

*Generalized Method of Moments  
with Partially Missing Data*

*Grigory Franguridi, Hyungsik Roger Moon*

*Paper No: 2026-003*

**CESR-SCHAEFFER  
WORKING PAPER SERIES**

*The Working Papers in this series have not undergone peer review or been edited by USC. The series is intended to make results of CESR and Schaeffer Center research widely available, in preliminary form, to encourage discussion and input from the research community before publication in a formal, peer-reviewed journal. CESR-Schaeffer working papers can be cited without permission of the author so long as the source is clearly referred to as a CESR-Schaeffer working paper.*

# Generalized Method of Moments with Partially Missing Data\*

GRIGORY FRANGURIDI<sup>1</sup> and HYUNGSIK ROGER MOON<sup>2</sup>

<sup>1</sup>Center for Economic and Social Research, University of Southern California

<sup>2</sup>Department of Economics, University of Southern California

January 7, 2026

## Abstract

We consider a generalized method of moments framework in which a part of the data vector is missing for some units in a completely unrestricted, potentially endogenous way. In this setup, the parameters of interest are usually only partially identified. We characterize the identified set for such parameters using the support function of the convex set of moment predictions consistent with the data. This identified set is sharp, valid for both continuous and discrete data, and straightforward to estimate. We also propose a statistic for testing hypotheses and constructing confidence regions for the true parameter, show that standard nonparametric bootstrap may not be valid, and suggest a fix using the bootstrap for directionally differentiable functionals of [Fang and Santos \(2019\)](#). A set of Monte Carlo simulations demonstrates that both our estimator and the confidence region perform well when samples are moderately large and the data have bounded supports.

**Keywords:** endogenous attrition, partial identification, panel data, support function

**JEL codes:** C14, C23

---

\*We appreciate helpful comments and suggestions from Andres Santos. Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health and in part by the Social Security Administration under Award Number U01AG077280. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

# 1 Introduction

Missing data are ubiquitous in empirical research, particularly in repeated-measurement settings. [Rubin \(1976\)](#) classified mechanisms of missingness into three categories: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). In this paper, we consider the last case, where missingness may be correlated with the data in an arbitrary manner. More specifically, we consider a generalized method of moments framework in which a portion of the data vector is missing for some units in a completely unrestricted, potentially endogenous way. A representative example is a two-period panel where, in the second period, some units drop out of the sample in a way that is correlated with both outcomes and covariates in both periods.

When no restriction on missingness is imposed, the parameters of interest are usually only partially identified. We characterize the identified set for the parameter of interest via the support function of the set of moments consistent with the observed data. We show that the identified set is sharp, is valid for both continuous and discrete data, and is straightforward to estimate.

We also show how to perform hypothesis testing and construct confidence regions using the estimate of the minimum of the support function as a test statistic. We derive the limit distribution of this test statistic and provide a procedure for computing critical values, following [Fang and Santos \(2019\)](#). We also show that the test controls size locally uniformly using an abstract result in [Fang and Santos \(2019\)](#).

Partial identification with missing data has been widely studied since Manski's seminal work, [Manski \(1989\)](#) (see, e.g., [Manski \(2005\)](#) and [Molinari \(2020\)](#) for a survey). Our paper contributes to this extensive literature, with its key feature being that the identified set of moment predictions is convex, and the support function of this set can be estimated using a simple sample analog.

Our model is neither a moment inequality model with a convex identified set (e.g., [Kaido and Santos \(2014\)](#)), nor the intersection bounds model of [Chernozhukov, Lee, and Rosen \(2013\)](#). It is similar to the model with convex moment predictions as in [Beresteanu, Molchanov, and Molinari \(2011\)](#), the main difference being that our characterization of the sharp identified set does not rely on a representation via random closed sets.

As mentioned above, the primary application of our model is a panel regression with attrition that may be endogenous. The analysis of panel data with attrition has a long history, and numerous techniques for handling attrition under various assumptions have been proposed. These include modelling the attrition process parametrically (e.g., [Hausman and Wise \(1979\)](#)), using auxiliary information in the form of refreshment samples (e.g., [Hirano, Imbens, Ridder, and Rubin \(2001\)](#)), imposing semiparametric assumptions on the attrition process (e.g., [Bhattacharya \(2008\)](#)), using inverse probability weighting (e.g., [Wooldridge](#)

(2002)), selection models with unobserved heterogeneity (e.g., Semykina and Wooldridge (2010)), and multiple imputation and Bayesian methods (e.g., Deng, Chang et al. (2016)).

The rest of the paper is organized as follows. Section 2 introduces the general framework and characterizes the sharp identified set. Section 3 suggests an estimator of the identified set. Section 4 develops a bootstrap procedure for hypothesis testing. Section 6 explores the performance of our methodology in a set of Monte Carlo simulations. Section 7 concludes. All proofs are given in the Appendix.

## 2 Setup and the identified set

We consider the generalized method of moments framework when some of the observed variables may be missing for some observations. Specifically, let  $Z_{1i}$  be a data vector that we observe for each unit  $i = 1, \dots, n$  of the sample and let  $Z_{2i}$  be a data vector that may be missing for some units. Let  $S_i \in \{0, 1\}$  be the sample selection indicator, which is equal to 1 when  $Z_{2i}$  is observed and 0 otherwise. Dropping the subscript  $i$ , the observed vector is then  $W = (S, Z_1', SZ_2')'$ . A key aspect of our framework is that we allow arbitrary dependence of  $S$  on both observed and unobserved features. In particular,  $S$  can be arbitrarily correlated with  $Z_2$ , representing endogenous sample selection. We do not allow for overlap between elements of  $Z_1$  and  $Z_2$ . More generally, no features of the distribution of  $Z_2$  are assumed to be identified from the random sample of  $Z_1$ . We are interested in a finite-dimensional parameter  $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$  that is defined by general moment conditions

$$\mathbb{E}_\pi[\phi(Z_1, Z_2, \theta)] = 0, \tag{1}$$

where  $\phi$  is a  $d_\phi$ -dimensional moment function and  $\pi$  is the joint distribution of  $Z_1, Z_2$ . Of course,  $\pi$  is not point-identified due to the presence of missing data (see below), and hence neither is  $\theta$ .

One special case of this framework is two-period panel data with unrestricted attrition in the second period. To see this, let  $Z_t = (X_t', Y_t) \in \mathbb{R}^d$  be the stacked vector of covariates and outcomes in period  $t = 1, 2$ . In period 1, a random sample from  $Z_1$  is observed. In period 2, there is attrition, and so  $Z_2$  is only observed when  $S = 1$ , where  $S \in \{0, 1\}$  is the indicator of staying in the sample. When the attrition is fully unrestricted and the parameter of interest is defined by a set of moment conditions (1) (such as, for instance, the slope coefficient in the fixed effects linear regression), this model becomes a special case of the GMM with missing data described above. For expositional clarity, we use the terminology related to this panel model throughout the rest of the paper.

We impose the following assumptions.

**Assumption 1.** (i) *The parameter space  $\Theta$  is compact.*

(ii) The distribution  $\pi$  of  $(Z_1, Z_2)$  belongs to the set  $\mathcal{P}(\mathcal{Z}_1 \times \mathcal{Z}_2)$  of probability distributions on  $\mathcal{Z}_1 \times \mathcal{Z}_2$ , where  $\mathcal{Z}_1, \mathcal{Z}_2 \subset \mathbb{R}^d$  are known compact sets.

(iii) For each  $\theta \in \Theta$ , the map  $(z_1, z_2) \mapsto \phi(z_1, z_2, \theta)$  is continuous.

(iv) For each  $\pi \in \mathcal{P}(\mathcal{Z}_1 \times \mathcal{Z}_2)$ , the map  $\theta \mapsto \mathbb{E}_\pi[\phi(Z_1, Z_2, \theta)]$  is continuous.

Assumptions 1(i), 1(iii) and 1(iv) are standard and impose compactness of the parameter space and the continuity of moments. Assumption 1(ii) posits that the researcher knows the compact set  $\mathcal{Z}_1 \times \mathcal{Z}_2$  that contains (but does not have to be equal to) the true support of the data. We impose no restrictions on the data-generating process beyond these basic requirements.

