

Sharp Identification for Regressions with Interval-Observed and Missing Covariates

Martin Fankhauser (Bocconi), Gil Jan Peled (Northwestern)

September 2026

Motivation

Consider a dataset with **missing** and **interval** data:

Y	X ₁	X ₂
0.4	2	[1,3]
[0.1,0.4]	?	4
1.2	-0.7	[0.5,1]
?	4	1
-0.2	3	?

Examples: sample attrition, survey non-response, top-coding, multiple price lists, unfolding brackets, bid-ask intervals, mixed-frequency data, ... HRS

How can we “run a regression” using this sample without imposing additional assumptions?

Imputation approaches can be misleading

Illustration: Linear model $Y = X'\theta + \epsilon$

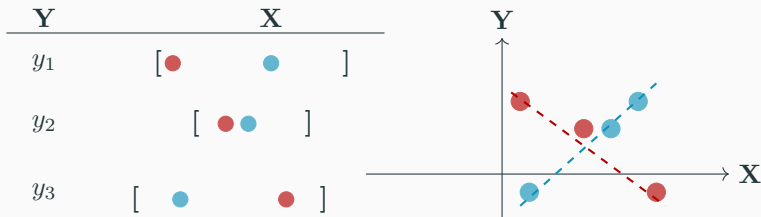


Figure 1: Regression with interval regressors ($y_1 > y_2 > y_3$)

Taking the midpoint of the observed intervals can be misleading (*wrong sign*), same for more sophisticated imputation rules.

Taking boundary points does not work [Details](#)

Incomplete data yields many admissible slopes, Θ_I .

Contribution

We consider **conditional moment models**

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

where the outcome Y and **the continuous conditioning X are missing or interval-valued.**

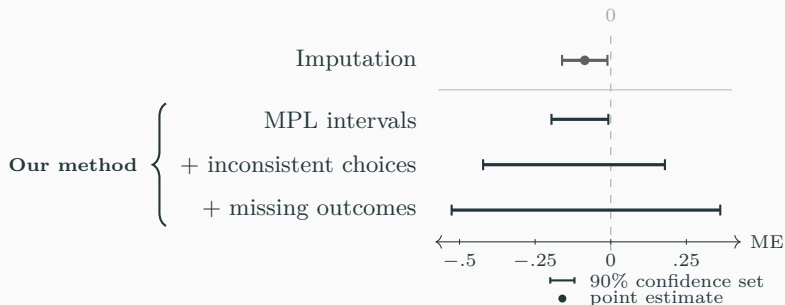
Examples: linear, probit, logit regressions, quantile regressions, structural equations, independence restrictions, . . .

Only assuming $P(Y \in [Y_l, Y_u]) = 1$ and $P(X \in [X_l, X_u]) = 1$, we

- **characterize the sharp identified set**, i.e. all θ consistent with the incomplete data and the model
- derive an **estimator** and valid **incomplete data robust confidence sets**
- provide an **R package**

Application preview

Castillo, Jordan, and Petrie (2018): Do experimentally elicited discount rates predict on-time high-school graduation?



Multiple price list (MPL) yields only interval responses for the discounts (MPL intervals) for all. 8% respond irrationally (+ inconsistent choices), for 30% graduation status is not observed (+ missing outcomes).

Partial identification:

Manski and Tamer (2002), Aucejo, Bugni, and Hotz (2017), Magnac and Maurin (2008), Asher, Novosad, and Rafkin (2018), Beresteanu and Molinari (2008), Beresteanu, Molchanov, and Molinari (2011), Schennach (2014), Chesher and Rosen (2017), Chesher and Rosen (2020), Lee (2009), Horowitz and Manski (2000), Franguridi and Moon (2026)

⇒ No auxiliary assumptions, sharp set, confidence sets, R implementation

Point identification / stronger assumptions / external info:

