariance-estimation errors, independently of $\mathcal{V}_r$. Thus these errors vanish along the sequence even when $\mathcal{V}_r$ changes with r . Moreover, $(\theta_r, P_r) \in \mathcal{F}_{\mathcal{V}_r}$ gives $h_{1,n_r,P_r}(\theta_r, u) \geq 0$ for every $u \in \mathcal{V}_r$. This is the nonnegativity used in the oracle-tail and GMS comparisons in Steps (a) and (b) of Lemma 14. The bootstrap-quantile comparison in Step (c) uses only $\bar{\varphi}_{n_r} \geq 0$. These comparisons therefore apply with every supremum restricted to $\mathcal{V}_r$, including in the Gaussian statistic $T(\cdot, \cdot)$ and its quantile c_0 . The bound $R_r = \sup_{u \in \mathcal{U}} |\nu_r(u)| = O_p(1)$ in Step (b) is still taken over the full index set. The fixed critical-value adjustment η supplies the same quantile margin as in that lemma.

It follows that the sequential coverage bound holds along every such triple sequence with uniformly convergent kernels. If the uniform bound in (a) failed, there would exist $\delta_0 > 0$, a strictly increasing sequence (n_r) , and triples as above whose coverage probabilities are at most $1 - \alpha - \delta_0$ for all r . Relative compactness supplies a kernel-convergent subsequence, contradicting the sequential bound. This proves (a).

(b) Fix P , $\theta \notin \Theta_O(\mathcal{V})(P)$, and a direction $u_0 \in \mathcal{V}$ with $\mathbf{E}_P[\tilde{m}(W^O, \theta, u_0)] < 0$. The law of large numbers makes the test statistic over $\mathcal{V}$ diverge at order n . Using the same per-

direction GMS adjustments and bootstrap draws, each bootstrap statistic over $\mathcal{V}$ is bounded by its counterpart over $\mathcal{U}$. The latter has an $O_p(1)$ conditional critical value under the fixed law P , by the bootstrap-process bound used in the proof of Theorem E.1 of Andrews and Shi (2017). Hence the critical value over $\mathcal{V}$ is $O_p(1)$, and the rejection probability tends to one.

(c) Coverage follows from (a), since $B_{K(n)} \subseteq \mathcal{U}$ and $\mathcal{F} \subseteq \mathcal{F}_{B_{K(n)}}$. For power, fix P and $\theta \notin \Theta_I^*(P)$, and choose $u_0 \in \mathcal{U}$ such that $\mathbf{E}_P[\tilde{m}(W^O, \theta, u_0)] = -2c < 0$. Put $u_n = \Pi_{K(n)} u_0 \in B_{K(n)}$. By (17),

$$|\mathbf{E}_P[\tilde{m}(W^O, \theta, u_0)] - \mathbf{E}_P[\tilde{m}(W^O, \theta, u_n)]| \leq d_m C_1 \sum_{k > K(n)} a_k \longrightarrow 0.$$

Thus $\mathbf{E}_P[\tilde{m}(W^O, \theta, u_n)] \leq -c$ for all sufficiently large n . PS4, applied to the constant sequence (θ, P) on $\mathcal{U}$, gives

$$\sup_{u \in \mathcal{U}} |\overline{m}_n(\theta, u) - \mathbf{E}_P[\tilde{m}(W^O, \theta, u)]| = O_p(n^{-1/2}),$$

so $\overline{m}_n(\theta, u_n) \leq -c/2$ with probability approaching one. Since $\overline{\Sigma}_n \leq M_0^2 + \epsilon$, the statistic over $B_{K(n)}$ is then at least $nc^2/\{4(M_0^2 + \epsilon)\}$.