To characterize the identified set for  $\theta$ , let  $p = P(S = 1) \in (0, 1)$  be the unconditional selection probability and notice that  $\pi$  can be decomposed into the point-identified distribution of  $(Z_1, Z_2)|S = 1$ , denoted by  $\pi^1$ , and the partially identified distribution of  $(Z_1, Z_2)|S = 0$ , denoted by  $\pi^0$ , viz.,

$$\pi = p\pi^1 + (1 - p)\pi^0.$$

Denote by  $\pi_1^0$  the point-identified probability distribution of  $Z_1|S = 0$ . The sharp identified set for  $\pi^0$  is the convex set of the distributions satisfying Assumption 1(ii) with first marginal  $\pi_1^0$ , i.e.,

$$\Pi^0 = \{ \pi^0 \in \mathcal{P}(\mathcal{Z}_1 \times \mathcal{Z}_2) : \pi^0(A \times \mathcal{Z}_2) = \pi_1^0(A) \text{ for all Borel sets } A \subset \mathcal{Z}_1 \}.$$

Therefore, the sharp identified set for  $\pi$  is

$$\Pi = \{ p\pi^1 + (1 - p)\pi^0, \pi^0 \in \Pi^0 \}.$$

Write

$$\mathbb{E}_\pi[\phi(Z_1, Z_2, \theta)] = p \mathbb{E}_{\pi^1}[\phi(Z_1, Z_2, \theta)] + (1 - p) \mathbb{E}_{\pi^0}[\phi(Z_1, Z_2, \theta)] =: \nu_{\pi^0}(\theta). \quad (2)$$

Then the sharp identified set for  $\theta$  is

$$\Theta_I = \{ \theta \in \Theta : \exists \pi^0 \in \Pi^0 \text{ such that } \nu_{\pi^0}(\theta) = 0 \}.$$

The next proposition establishes a convenient characterization of  $\Theta_I$  in terms of the support

function of the compact, convex set<sup>1</sup> of moment predictions

$$N(\theta) := \{\nu_{\pi^0}(\theta) : \pi^0 \in \Pi^0\}.$$

**Proposition 1** (identified set via support function). *Suppose Assumption 1 holds and let*

$$\psi_{N(\theta)}(u) = \max_{\nu \in N(\theta)} u' \nu$$

be the support function of  $N(\theta)$ . Define the criterion function

$$Q(\theta) = \min_{u \in \mathbb{B}} \psi_{N(\theta)}(u),$$

where  $\mathbb{B}$  is the unit ball in  $\mathbb{R}^{d_\phi}$ . Then  $\Theta_I$  is compact and given by

$$\Theta_I = \{\theta \in \Theta : Q(\theta) = 0\}. \quad (3)$$

*Proof.* See Appendix A. □

**Remark 2.1.** *In the above characterization, the Euclidean unit ball  $\mathbb{B}$  can be replaced by any compact set containing the origin in its interior, such as the unit ball in the  $\ell^1$  norm or the uniform norm. This may turn out to be convenient computationally.*

**Remark 2.2.** *A similar strategy was recently employed by Franguridi and Liu (2025) for characterizing the identified set in a partially identified GMM framework based on optimal transport.*

The support function  $\psi_{N(\theta)}$  can be written as

$$\begin{aligned} \psi_{N(\theta)}(u) &= \max_{\pi \in \Pi} u' \mathbb{E}_\pi[\phi(Z_1, Z_2, \theta)] \\ &= p \mathbb{E}[u' \phi(Z_1, Z_2, \theta) | S = 1] + (1 - p) \max_{\pi^0 \in \Pi^0} \mathbb{E}_{\pi^0}[u' \phi(Z_1, Z_2, \theta)]. \end{aligned}$$

The last term involves optimization over an infinite-dimensional set of distributions. Fortunately, we can convert it into a finite-dimensional program for every value of  $Z_1$ , as the following proposition shows.

**Proposition 2** (reduction formula). *Suppose Assumption 1 holds. Then*

$$\max_{\pi^0 \in \Pi^0} \mathbb{E}_{\pi^0}[u' \phi(Z_1, Z_2, \theta)] = \mathbb{E} \left[ \max_{z_2 \in \mathcal{Z}_2} u' \phi(Z_1, z_2, \theta) | S = 0 \right].$$

*Proof.* See Appendix B. □

---

<sup>1</sup>See Step 1 in Appendix A for the proof of convexity and compactness of  $N(\theta)$ .

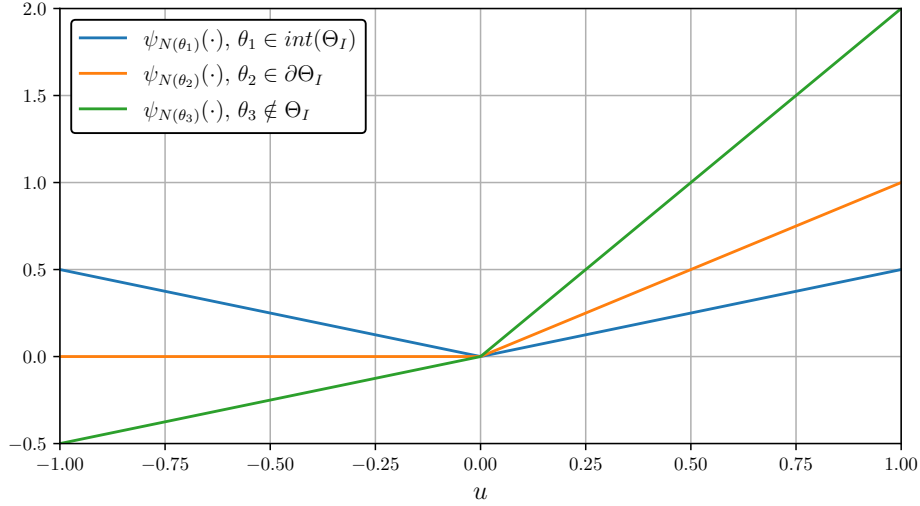


Figure 1: Set identification via support functions

**Example 1.** Suppose there is only one moment condition, i.e.,  $\phi$  is scalar-valued. In this case, the set of moment predictions is the closed interval

$$N(\theta) = \left[ \min_{\pi \in \Pi} \mathbb{E}_{\pi}[\phi(Z_1, Z_2, \theta)], \max_{\pi \in \Pi} \mathbb{E}_{\pi}[\phi(Z_1, Z_2, \theta)] \right],$$

The support function of  $N(\theta)$  is a piecewise linear function given by

$$\psi_{N(\theta)}(u) = \max_{\pi \in \Pi} u \mathbb{E}_{\pi}[\phi(Z_1, Z_2, \theta)] = \begin{cases} u \cdot \min_{\pi \in \Pi} \mathbb{E}_{\pi}[\phi(Z_1, Z_2, \theta)], & \text{if } u < 0, \\ u \cdot \max_{\pi \in \Pi} \mathbb{E}_{\pi}[\phi(Z_1, Z_2, \theta)], & \text{if } u \geq 0, \end{cases}$$

and hence the criterion function is

$$Q(\theta) = \min_{u \in \mathbb{B}} \psi_{N(\theta)}(u) = \min \left\{ 0, \max_{\pi \in \Pi} \mathbb{E}_{\pi}[\phi(Z_1, Z_2, \theta)], -\min_{\pi \in \Pi} \mathbb{E}_{\pi}[\phi(Z_1, Z_2, \theta)] \right\}.$$

The identification condition  $Q(\theta) = 0$  can be easily seen to be equivalent to

$$\min_{\pi \in \Pi} \mathbb{E}_{\pi}[\phi(Z_1, Z_2, \theta)] \leq 0 \leq \max_{\pi \in \Pi} \mathbb{E}_{\pi}[\phi(Z_1, Z_2, \theta)],$$

which, of course, is just the condition that zero belongs to the set of moment predictions. See Figure 1 for an illustration of how the argmax set of  $\psi_{N(\theta)}$  determines the position of  $\theta$  relative to the identified set  $\Theta_I$ . Notice also that using the reduction formula in Proposition 2,

we can write

$$\begin{aligned}\min_{\pi \in \Pi} \mathbb{E}_\pi[\phi(Z_1, Z_2, \theta)] &= \mathbb{E}[f_{\min}(W, \theta)], \\ \max_{\pi \in \Pi} \mathbb{E}_\pi[\phi(Z_1, Z_2, \theta)] &= \mathbb{E}[f_{\max}(W, \theta)],\end{aligned}$$

where  $W = (S, Z_1', SZ_2')'$  is the data vector and

$$\begin{aligned}f_{\min}(W, \theta) &= S\phi(Z_1, SZ_2, \theta) + (1 - S) \min_{z_2 \in \mathcal{Z}_2} \phi(Z_1, z_2, \theta), \\ f_{\max}(W, \theta) &= S\phi(Z_1, SZ_2, \theta) + (1 - S) \max_{z_2 \in \mathcal{Z}_2} \phi(Z_1, z_2, \theta).\end{aligned}$$