Hsiao (1983), Robins, Rotnitzky, and Zhao (1994), Rigobon and Stoker (2009), Duan and Pelger (2025), Freyberger et al. (2025), Bretschneider and Kaddoura (2026), Pacini (2018), Breunig and Haan (2021), Miao and E. (2018), Fang, Zhao, and Shao (2018)

⇒ Complementary, our result shows how informative these methods are

Partial identification reviews:

Molinari (2020), Canay and Shaikh (2017), Kline and Tamer (2023)

Outline

Theory

1. Identification
2. Estimator and Inference
3. Implementation

Simulation

1. Comparison with naive approaches
2. Potential information gain

Application

1. Revisit Castillo, Jordan, and Petrie (2018) (Multiple Price List survey design)
2. Simplifies robustness checks

Identification

We study the following model

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

where Y, X are scalar random variables living on $(\Omega, \mathcal{A}, F_0)$. $\theta_0 \in \Theta$ is the true unknown parameter.

Example:

Suppose $m(Y, X, \theta) = Y - X'\theta$, which summarizes our usual linear (mean) regression model:

$$Y = X'\theta + \epsilon \quad \mathbf{E}[\epsilon | X] = \mathbf{E}[\epsilon] = 0$$

Incomplete data

We assume to know nothing about the censoring or missing data generating process, except a basic consistency condition.

Assumption 1

Let $W^o \equiv (Y_l, Y_u, X_l, X_u)$ be the observed random variables defined on the same probability space as the latent $W = (Y, X)$. Assume

$$P(Y \in [Y_l, Y_u], X \in [X_l, X_u]) = 1$$

Missing data case: $[X_l, X_u] = \mathcal{X}$, i.e. intervals cover full support.
Some observations can be fully observed e.g. $X_l = X_u$.

Apart from that we only need regularity conditions: $m(\cdot)$ is continuous and uniformly bounded, Θ compact.

Regularity conditions

Defining the sharp identified set

Define the **set of admissible completion laws**:

$$\Gamma \equiv \left\{ \eta \in \Delta(\mathcal{W} \times \mathcal{W}^O) : \begin{aligned} &\eta_{W^O} = P_{W^O}, \\ &\eta(Y \in [Y_l, Y_u], X \in [X_l, X_u]) = 1 \end{aligned} \right\}.$$

The **sharp identified set** contains all parameter values for which there **exists an admissible completion of the data fulfilling the conditional moment restriction**.

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

From conditional to unconditional moments

We can transform conditional moments into **infinitely many unconditional moments** via **instrument/basis functions**

$$\begin{aligned} \mathbf{E}[m(Y, X, \theta)|X] &= 0 \quad \text{a.s.} \\ \iff \mathbf{E}[m(Y, X, \theta)g(X)] &= 0 \quad \forall g \in \mathcal{G} \end{aligned}$$

where $\mathcal{G}$ is the properly chosen class of instrument functions.

Intuition:

Suppose $X \in \{\text{male, female}\}$, $\mathbf{E}[m(Y, X, \theta)|X] = 0$ requires:

$$\mathbf{E}[m(Y, X, \theta)\mathbf{1}\{X = \text{male}\}] = 0$$

$$\mathbf{E}[m(Y, X, \theta)\mathbf{1}\{X = \text{female}\}] = 0$$

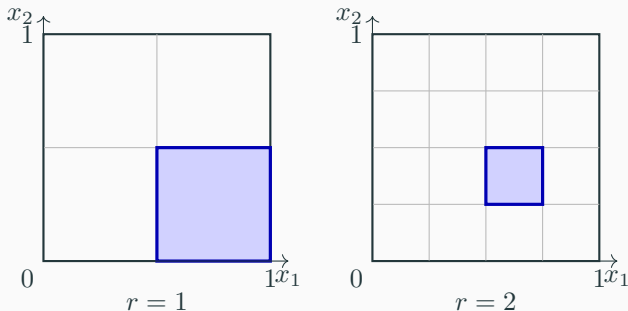
for continuous X we cannot list them, so check rich collection of regions of $\mathcal{X}$.