With the same per-direction GMS adjustments and bootstrap draws, each bootstrap statistic over $B_{K(n)}$ is bounded by the corresponding statistic over $\mathcal{U}$. Their conditional critical values have the same ordering, and the full-space critical value is $O_p(1)$ by the bound used in the proof of Theorem E.1 of Andrews and Shi (2017). Therefore θ is excluded with probability approaching one, without a rate restriction on $K(n)$. $\square$

E Relation to Alternative Outer Sets

Aucejo, Bugni, and Hotz (2017) are concerned with the same conditional moment model defined by equation (1) and Assumptions 1 above. However, their focus is on the missing

data problem, i.e. $[Y_l, Y_u] = \mathcal{Y}$ and $[X_l, X_u] = \mathcal{X}$. Nevertheless, comparing our results with their construction develops further intuition for the identification problem, thus in this section we discuss both approaches and their relation in detail. For simplicity, let us focus on the one moment case, thus $d_m = 1$, although both their paper and the present work can deal with the more general case. Similarly, let us only cover the missing covariate case and abstract from fully observed additional covariates. Define $Z = 1$ if the covariate is missing and $Z = 0$ otherwise.

As mentioned in the main text, Aucejo, Bugni, and Hotz (2017) note that the sharp identified set, Θ_I^* , is generally a too complicated object to compute and thus they resort to constructing outer sets, i.e. supersets of the sharp identified set. The first step in Aucejo, Bugni, and Hotz (2017) transforms the conditional moment conditions into unconditional moments via instrument functions as we propose for our problem in Lemma 1. They use the class of boxes $\mathcal{G}_{\text{box}} \equiv \{\mathbb{I}[X \in G(x, v)] : \forall (x, v) \in \mathcal{X} \times \mathbb{R}_{++}^{d_x}\}$ with $G(x, v) \equiv \{\tilde{x} \in \mathcal{X} : \{x_j - v_j \leq \tilde{x}_j \leq x_j + v_j\}_{j=1}^{d_x}\}$ as instrument functions.⁹ Let us denote the collection of boxes by $\{G_k\}_{k \geq 0} \equiv \{G(x, v) \mid \forall (x, v) \in \mathcal{X} \times \mathbb{R}_{++}^{d_x}\}$. Note that the main point of the following conceptual comparison does not rely on the specific instrument class. They propose two outer sets.

The first outer set in Aucejo, Bugni, and Hotz (2017) uses the following vector-valued auxiliary functions for each member of $\{G_k\}_{k \geq 0}$:

$$M_{k,1}(W^O) \equiv \left(\begin{array}{c} - \left(\inf_{y \in \mathcal{Y}, x \in G_k} m(y, x, \theta) \mathbb{I}[Z = 1] + m(Y, X, \theta) \mathbb{I}[Z = 0, X \in G_k] \right) \\ + \left(\sup_{y \in \mathcal{Y}, x \in G_k} m(y, x, \theta) \mathbb{I}[Z = 1] + m(Y, X, \theta) \mathbb{I}[Z = 0, X \in G_k] \right) \end{array} \right)$$

They proceed by defining their outer set as $\Theta_{I,1}^{ABH} \equiv \{\theta \in \Theta : \mathbf{E}[M_{k,1}(W^O)] \geq 0 \quad \forall k \geq 0\}$. This solves the moment problem for each box, or in general, each subspace defined by the instrument functions, but not simultaneously for all of them. As Aucejo, Bugni, and Hotz

⁹Note that in the definition of $B(x, v)$ we use $\tilde{x} \in \mathcal{X}$ directly, thus $B(x, v) \cap \mathcal{X} \neq \emptyset$ is always true and we can avoid carrying the intersection in the definition of M_k below.

(2017) mention, the difficult part of characterizing the sharp identified set is to ensure that the moment conditions hold *simultaneously* for all boxes and not only individually, thus Θ_I^{ABH} is an outer set.