### 3 Estimation

Combining Propositions 1 and 2 suggests the following estimator of the identified set:

$$\hat{\Theta}_I = \left\{ \theta \in \Theta : \hat{Q}(\theta) \geq -\eta_n \right\},$$

where  $\eta_n > 0$  is a tuning parameter and the sample criterion function

$$\hat{Q}(\theta) = \min_{u \in \mathbb{B}} \hat{\psi}_{N(\theta)}(u) = \frac{1}{n} \sum_{i=1}^n \left( s_i u' \phi(z_{i1}, z_{i2}, \theta) + (1 - s_i) \max_{z_2 \in \mathcal{Z}_2} u' \phi(z_{i1}, z_2, \theta) \right).$$

Define the distance from a point  $x \in \mathbb{R}^{d_\phi}$  to a set  $A \subset \mathbb{R}^{d_\phi}$  by

$$d(x, A) = \inf_{a \in A} \|a - x\|.$$

For any sets  $A, B \subset \mathbb{R}^{d_\theta}$ , define the Hausdorff distance

$$d_H(A, B) = \max \left\{ \sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A) \right\}.$$

We impose the following assumptions.

**Assumption 2.** *There exists a weakly increasing function  $m : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that  $m(0) = 0$ ,  $m(\delta) > 0$  for  $\delta > 0$  and  $\min_{u \in \mathbb{B}} \psi_{N(\theta)}(u) \leq -m(d(\theta, \Theta_I))$ .*

**Assumption 3.**  *$\eta_n \downarrow 0$  and  $n^{-1/2} = o(\eta_n)$ .*

Assumption 2 imposes separation of the identified set  $\Theta_I$  in terms of the criterion function  $Q(\theta) = \min_{u \in \mathbb{B}} \psi_{N(\theta)}(u)$ . A sufficient condition can be formulated in terms of the gradient of  $Q(\theta)$  away from the identified set. Assumption 3 states that  $\eta_n$  has to converge to zero slower than  $O(n^{-1/2})$ .

The following proposition establishes the consistency of our estimator in the Hausdorff distance.

**Theorem 1** (Hausdorff consistency). *Under Assumptions 1 to 3,  $d_H(\Theta_I, \hat{\Theta}_I) = o_p(1)$ .*

*Proof.* See Appendix C. □

## 4 Inference

We now wish to develop a test of the hypothesis  $H_0 : \theta = \theta_0$ . We suggest using the statistic

$$\hat{T}(\theta_0) = \sqrt{n} \min_{u \in \mathbb{B}} \hat{\psi}_{N(\theta_0)}(u),$$

which is the estimator of the scaled negative distance to the identified set of moment predictions  $N(\theta_0)$ . Note that more negative values of  $\hat{T}(\theta_0)$  indicate stronger evidence against the null. To establish the asymptotic distribution of this statistic, we first prove that the function  $\hat{\psi}_{N(\theta)}(u)$  converges to a Gaussian process uniformly over  $u \in \mathbb{B}$  and  $\theta \in \Theta$ .

**Assumption 4.**  $\mathbb{E} [\sup_{\theta \in \Theta} \sup_{z_2 \in \mathcal{Z}_2} \|\phi(Z_1, z_2, \theta)\|^2] < \infty$ .

**Assumption 5.** *There exists a random variable  $L(Z_1)$  such that  $\mathbb{E} L(Z_1)^2 < \infty$  and*

$$\sup_{z_2 \in \mathcal{Z}_2} \|\phi(Z_1, z_2, \theta_1) - \phi(Z_1, z_2, \theta_2)\| \leq L(Z_1) \|\theta_1 - \theta_2\|$$

for all  $\theta_1, \theta_2 \in \Theta$ , almost surely in  $Z_1$ .

**Theorem 2** (uniform CLT for the support function). *Suppose Assumptions 1, 4 and 5 hold. Then there exists a Gaussian process  $\mathbb{G}$  with trajectories in  $C(\mathbb{B} \times \Theta)$  such that*

$$\sqrt{n}(\hat{\psi}_{N(\theta)}(u) - \psi_{N(\theta)}(u)) \rightsquigarrow \mathbb{G}(u, \theta).$$

*Proof.* See Appendix D. □

**Remark 4.1.** *The results of [Kaido and Santos \(2014\)](#) suggest that the estimator  $\hat{\psi}_{N(\theta)}(u)$  is asymptotically semiparametrically efficient for  $\psi_{N(\theta)}(u)$ .*

An application of the functional delta method to Theorem 2 yields the following result for the asymptotic distribution of the test statistic  $\hat{T}(\theta)$  over  $\theta \in \Theta_I$ .

**Corollary 1** (asymptotic distribution of test statistic). *Under the assumptions of Theorem 2,*

$$\hat{T}(\theta) \rightsquigarrow \min_{u \in U(\theta)} \mathbb{G}(u, \theta)$$

uniformly over  $\theta \in \Theta_I$ , where

$$U(\theta) = \underset{u \in \mathbb{B}}{\operatorname{argmin}} \psi_{N(\theta)}(u).$$

*Proof.* See Appendix E. □

**Remark 4.2.** Using this uniform convergence result, it should be possible to develop a specification test based on the statistic  $\max_{\theta \in \Theta} \hat{T}(\theta)$ , the population analog of which is zero if and only if the identified set  $\Theta_I$  is nonempty. The critical values can be obtained using the bootstrap of Fang and Santos (2019), in analogy to our treatment of the test statistic  $\hat{T}(\theta_0)$  for a fixed  $\theta_0$  below. Similar specification tests have been developed for moment inequality models, see, e.g., Bugni, Canay, and Shi (2015) and references therein.

Since the asymptotic distribution of  $\hat{T}(\theta)$  is neither available in closed form nor is easy to simulate from, we need an alternative procedure for obtaining the critical values. The results of Fang and Santos (2019) imply that, unless the Hadamard derivative  $\min_{u \in U(\theta)} \mathbb{G}(u, \theta)$  is a linear functional of  $\mathbb{G}$ , the standard nonparametric bootstrap is invalid.

**Example 1** (continued). In the univariate case, the sample criterion function is

$$\hat{Q}(\theta) = \min_{u \in \mathbb{B}} \hat{\psi}_{N(\theta)}(u) = \min \left\{ 0, \hat{\mathbb{E}}[f_{\max}(W, \theta)], -\hat{\mathbb{E}}[f_{\min}(W, \theta)] \right\}.$$

Consider a data generating process and a parameter value  $\theta$  satisfying  $\mathbb{E}[f_{\min}(W, \theta)] < 0$ , making the sample moment  $\hat{\mathbb{E}}[f_{\min}(W, \theta)]$  irrelevant asymptotically. We have

$$\begin{aligned} Q(\theta) &= \min \{ 0, \mathbb{E}[f_{\max}(W, \theta)] \}, \\ \hat{Q}(\theta) &\overset{asy}{\approx} \min \left\{ 0, \hat{\mathbb{E}}[f_{\max}(W, \theta)] \right\}. \end{aligned}$$

This is known to be an irregular problem when  $\mathbb{E}[f_{\max}(W, \theta)] = 0$ , which corresponds to  $\theta$  being on the boundary of the identified set. In particular, the standard nonparametric bootstrap is invalid for the asymptotic distribution  $\min(0, N(0, 1))$  of  $\sqrt{n}\hat{Q}(\theta)$ , see, e.g., Andrews (2000).

We show how to leverage the bootstrap for directionally differentiable functionals of Fang and Santos (2019). To this end, we propose an estimator of the derivative that satisfies their key Assumption 4.

For  $\varepsilon_n > 0$ , define

$$\hat{\chi}'(h) = \min_{u \in \hat{U}(\varepsilon_n)} h(u), \tag{4}$$

where

$$\hat{U}(\varepsilon_n) = \left\{ u \in \mathbb{B} : \hat{\psi}_{N(\theta_0)}(u) \leq \min_{v \in \mathbb{B}} \hat{\psi}_{N(\theta_0)}(v) + \varepsilon_n \right\}$$

is the “ $\varepsilon_n$ -enlarged” argmin set of  $\hat{\psi}_{N(\theta_0)}$ . We impose the following assumptions.