Example: weighted cubes illustration

Considers all resolutions r simultaneously.

Definition

Assumptions



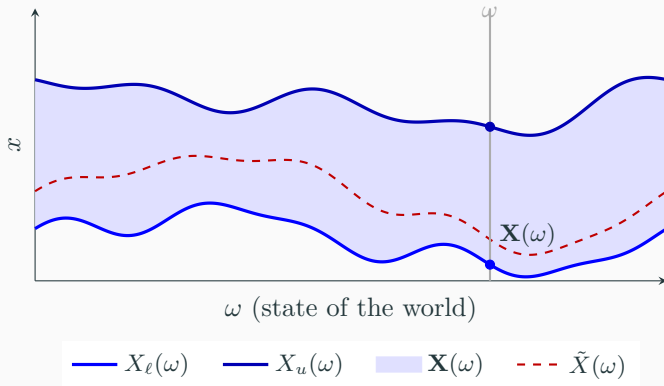
Formally, define the map

$$M_{\theta}(y, x) \equiv \begin{pmatrix} m(y, x, \theta)g_1(x) \\ m(y, x, \theta)g_2(x) \\ \vdots \end{pmatrix} \in \ell_1$$

In words, $M_{\theta}(y, x)$ lists moments for all cubes.

Random set and its selections

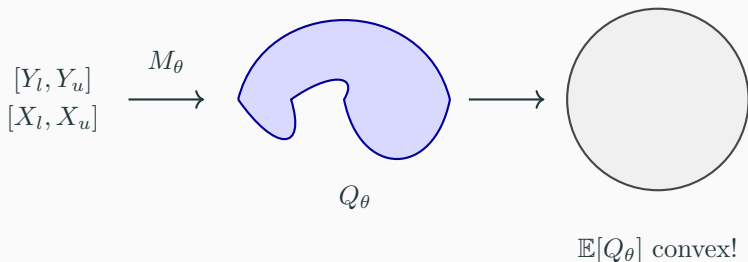
Random set $\mathbf{X}(\omega) = [X_\ell(\omega), X_u(\omega)]$ and a selection $\tilde{X}(\omega) \in \mathbf{X}(\omega)$



References: Molchanov (2017), Molchanov and Molinari (2018), Beresteanu and Molinari (2008), Beresteanu, Molchanov, and Molinari (2011), Beresteanu and Molchanov (2012)

Intuition and proof idea

For each candidate $\theta \in \Theta$:



In a point identified case we would require $\mathbf{E}[m(Y, X, \theta)|X] = 0$.

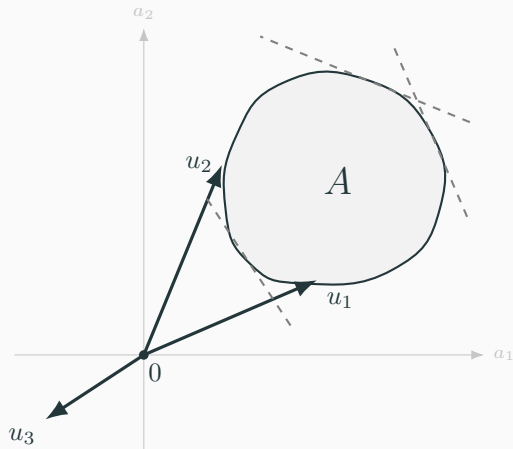
We need to check whether $\mathbf{0}$ is in the **Aumann expectation** of Q_θ :

$$\mathbf{0} \in \mathbb{E}[Q_\theta] \equiv \text{cl}\{\mathbf{E}[q] : q \in Q_\theta\}$$

$\Rightarrow$ Implies $\exists Y, X$ with $\eta_{Y,X}$ such that $\mathbf{E}_{\eta_{Y,X}}[m(Y, X, \theta)|X] = 0$.