The second outer set defined in Aucejo, Bugni, and Hotz (2017) is based on moments integrating out the missing covariate. Their auxiliary moment conditions do not depend on k in the case of no additional fully observed covariates and are given by:

$$M_{k,2}(W^O) \equiv \begin{pmatrix} - \left(\inf_{x \in \mathcal{X}} m(Y, x, \theta) \mathbb{I}[Z = 1] + m(Y, X, \theta) \mathbb{I}[Z = 0] \right) \\ + \left(\sup_{x \in \mathcal{X}} m(Y, x, \theta) \mathbb{I}[Z = 1] + m(Y, X, \theta) \mathbb{I}[Z = 0] \right) \end{pmatrix}$$

The second outer identified set is the collection $\Theta_{I,2}^{ABH} \equiv \{\theta \in \Theta : \mathbf{E}[M_{k,2}(W^O)] \geq 0\}$.

Let us now show the relation between $\Theta_{I,1}^{ABH}$, $\Theta_{I,2}^{ABH}$ and our sharp identified set Θ_I . First note that the class of boxes is comprehensively revealing (see Andrews and Shi (2013)), i.e. it fulfills Assumptions 2 (i)¹⁰ thus our identification main result applies. Further, we know that any subset $\mathcal{U} \subseteq B_1^\infty$ is defining an outer set. In particular, choose $\mathcal{U}_{point} := \{\pm e_j : j \geq 1\}$, where $e_j := \{0, \dots, 1, 0, \dots\}$ is the j th dimensional basis vector. The outer set implied by this choice is

$$\Theta_O(\mathcal{U}_{point}) = \{\theta \in \Theta : \mathbf{E}[h(Q_{\theta,0}, u) \mathbb{I}[Z = 0]] + \mathbf{E}[h(Q_{\theta,1}, u) \mathbb{I}[Z = 1]] \geq 0 \quad \forall u \in \mathcal{U}_{point}\}$$

where we use the law of total expectations and divide our theorem into fully observed and missing covariates and denote with a slight abuse of notation the respective random sets as $Q_{\theta,z}$.

Further, considering the class of boxes, the support function for the missing covariate case only is given by $h(Q_{\theta,z}, u) = \sup_{x \in [X_l, X_u]} \sum_{k \geq 0} u_k m(Y, x, \theta) \mathbb{I}[x \in G_k]$. Any choice of $\tilde{u} \in \mathcal{U}_{point}$ simplifies the support function considerably. In particular, choosing $\tilde{u} = e_k$ yields

¹⁰Condition Assumptions 2 (ii) is not fulfilled by the class of boxes alone. Instead we would need to augment them with positive and decaying weights. For the identification argument here, the choice of such weights is immaterial so we abstract from them.

$h(Q_{\theta,z}, \tilde{u}) = \sup_{x \in [X_l, X_u]} m(Y, x, \theta) \mathbb{I}[x \in G_k]$. For the cases for which we fully observe the covariates, $Z = 0$, this further collapses to $h(Q_{\theta,0}, \tilde{u}) = m(Y, X, \theta) \mathbb{I}[X \in G_k]$ with $X = X_l = X_u$. Thus, the condition summed over missingness patterns for the specific choice e_k is

$$\begin{aligned} & \mathbf{E} \left[\sup_{x \in [X_l, X_u]} m(Y, x, \theta) \mathbb{I}[x \in G_k] \mathbb{I}[Z = 1] \right] + \mathbf{E} [m(Y, X, \theta) \mathbb{I}[X \in G_k] \mathbb{I}[Z = 0]] \geq 0 \\ \iff & \mathbf{E} \left[\sup_{x \in \mathcal{X}} m(Y, x, \theta) \mathbb{I}[x \in G_k] \mathbb{I}[Z = 1] + m(Y, X, \theta) \mathbb{I}[Z = 0, X \in G_k] \right] \geq 0 \end{aligned}$$

Where we used $[X_l, X_u] = \mathcal{X}$ for the missing data case. Simultaneously, using $\sup_{a \in A} f(a) = -\inf_{a \in A} (-f(a))$, and considering $\tilde{u} = -e_k$ we obtain:

$$\mathbf{E} \left[- \left(\inf_{x \in \mathcal{X}} m(Y, x, \theta) \mathbb{I}[x \in G_k] \mathbb{I}[Z = 1] + m(Y, X, \theta) \mathbb{I}[Z = 0, X \in G_k] \right) \right] \geq 0$$