**Assumption 6** (sharp minima of  $\psi_{N(\theta_0)}$ ). *There exists  $\kappa > 0$  such that*

$$\psi_{N(\theta_0)}(u) \geq \min_{v \in \mathbb{B}} \psi_{N(\theta_0)}(v) + \kappa \cdot d(u, U(\theta_0)) \text{ for all } u \in \mathbb{B}.$$

**Assumption 7** (bandwidth). *The sequence  $\varepsilon_n$  satisfies  $\varepsilon_n \downarrow 0$  and  $\|\hat{\psi}_{N(\theta_0)} - \psi_{N(\theta_0)}\|_{\mathbb{B}} = o_p(\varepsilon_n)$ .*

Assumption 6 posits that  $\psi_{N(\theta_0)}$  grows sufficiently fast away from its argmin set  $U(\theta_0)$ . The latter does not have to be a singleton, as illustrated in Example 1 for which Assumption 6 holds. This assumption is equivalent to every subgradient  $\nabla \psi_{N(\theta_0)}(u)$  being bounded away from zero for  $u \notin U(\theta_0)$ . It also suffices that it holds in a small neighborhood around  $U(\theta_0)$  rather than on the entire set  $\mathbb{B}$ . Assumption 7 restricts how fast  $\varepsilon_n$  should converge to zero. Since  $\|\hat{\psi}_{N(\theta_0)} - \psi_{N(\theta_0)}\|_{\mathbb{B}} = O_p(1/\sqrt{n})$  by Theorem 2, we can take  $\varepsilon_n \sim \log n/\sqrt{n}$ .

Given our estimator (4), the Fang and Santos (2019) bootstrap algorithm for testing  $H_0 : \theta = \theta_0$  at the nominal size  $\alpha \in (0, 1)$  is as follows.

1. Compute  $\hat{T}(\theta_0) = \sqrt{n} \min_{u \in \mathbb{B}} \hat{\psi}_{N(\theta_0)}(u)$ .
2. Define  $\hat{U}_n = \left\{ u \in \mathbb{B} : \hat{\psi}_{N(\theta_0)}(u) \leq \min_{v \in \mathbb{B}} \hat{\psi}_{N(\theta_0)}(v) + \varepsilon_n \right\}$ .
3. For each  $b = 1, \dots, B$ :
  - (a) Draw a random sample  $W_i^b$ ,  $i = 1, \dots, n$ , with replacement from the data  $W_i = (S_i, Z_{1i}, S_i Z_{2i})$ ,  $i = 1, \dots, n$ .
  - (b) Compute  $\hat{\psi}_{N(\theta_0)}^{*b}(u)$  on the sample  $W_i^b$ ,  $i = 1, \dots, n$ .
  - (c) Compute

$$\hat{T}^{*b} = \min_{u \in \hat{U}_n} \sqrt{n} \left( \hat{\psi}_{N(\theta_0)}^{*b}(u) - \hat{\psi}_{N(\theta_0)}(u) \right).$$

4. Denote by  $\hat{c}_\alpha^*$  the  $\alpha$ -quantile of  $\hat{T}^{*1}, \dots, \hat{T}^{*B}$ .
5. Reject  $H_0$  if  $\hat{T}(\theta_0) < \hat{c}_\alpha^*$ .

The following theorem establishes that this bootstrap-based testing procedure controls size for a fixed data generating process.

**Theorem 3** (test validity, pointwise). *Suppose Assumptions 6 and 7 hold and consider the testing procedure above with  $B = \infty$ . Then for every fixed data generating process, the asymptotic probability of rejecting the true null hypothesis is less than or equal to  $\alpha$ .*

*Proof.* See Appendix F. □

Although the previous theorem establishes that our test asymptotically controls size for a fixed data generating process, it is silent about uniform size control. In other words, it could be possible that even for a large sample size, one can find a particularly unfavorable DGP that would lead our test to exhibit significant size distortion. Such lack of uniform size control has been extensively studied in the context of moment inequality models, e.g., [Imbens and Manski \(2004\)](#); [Romano and Shaikh \(2008\)](#); [Andrews and Guggenberger \(2009\)](#); [Moon and Schorfheide \(2009\)](#); [Woutersen \(2006\)](#).

Fortunately, we can rely on the abstract results of [Fang and Santos \(2019\)](#) to establish local size control. Denote by  $P$  the joint distribution of the data vector  $W = (S, Z'_1, SZ'_2)'$  and let  $\psi(P)$  be the corresponding population support function of  $N(\theta_0)$ , where  $\theta_0$  is any value in the identified set  $\Theta_I = \Theta_I(P)$ . We consider a sequence of DGPs  $P_t$ ,  $t \geq 0$ , that is local to a given DGP  $P_0$  in the following sense, cf. Assumption 5 in [Fang and Santos \(2019\)](#).

**Assumption 8** (local sequence of DGPs). *The path  $t \mapsto P_t$  is quadratic mean differentiable and satisfies for any  $\lambda > 0$  the following.*

(i) *There exists  $\psi'(\lambda)$  such that*

$$\|\sqrt{n}(\psi(P_{\lambda/\sqrt{n}}) - \psi(P_0)) - \psi'(\lambda)\|_{\mathbb{B}} \rightarrow 0.$$

(ii)  $\sqrt{n}(\hat{\psi}_n - \psi(P_{\lambda/\sqrt{n}})) \xrightarrow{\lambda} \mathbb{G}_0$ , where  $\xrightarrow{\lambda}$  denotes convergence in distribution under  $\{W_i\}_{i=1}^n$  i.i.d., with each  $W_i$  distributed as  $P_{\lambda/\sqrt{n}}$ , and  $\hat{\psi}_n$  is our estimator of  $\psi(P_{\lambda/\sqrt{n}})$  based on  $\{W_i\}_{i=1}^n$ .

(iii)  $\mathbb{G}_0$  is tight and supported on  $C(\mathbb{B})$ .

Assumption 8(i) imposes differentiability of the parameter  $\psi(P)$  along the path  $P_{\lambda/\sqrt{n}}$ . It is expected to hold for any quadratic mean differentiable path in our model under mild regularity conditions. Assumption 8(ii) requires that the distribution of the estimator  $\hat{\psi}_n$  is unaffected by local perturbations of the DGP, i.e. that  $\hat{\psi}_n$  is *regular*. Finally, Assumption 8(iii) is expected to hold due to [Theorem 2](#).

**Theorem 4** (test validity, locally uniform). *Suppose Assumptions 6 to 8 hold and consider the testing procedure above with  $B = \infty$ . Then this procedure controls size locally uniformly*

in the sense that

$$\limsup_{n \rightarrow \infty} P_{\lambda/\sqrt{n}} \left( \hat{T}_n(\theta_0) < \hat{c}_\alpha^* \right) \leq \alpha.$$

*Proof.* See Appendix H. □

## 5 Extension to more general missingness patterns

Our methodology can be extended to more general missingness patterns, as long as missingness is “monotone” (see below) and data contains no information on the distribution of the missing variables (i.e., missingness is “fully unrestricted”). We will illustrate this extension using the setup of panel data with more than two periods, where attrition happens monotonically, i.e., once the units drop out of the sample, they never return. For expositional simplicity, we consider the case of  $T = 3$  periods.

Denote the data in the three periods by  $Z_1, Z_2, Z_3$  and denote the indicators of staying in the sample in periods 2 and 3 by  $S_2$  and  $S_3$ , respectively. Since there is no return to the sample,  $S_2 = 0$  implies  $S_3 = 0$ . Also, let

$$p_2 = P(S_2 = 1), \quad p_3 = P(S_3 = 1), \quad p_{2|0} = p(S_2 = 1 | S_3 = 0).$$

The parameter of interest  $\theta$  is defined by the moment conditions

$$\mathbb{E}_\pi[\phi(Z_1, Z_2, Z_3, \theta)] = 0,$$

where the expectation is taken with respect to the latent, partially identified distribution  $\pi$  of  $(Z_1, Z_2, Z_3)$ .

**Proposition 3.** *The identified set is  $\Theta_I = \{\theta \in \Theta : \min_{u \in \mathbb{B}} \psi_\theta(u) = 0\}$ , where*

$$\begin{aligned} \psi_\theta(u) = & p_3 \mathbb{E}[\phi(Z_1, Z_2, Z_3, \theta) | S_3 = 1] + (1 - p_3) \left[ p_{2|0} \mathbb{E}[\max_{z_3} u' \phi(Z_1, Z_2, z_3, \theta) | S_2 = 1, S_3 = 0] \right. \\ & \left. + (1 - p_{2|0}) \mathbb{E}[\max_{z_2, z_3} u' \phi(Z_1, z_2, z_3, \theta) | S_2 = S_3 = 0] \right]. \end{aligned}$$

*Proof.* See Appendix G. □

## 6 Monte Carlo simulation

We consider a simple cross-sectional linear regression

$$Y_i = \theta_0 + \theta_1 X_i + \varepsilon_i, \quad i = 1, \dots, n,$$

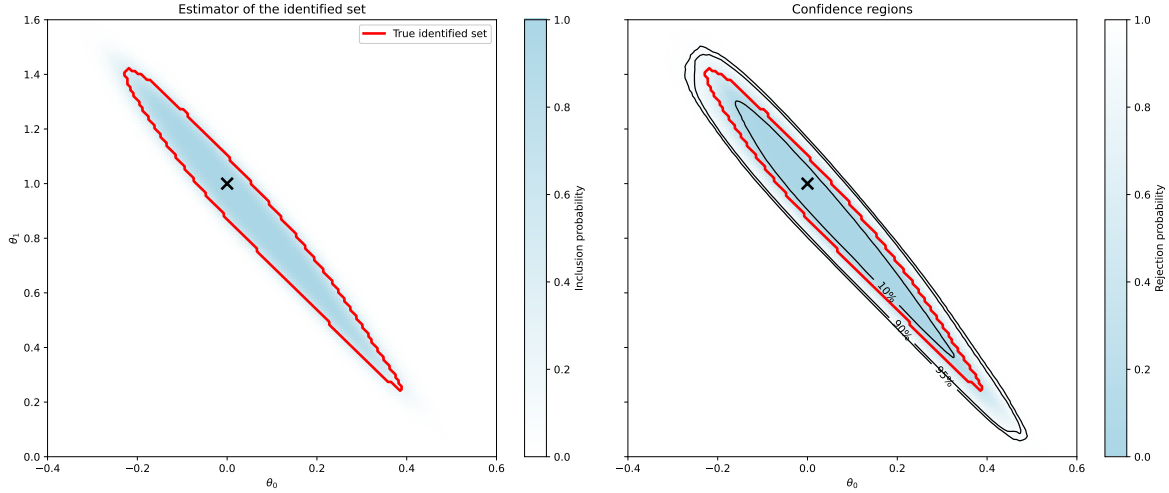


Figure 2: Estimates of the identified set and confidence regions with nominal sizes  $1 - \alpha \in \{0.10, 0.90, 0.95\}$  averaged over 1000 simulations.

where  $Y_i$  is observed for the entire sample, whereas  $X_i$  is missing for observations  $i$  with  $S_i = 0$ , where  $S_i \in \{0, 1\}$  is the sample selection indicator. The parameter of interest is  $\theta = (\theta_0, \theta_1)$ . We assume that the support of  $X_i$  is known, in agreement with Assumption 1(ii). Dropping the subscript  $i$ , the moment conditions are

$$\begin{aligned}\mathbb{E}(Y - \theta_0 - \theta_1 X) &= 0, \\ \mathbb{E}[X(Y - \theta_0 - \theta_1 X)] &= 0,\end{aligned}$$

and hence the support function is

$$\begin{aligned}\psi_\theta(u_1, u_2) &= p \mathbb{E}[u_1(Y - \theta_0 - \theta_1 X) + u_2 X(Y - \theta_0 - \theta_1 X) | S = 1] \\ &\quad + (1 - p) \mathbb{E} \left[ \max_{x \in [0, 1]} \{u_1(Y - \theta_0 - \theta_1 x) + u_2 x(Y - \theta_0 - \theta_1 x)\} \right].\end{aligned}$$

Notice that the inner maximization is a constrained quadratic problem. We assume that  $X_i$  has a uniform $[0, 1]$  distribution and  $\varepsilon_i$  has a uniform $[-1, 1]$  distribution independently of  $X_i$ . Selection is completely random with probability  $p = P(S_i = 1) = 0.9$ . We set the sample size  $n = 1000$  and the tuning parameters  $\eta_n = \varepsilon_n = 0.1 \log n / \sqrt{n} \approx 0.022$ .

The simulation results are illustrated in Figure 2. The true identified set (red contour line) is approximated using a sample of 10,000 observations. Our estimates of the identified set track the true identified set closely. The confidence regions (three black contour lines) collect the hypothesized parameter values that are not rejected by the bootstrap test described in Section 4 with the size  $\alpha \in \{0.05, 0.10, 0.90\}$  and the number of bootstrap samples  $B = 1000$ . They are found to be quite narrow even at high nominal confidence levels  $1 - \alpha$ .

## 7 Conclusion

In this paper, we study the partial identification of parameters defined by GMM moment conditions when components of the data vector are missing in a completely unrestricted manner. A leading motivation is panel data subject to endogenous attrition.

We characterize the sharp identified set through the support function of the convex set of moment predictions consistent with the observed data. For inference, we employ the minimum of the sample analog of the support function as a test statistic and develop a valid inference procedure, drawing on the bootstrap for directionally differentiable functionals of Fang and Santos (2019). We demonstrate that our estimator and confidence regions perform well in a set of Monte Carlo simulations.

## References

- ALIPRANTIS, C. D. AND K. C. BORDER (2006): *Infinite dimensional analysis: a hitchhiker's guide*, Springer.
- ANDREWS, D. W. (2000): "Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space," *Econometrica*, 399–405.
- ANDREWS, D. W. AND P. GUGGENBERGER (2009): "Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities," *Econometric Theory*, 25, 669–709.
- BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2011): "Sharp identification regions in models with convex moment predictions," *Econometrica*, 79, 1785–1821.
- BHATTACHARYA, D. (2008): "Inference in panel data models under attrition caused by unobservables," *Journal of Econometrics*, 144, 430–446.
- BUGNI, F. A., I. A. CANAY, AND X. SHI (2015): "Specification tests for partially identified models defined by moment inequalities," *Journal of Econometrics*, 185, 259–282.
- CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): "Intersection bounds: Estimation and inference," *Econometrica*, 81, 667–737.
- DENG, Y., C. CHANG, ET AL. (2016): "Statistical Methods for Nonignorable Missing Longitudinal Data," *Statistical Science*.
- DUDLEY, R. M. (2018): *Real analysis and probability*, Chapman and Hall/CRC.
- FANG, Z. AND A. SANTOS (2019): "Inference on directionally differentiable functions," *The Review of Economic Studies*, 86, 377–412.
- FRANGURIDI, G. AND L. LIU (2025): "Inference in partially identified moment models via regularized optimal transport," *Working paper*.

- HAUSMAN, J. A. AND D. WISE (1979): “Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment,” *Econometrica*, 47, 455–474.
- HIRANO, K., G. IMBENS, G. RIDDER, AND D. B. RUBIN (2001): “Combining panel data sets with attrition and refreshment samples,” *Econometrica*, 69, 1645–1659.
- IMBENS, G. W. AND C. F. MANSKI (2004): “Confidence intervals for partially identified parameters,” *Econometrica*, 72, 1845–1857.
- KAIDO, H. AND A. SANTOS (2014): “Asymptotically efficient estimation of models defined by convex moment inequalities,” *Econometrica*, 82, 387–413.
- MANSKI, C. F. (1989): “Anatomy of the selection problem,” *Journal of Human resources*, 343–360.
- (2005): “Partial identification with missing data: concepts and findings,” *International Journal of Approximate Reasoning*, 39, 151–165.
- MOLINARI, F. (2020): “Microeconometrics with partial identification,” *Handbook of econometrics*, 7, 355–486.
- MOON, H. R. AND F. SCHORFHEIDE (2009): “Estimation with overidentifying inequality moment conditions,” *Journal of Econometrics*, 153, 136–154.
- ROMANO, J. P. AND A. M. SHAIKH (2008): “Inference for identifiable parameters in partially identified econometric models,” *Journal of Statistical Planning and Inference*, 138, 2786–2807.
- RUBIN, D. B. (1976): “Inference and missing data,” *Biometrika*, 63, 581–592.
- SEMYKINA, A. AND J. WOOLDRIDGE (2010): “Estimating Panel Data Models in the Presence of Endogeneity and Selection: Theory and Application,” *Journal of Econometrics*, 157, 375–380.
- SHAPIRO, A. (1991): “Asymptotic analysis of stochastic programs,” *Annals of Operations Research*, 30, 169–186.
- VAN DER VAART, A. AND J. WELLNER (2023): *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer.
- WOOLDRIDGE, J. (2002): *Econometric Analysis of Cross Section and Panel Data*, MIT Press.
- WOUTERSEN, T. (2006): “A simple way to calculate confidence intervals for partially identified parameters,” *Unpublished Manuscript, Johns Hopkins University*.