These expressions are almost equivalent to the elements of the vector $M_{k,1}$ defined above, however whereas we take the supremum over $x \in \mathcal{X}$, Aucejo, Bugni, and Hotz (2017) are considering $y \in \mathcal{Y}, x \in G_k$. Additionally, our result includes the indicator $\mathbb{I}[x \in G_k]$ in the objective for the missing observations. Notice that taking the supremum over the entire support of Y even for observed values, is not using the entire information provided from the data.

The second difference is that instead of maximizing over the full support as in Theorem 1 above, Aucejo, Bugni, and Hotz (2017) maximize over the individual boxes. In combination with also varying Y for the missing X observations this still yields an outer set. However, conceptually it is assuming that for each box, the value of the missing covariate must lie within the specific box. This is stronger than required by the conditional moment restriction. The conditional moment model is only implying that the moment condition has to hold over all subspaces, not that it must hold under all subspaces assuming that all realized values are within that space. Instead, our construction maximizes $m(Y, x, \theta) \mathbb{I}[x \in G_k]$ over $x \in \mathcal{X}$,

thus it allows for a specific observation not contributing to G_k .¹¹

Taking all of the above into account, the preceding comparison does not by itself establish a general ranking between $\Theta_{I,1}^{ABH}$ and $\Theta_O(\mathcal{U}_{point})$. The main takeaway is that our identification result also considers combinations of boxes. It is the requirement that the moment conditions hold simultaneously across boxes that makes the sharp identified set difficult to compute. Adding directions to our support-function approach yields tighter outer sets which, under Theorem 1, converge to the sharp identified set. Since the sharp set is contained in $\Theta_{I,1}^{ABH}$ under the same maintained assumptions, the limiting set is contained in $\Theta_{I,1}^{ABH}$, although the inclusion need not be strict.

For $\Theta_{I,2}^{ABH}$, an equivalence follows from a single box. Because $\mathcal{X}$ is compact, we can include a box $G_{k_0} = \mathcal{X}$ in our instrument class. The directions e_{k_0} and $-e_{k_0}$ then yield, after dividing by the positive instrument weight,

$$\begin{aligned} \mathbf{E} \left[\sup_{x \in \mathcal{X}} m(Y, x, \theta) \mathbb{I}[Z = 1] + m(Y, X, \theta) \mathbb{I}[Z = 0] \right] &\geq 0, \\ \mathbf{E} \left[- \inf_{x \in \mathcal{X}} m(Y, x, \theta) \mathbb{I}[Z = 1] - m(Y, X, \theta) \mathbb{I}[Z = 0] \right] &\geq 0. \end{aligned}$$

These are precisely the two inequalities defining $\Theta_{I,2}^{ABH}$ in the absence of additional fully observed covariates. Thus $\Theta_O(\{e_{k_0}, -e_{k_0}\}) = \Theta_{I,2}^{ABH}$ and, in particular, $\Theta_O(\mathcal{U}_{point}) \subseteq \Theta_{I,2}^{ABH}$.

¹¹Another way to see this point is to consider a simpler setting. Suppose $m(y, x, \theta) = x$ with $\mathcal{X} = [0, 1]$, thus the moment condition boils down to $\mathbb{E}[X|X] = 0$. If X is missing for all observations we cannot rule out that $P_x = \delta_0$, a mass point at zero. The latter is the only distribution of X fulfilling the moment condition. For any box in $[0, 1]$ and u vector a feasible solution to our support function approach is zero as we can either choose $x = 0$ or another value with associated $u = 0$. Instead, considering any box $[a, b]$ with $a \neq 0$, the first moment in Aucejo, Bugni, and Hotz (2017) first outer set is negative, thus violated. We thank Federico Bugni for bringing up this example.