# Appendices

## A Proof of Proposition 1

**Step 1:  $N(\theta)$  is compact and convex.**

Convexity is trivial. Since  $\mathcal{Z}_1 \times \mathcal{Z}_2$  is compact, Theorem 15.11 and Corollary 15.7 in [Aliprantis and Border \(2006\)](#) imply that  $\Pi^0$  is compact in the weak topology and the map  $\pi^0 \mapsto \nu_{\pi^0}(\theta)$  is continuous. Hence,  $N(\theta)$  is compact as the image of a compact set under a continuous map.

**Step 2:  $\Theta_I = Q^{-1}(\{0\})$ .**

Since  $N(\theta)$  is closed and convex,  $0 \in N(\theta)$  if and only if its support function is everywhere nonnegative:  $\psi_{N(\theta)}(u) \geq 0$  for all  $u \in \mathbb{R}^{\dim(\phi)}$ . Indeed, if  $0 \in N(\theta)$ , then  $\psi_{N(\theta)}(u) = \max_{\nu \in N(\theta)} u'\nu \geq 0$ . Conversely, if  $0 \notin N(\theta)$ , then by strong separation of a point from the closed convex set  $N(\theta)$ , there exists  $u \neq 0$  and  $\alpha > 0$  such that  $u'\nu \leq \alpha$  for all  $\nu \in N(\theta)$ . This implies  $\psi_{N(\theta)}(u) \leq -\alpha < 0$ .

Since  $\psi_{N(\theta)}(0) = 0$  for all  $\theta$ , nonnegativity of  $\psi_{N(\theta)}$  is equivalent to the equality of

$$\min_{u \in \mathbb{R}^{\dim(\phi)}} \psi_{N(\theta)}(u) \tag{5}$$

to zero. Let us show that the latter is equivalent to the equality of

$$\min_{u \in \mathbb{B}} \psi_{N(\theta)}(u) \tag{6}$$

to zero. Indeed, if (5) is zero, then the minimum is achieved at  $u = 0$ , and hence (6) is zero. Conversely, if (5) is nonzero, then there exists  $u$  such that  $\psi_{N(\theta)}(u) < 0$ . By the positive homogeneity of the support function,  $\psi_{N(\theta)}(u/\|u\|) = \psi_{N(\theta)}(u)/\|u\| < 0$ , and hence the expression (6) is negative.

**Step 3:  $\Theta_I$  is compact.**

Since  $\Theta_I = Q^{-1}(\{0\})$ , it suffices to establish the continuity of  $Q(\theta)$ . For this, notice that

$$(u, \theta, \pi^0) \mapsto u'\nu_{\pi^0}(\theta)$$

is a continuous function on the compact set  $\mathbb{B} \times \Theta \times \Pi^0$ , where  $\Pi^0$  is equipped with the weak topology. Berge's maximum theorem (see, e.g., Theorem 17.31 in [Aliprantis and Border \(2006\)](#)) implies that the function

$$(u, \theta) \mapsto \max_{\pi^0 \in \Pi^0} u'\nu_{\pi^0}(\theta) = \psi_{N(\theta)}(u)$$

is continuous on the compact set  $\mathbb{B} \times \Theta$ . Applying Berge's theorem again implies that  $Q(\theta)$  is continuous, completing the proof.

## B Proof of Proposition 2

Denote  $c(z_1, z_2) = u' \phi(z_1, z_2, \theta)$  and let  $\Pi^0$  be the set of probability measures  $\pi^0$  on  $\mathcal{Z}_1 \times \mathcal{Z}_2$  with the first marginal  $\pi_1^0(\cdot) = \mathbb{P}(Z_1 \in \cdot \mid S = 0)$ .

Fix any  $\pi^0 \in \Pi^0$ . There exists a Markov kernel  $\pi^0(dz_2 \mid z_1)$  such that

$$\pi^0(dz_1, dz_2) = \pi_1^0(dz_1) \pi^0(dz_2 \mid z_1),$$

see, e.g., Theorems 10.2.1 and 10.2.2 in [Dudley \(2018\)](#). Hence

$$\mathbb{E}_{\pi^0}[c(Z_1, Z_2)] = \int_{\mathcal{Z}_1} \left( \int_{\mathcal{Z}_2} c(z_1, z_2) \pi^0(dz_2 \mid z_1) \right) \pi_1^0(dz_1).$$

For each  $z_1$ ,

$$\int_{\mathcal{Z}_2} c(z_1, z_2) \pi^0(dz_2 \mid z_1) \leq \sup_{z_2 \in \mathcal{Z}_2} c(z_1, z_2) = \max_{z_2 \in \mathcal{Z}_2} c(z_1, z_2),$$

where the equality follows from continuity of  $c(z_1, \cdot)$  and compactness of  $\mathcal{Z}_2$ . Integrating and taking the maximum over  $\pi^0 \in \Pi^0$  yields

$$\max_{\pi^0 \in \Pi^0} \mathbb{E}_{\pi^0}[c(Z_1, Z_2)] \leq \int \max_{z_2 \in \mathcal{Z}_2} c(z_1, z_2) \pi_1^0(dz_1). \quad (7)$$

We now establish a matching lower bound. Let  $\{q(k)\}_{k \geq 1}$  be a countable dense subset of  $\mathcal{Z}_2$ . For each  $N \geq 1$ , define

$$m_N(z_1) := \max_{1 \leq k \leq N} c(z_1, q(k))$$

and let  $k_N(z_1)$  be an arbitrary element of the set

$$\operatorname{argmax}_{1 \leq k \leq N} c(z_1, q(k)) = \{k \in \{1, \dots, N\} : c(z_1, q(k)) = m_N(z_1)\}.$$

Denote  $q_N(z_1) = q(k_N(z_1))$  so that  $m_N(z_1) = c(z_1, q_N(z_1))$ . Let  $\delta_{q_N(z_1)}(dz_2)$  be the point mass at  $q_N(z_1)$ . Define the joint distribution

$$\pi_N(dz_1, dz_2) := \pi_1^0(dz_1) \delta_{q_N(z_1)}(dz_2).$$

Then  $\pi_N \in \Pi^0$  and

$$\mathbb{E}_{\pi_N}[c(Z_1, Z_2)] = \int m_N(z_1) \pi_1^0(dz_1).$$

Since  $\{q(k)\}_{k \geq 1}$  is dense in the compact set  $\mathcal{Z}_2$  and  $c(z_1, \cdot)$  is continuous, we have, as  $N \rightarrow \infty$ ,

$$m_N(z_1) = \max_{1 \leq k \leq N} c(z_1, q(k)) \uparrow \max_{1 \leq k < \infty} c(z_1, q(k)) = \max_{z_2 \in \mathcal{Z}_2} c(z_1, z_2) \quad \text{for all } z_1 \in \mathcal{Z}_1.$$

Moreover,  $|m_N(z_1)| \leq \sup_{z_2 \in \mathcal{Z}_2} |c(z_1, z_2)| \leq C$  for some constant  $C < \infty$  independent of  $z_1$  due to continuity of  $c(\cdot, \cdot)$  and compactness of  $\mathcal{Z}_1$  and  $\mathcal{Z}_2$ . Hence, by dominated convergence,

$$\lim_{N \rightarrow \infty} \int m_N(z_1) \pi_1^0(dz_1) = \int \max_{z_2 \in \mathcal{Z}_2} c(z_1, z_2) \pi_1^0(dz_1).$$

Therefore,

$$\max_{\pi^0 \in \Pi^0} \mathbb{E}_{\pi^0}[c(Z_1, Z_2)] \geq \sup_{N \geq 1} \mathbb{E}_{\pi_N}[c(Z_1, Z_2)] = \int \max_{z_2 \in \mathcal{Z}_2} c(z_1, z_2) \pi_1^0(dz_1).$$

Combining with the upper bound (7) yields

$$\max_{\pi^0 \in \Pi^0} \mathbb{E}_{\pi^0}[c(Z_1, Z_2)] = \int \max_{z_2 \in \mathcal{Z}_2} c(z_1, z_2) \pi_1^0(dz_1) = \mathbb{E} \left[ \max_{z_2 \in \mathcal{Z}_2} c(Z_1, z_2) \mid S = 0 \right],$$

which completes the proof.

## C Proof of Theorem 1

We will follow the proof strategy in [Franguridi and Liu \(2025\)](#) and show that  $\Theta_I \subset \hat{\Theta}_I$  w.p.a. 1 and that for every  $\delta > 0$ ,  $\hat{\Theta}_I \subset \Theta_I^\delta$  w.p.a. 1, where  $\Theta_I^\delta$  is the  $\delta$ -enlargement of  $\Theta_I$ , i.e.

$$\Theta_I^\delta = \{\theta \in \Theta : d(\theta, \Theta_I) \leq \delta\}.$$

Compactness of  $\Theta_I$  will then imply the required Hausdorff convergence.

First, let us show that  $\Theta_I \subset \hat{\Theta}_I$  w.p.a. 1. We have

$$\sup_{\theta \in \Theta_I} \hat{Q}_n(\theta) \leq \sup_{\theta \in \Theta_I} Q(\theta) + \|\hat{Q}_n - Q\|_\infty = \|\hat{Q}_n - Q\|_\infty.$$

By [Theorem 2](#) and [Assumption 3](#), the right-hand side is smaller than  $\eta_n$  w.p.a. 1, and hence  $\Theta_I \subset \hat{\Theta}_I$  w.p.a. 1.

Now let us show that for every  $\delta > 0$ ,  $\hat{\Theta}_I \subset \Theta_I^\delta$  w.p.a. 1. By [Assumption 2](#), we have  $\inf_{\theta \notin \Theta_I^\delta} Q(\theta) \geq m(\delta) > 0$ . On the event  $\|\hat{Q}_n - Q\|_\infty \leq r_n$ , we have

$$\inf_{\theta \notin \Theta_I^\delta} \hat{Q}_n(\theta) \geq \inf_{\theta \notin \Theta_I^\delta} Q(\theta) - \|\hat{Q}_n - Q\|_\infty \geq m(\delta) - r_n.$$

Choose  $n$  large enough so that  $r_n \leq \eta_n < m(\delta)/2$ . Then

$$\inf_{\theta \notin \Theta_I^\delta} \hat{Q}_n(\theta) \geq m(\delta)/2 > \eta_n,$$

and hence no  $\theta$  outside of  $\Theta_I^\delta$  belongs to  $\hat{\Theta}_I$ . Therefore,  $\hat{\Theta}_I \subset \Theta_I^\delta$  w.p.a. 1.

## D Proof of Theorem 2

Define the data vector  $W = (S, Z_1', SZ_2')'$  and let

$$f_{u,\theta}(W) = Su'\phi(Z_1, SZ_2, \theta) + (1 - S) \max_{z_2 \in \mathcal{Z}_2} u'\phi(Z_1, z_2, \theta),$$

so that  $\hat{\psi}_{N(\theta)}(u) = \frac{1}{n} \sum_{i=1}^n f_{u,\theta}(W_i)$  and  $\psi_{N(\theta)}(u) = \mathbb{E} f_{u,\theta}(W)$ . For brevity, drop the arguments  $Z_1$  and  $SZ_2$  in  $\phi(Z_1, SZ_2, \theta)$  and  $\max_{z_2} \phi(Z_1, z_2, \theta)$ . Write

$$\begin{aligned} & f_{u_1, \theta_1}(W) - f_{u_2, \theta_2}(W) \\ &= S(u_1'\phi(\theta_1) - u_2'\phi(\theta_2)) + (1 - S) \left( \max_{z_2} u_1'\phi(z_2, \theta_1) - \max_{z_2} u_2'\phi(z_2, \theta_2) \right) \\ &= S((u_1 - u_2)'\phi(\theta_1) - u_2'(\phi(\theta_2) - \phi(\theta_1))) \\ &+ (1 - S) \left( \max_{z_2} u_1'\phi(z_2, \theta_1) - \max_{z_2} u_2'\phi(z_2, \theta_1) + \max_{z_2} u_2'\phi(z_2, \theta_1) - \max_{z_2} u_2'\phi(z_2, \theta_2) \right). \end{aligned}$$

By Assumption 5, the term multiplied by  $S$  can be bounded in absolute value by

$$\|u_1 - u_2\| \|\phi(Z_1, SZ_2, \theta_1)\| + \|u_2\| L(Z_1) \|\theta_1 - \theta_2\|.$$

For the first term multiplying  $(1 - S)$ , we have

$$\left| \max_{z_2} u_1'\phi(z_2, \theta_1) - \max_{z_2} u_2'\phi(z_2, \theta_1) \right| \leq \max_{z_2} |(u_1 - u_2)'\phi(z_2, \theta_1)| \leq \|u_1 - u_2\| \max_{z_2} \|\phi(Z_1, z_2, \theta_1)\|.$$

Finally, for the second term multiplying  $(1 - S)$ , we have

$$\begin{aligned} & \left| \max_{z_2} u_2'\phi(z_2, \theta_1) - \max_{z_2} u_2'\phi(z_2, \theta_2) \right| \leq \max_{z_2} |u_2'(\phi(z_2, \theta_1) - \phi(z_2, \theta_2))| \\ & \leq \|u_2\| \max_{z_2} \|\phi(z_2, \theta_1) - \phi(z_2, \theta_2)\| \leq L(Z_1) \|\theta_1 - \theta_2\|, \end{aligned}$$

where the last inequality follows from  $\|u_2\| \leq 1$  and Assumption 5.

Combining the inequalities above with Assumption 4 implies that the function class  $\mathcal{F} = \{f_{u,\theta} : u \in \mathbb{B}, \theta \in \Theta\}$  is Lipschitz in the parameter with a square integrable constant. By Theorem 2.7.17 in [van der Vaart and Wellner \(2023\)](#),  $\mathcal{F}$  is Donsker.

## E Proof of Corollary 1

For any  $\psi \in C(\mathbb{B})$ , the functional  $\psi \mapsto \min_{u \in \mathbb{B}} \psi(u)$  is Hadamard directionally differentiable with the derivative

$$h \mapsto \min_{u \in U(\psi)} h(u), \quad h \in C(\mathbb{B}),$$

where

$$U(\psi) = \arg \min_{u \in \mathbb{B}} \psi(u),$$

as established in Theorem 3.1 in Shapiro (1991). Applying the functional delta method (e.g., Theorem 3.10.5 in van der Vaart and Wellner (2023)) to Theorem 2 completes the proof.

## F Proof of Theorem 3

We need the following two technical lemmas, which are completely abstract and independent of our framework.

**Lemma 1.** *Suppose Assumptions 6 and 7 hold. Let  $\hat{\psi}$  be random functions on  $\mathbb{B}$  converging uniformly in probability to  $\psi$ . Then*

$$d_H(\hat{U}(\varepsilon_n), U) = o_p(1).$$

*Proof.* The Hausdorff convergence is implied by

$$\sup_{u \in \hat{U}(\varepsilon_n)} d_H(u, U) = o_p(1) \tag{8}$$

and

$$\sup_{u \in U} d_H(u, \hat{U}(\varepsilon_n)) = o_p(1). \tag{9}$$

First, let us establish (8). Let  $\delta_n = \|\hat{\psi} - \psi\|_{\mathbb{B}}$ . For  $v \in \hat{U}(\varepsilon_n)$ , we have

$$\psi(v) \leq \hat{\psi}(v) + \delta_n \leq \min_{u \in \hat{U}(\varepsilon_n)} \hat{\psi}(u) + \varepsilon_n + \delta_n \leq \min_{u \in U} \psi(u) + 2\delta_n + \varepsilon_n,$$

where the second inequality holds by the definition of  $\hat{U}(\varepsilon_n)$ . Combining with Assumption 6 yields

$$\kappa \cdot d_H(v, U) \leq \psi(v) - \min_{u \in U} \psi(u) \leq 2\delta_n + \varepsilon_n.$$

Taking the supremum over  $v \in \hat{U}(\varepsilon_n)$  and noting that  $2\delta_n + \varepsilon_n \rightarrow 0$  establishes (8).

Now let us establish (9). Indeed, for any  $u \in U$ ,

$$\hat{\psi}(u) \leq \psi(u) + \delta_n = \min_{u \in U} \psi(u) + \delta_n \leq \min_{u \in \hat{U}(\varepsilon_n)} \hat{\psi}(u) + 2\delta_n.$$

Consider the event  $E_n = \{\delta_n \leq \varepsilon_n/2\}$ . Using the inequality above, on this event,

$$\hat{\psi}(u) \leq \min_{u \in \hat{U}(\varepsilon_n)} \hat{\psi}(u) + \varepsilon_n,$$

which implies  $u \in \hat{U}(\varepsilon_n)$ , and hence  $U \subset \hat{U}(\varepsilon_n)$ . By Assumption 7,  $P(E_n) \rightarrow 1$ , and so  $U \subset \hat{U}(\varepsilon_n)$  w.p.a. 1. Therefore,  $\sup_{u \in U} d_H(u, \hat{U}(\varepsilon_n)) = 0$  w.p.a. 1, establishing (9).  $\square$

**Lemma 2.** *Suppose  $d_H(\hat{U}, U) = o_p(1)$ , where  $\hat{U}, U$  are nonempty compact subsets of  $\mathbb{B}$ . Then for any continuous function  $h$  on  $\mathbb{B}$ ,*

$$\min_{u \in \hat{U}} h(u) - \min_{u \in U} h(u) = o_p(1).$$

*Proof.* The modulus of continuity of  $h$  is

$$\omega_h(\varepsilon) = \max_{u_1, u_2 \in \mathbb{B}: \|u_1 - u_2\| \leq \varepsilon} |h(u_1) - h(u_2)|.$$

Since  $h$  is continuous and  $\mathbb{B}$  is compact,  $\omega_h(\varepsilon) \downarrow 0$  as  $\varepsilon \downarrow 0$ .

Pick  $u^* \in \arg \min_{u \in U} h(u)$ . By definition of the Hausdorff distance, there exists  $\hat{u} \in \hat{U}$  such that  $\|\hat{u} - u^*\| \leq d_H(\hat{U}, U)$ . Then

$$\min_{u \in \hat{U}} h(u) \leq h(\hat{u}) \leq h(u^*) + \omega_h(d_H(\hat{U}, U)),$$

which implies

$$\min_{u \in \hat{U}} h(u) - \min_{u \in U} h(u) \leq \omega_h(d_H(\hat{U}, U)).$$

Similarly, pick any  $\hat{u}^* \in \arg \min_{u \in \hat{U}} h(u)$ . There exists  $u \in U$  such that  $\|\hat{u}^* - u\| \leq d_H(\hat{U}, U)$ . Then

$$\min_{u \in U} h(u) \leq h(\hat{u}^*) + \omega_h(d_H(\hat{U}, U)),$$

which implies

$$\min_{u \in U} h(u) - \min_{u \in \hat{U}} h(u) \leq \omega_h(d_H(\hat{U}, U)).$$

But since  $\omega_h(\varepsilon)$  is continuous at  $\varepsilon = 0$  and  $d_H(\hat{U}, U) = o_p(1)$ , by the continuous mapping theorem,  $\omega_h(d_H(\hat{U}, U)) = o_p(1)$ , completing the proof.  $\square$

Let us show that the estimator  $\hat{\chi}'$  is amenable to Remark 3.4 in Fang and Santos (2019). Indeed, let  $\hat{\Delta} = \|h_1 - h_2\|_{\hat{U}(\varepsilon_n)}$  (uniform norm on  $\hat{U}(\varepsilon_n)$ ). Then  $h_2(u) \geq h_1(u) - \hat{\Delta}$ , and so

$$\min_{u \in \hat{U}(\varepsilon_n)} h_2(u) \geq \min_{u \in \hat{U}(\varepsilon_n)} h_1(u) - \hat{\Delta}.$$

Swapping  $h_1$  and  $h_2$  yields the reverse inequality. Combining the two inequalities yields

$$|\hat{\chi}'(h_1) - \hat{\chi}'(h_2)| = \left| \min_{u \in \hat{U}(\varepsilon_n)} h_1(u) - \min_{u \in \hat{U}(\varepsilon_n)} h_2(u) \right| \leq \hat{\Delta} \leq \|h_1 - h_2\|_B,$$

i.e.  $\hat{\chi}'$  is 1-Lipschitz. Therefore, according to Remark 3.4, it suffices to establish pointwise consistency of  $\hat{\chi}'(h)$ , i.e., for any  $h \in C(\mathbb{B})$ ,

$$|\hat{\chi}'(h) - \chi'(h)| = o_p(1).$$

Indeed, Theorem 2 implies that  $\hat{\psi}$  converges uniformly in probability to  $\psi$  on  $\mathbb{B}$ . By Lemma 1, the argmin sets  $\hat{U}(\varepsilon_n)$  converge to  $U$  in the Hausdorff distance in probability. Using Lemma 2 completes the proof.

## G Proof of Proposition 3

Dropping the argument  $\theta$  for brevity,

$$\mathbb{E}[\phi(Z_1, Z_2, Z_3)] = p_3 \mathbb{E}[\phi(Z_1, Z_2, Z_3)|S_3 = 1] + (1 - p_3) \mathbb{E}[\phi(Z_1, Z_2, Z_3)|S_3 = 0].$$

Notice that  $\mathbb{E}[\phi(Z_1, Z_2, Z_3)|S_3 = 1]$  is point identified since  $S_3 = 1$  implies  $S_2 = 1$ . Then

$$\begin{aligned} \mathbb{E}[\phi(Z_1, Z_2, Z_3)|S_3 = 0] &= p_{2|0} \mathbb{E}[\phi(Z_1, Z_2, Z_3)|S_2 = 1, S_3 = 0] \\ &\quad + (1 - p_{2|0}) \mathbb{E}[\phi(Z_1, Z_2, Z_3)|S_2 = S_3 = 0]. \end{aligned}$$

Arguing as in the proof of Proposition 2 completes the proof.

## H Proof of Theorem 4

We map our testing framework into Section 3.4.2 of [Fang and Santos \(2019\)](#) and then apply their Corollary 3.2. To this end, multiply our statistic by  $-1$  to obtain

$$\begin{aligned} T(\theta) &= \chi(\psi_\theta) = -\min_{u \in \mathbb{B}} \psi_\theta(u), \\ \hat{T}(\theta) &= \chi(\hat{\psi}_\theta) = -\min_{u \in \mathbb{B}} \hat{\psi}_\theta(u), \end{aligned}$$

where

$$\chi(h) = -\min_{u \in \mathbb{B}} h(u).$$

This functional is Hadamard differentiable at  $\psi_\theta$  with the derivative

$$\chi'_{\psi_\theta}(h) = -\min_{u \in U} h(u),$$

where  $U = \arg \min_{u \in \mathbb{B}} \psi_\theta(u)$ . Notice that  $T(\theta) = 0$  if  $\theta \in \Theta_I$  and  $T(\theta) > 0$  otherwise. Hence, we are testing the hypothesis  $H_0 : \chi(\psi_{\theta_0}) \leq 0$  against  $H_1 : \chi(\psi_{\theta_0}) > 0$ , as in equation (31) of [Fang and Santos \(2019\)](#).

Let us show now that the functional  $\chi'(h)$  is convex. Indeed, for all  $h_1, h_2$  and  $u \in U$ , we have

$$th_1(u) + (1-t)h_2(u) \geq t \min_{u \in U} h_1(u) + (1-t) \min_{u \in U} h_2(u),$$

and hence

$$\begin{aligned} -\chi'(th_1 + (1-t)h_2) &= \min_{u \in U} \{th_1(u) + (1-t)h_2(u)\} \\ &\geq t \min_{u \in U} h_1(u) + (1-t) \min_{u \in U} h_2(u) = -t\chi'(h_1) - (1-t)\chi'(h_2). \end{aligned}$$

Multiplying this inequality by  $-1$  establishes the convexity of  $\chi'(h)$ . Applying Corollary 3.2 in [Fang and Santos \(2019\)](#) completes the proof.