Support functions trace out convex sets

The support function for a set A is $h(A, u) \equiv \sup_{a \in A} u'a$.



For u_3 support function is negative, thus 0 is not included in A .

Identification theorem

With this background, the condition

$$\mathbf{0} \in \mathbb{E}[Q_\theta]$$

together with standard properties of Aumann expectations yields:

Theorem 1

Let Assumptions 1-4 hold, then

$$\Theta_I^* = \left\{ \theta \in \Theta : \inf_{u \in B_1^\infty} \mathbf{E}[h(Q_\theta, u)] \geq 0 \right\}.$$

$B_1^\infty = \{u \in \ell_\infty : \|u\|_\infty \leq 1\}$. We show Q lives in ℓ_1 (abs sum sequences), whose dual is ℓ_∞ (bounded sequences) so that the separation argument works.

Expectation Properties

Why is it sharp?

Assumptions

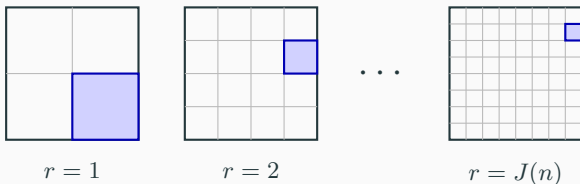
Consistency and inference

Constructing an estimator

Let $T(\theta) \equiv - \inf_{u \in B_1^\infty} \mathbf{E}[h(Q_\theta, u)]$, $T_{K(n)}(\theta) \equiv - \inf_{u \in B_{K(n)}} \mathbf{E}[h(Q_\theta, u)]$ and define the sample criterion as

$$\hat{T}_{n,K}(\theta) = - \inf_{u \in B_{K(n)}} \frac{1}{n} \sum_{i=1}^n h(Q_{\theta,K(n)}(W_i^O), u)$$

where $K(n) = \sum_{r=\bar{r}}^{J(n)} r$ is the number of cubes.



We need to balance two errors:

$$\underbrace{T(\theta) - T_{K(n)}(\theta)}_{\text{sieve/truncation}} \quad \underbrace{T_{K(n)}(\theta) - \hat{T}_{n,K(n)}(\theta)}_{\text{sampling error}}$$

Theorem 2

Let A1-A5 hold. Suppose that $\Theta_I^* \neq \emptyset$. Let $K(n) \rightarrow \infty$, and let $a_n \asymp \sqrt{\frac{n}{K(n)}}$. Let $\frac{K(n)}{n} \rightarrow 0$, $\tau_n \rightarrow \infty$ and $\frac{\tau_n}{a_n} \rightarrow 0$. Define

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

Then $\rho_H(\hat{\Theta}_n, \Theta_I^*) \xrightarrow{p} 0$.

Proof: establish triangular law of large numbers for RS, verify conditions in Chernozhukov, Hong, and Tamer (2007).

Requires: $K(n) \rightarrow \infty$, i.e. approximation becomes *rich enough*,
 $K(n)/n \rightarrow 0$, i.e. slowly enough relative to n .

Assumption 5

Hausdorff distance ρ_H

Incomplete-data robust confidence sets

The Andrews and Shi (2017) bootstrap critical value $c_{n,1-\alpha}^{AS}(\theta)$ is used to construct confidence sets:

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

Theorem 3

Under A1-A4, for fixed P_{W_0}

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

Moreover, every fixed $\theta \notin \Theta_I^*$ is rejected with probability approaching one.

Proof: verify manageability condition in Andrews and Shi (2017) using similar arguments as for Theorem 2.

Details

Confidence for point

Uniformity

Implementation

Nested optimization

Recall from Theorem 1:

$$\theta \in \Theta_I^* \iff \inf_{u \in B_1^\infty} \mathbf{E}[h(Q_\theta(W^O), u)] \geq 0$$

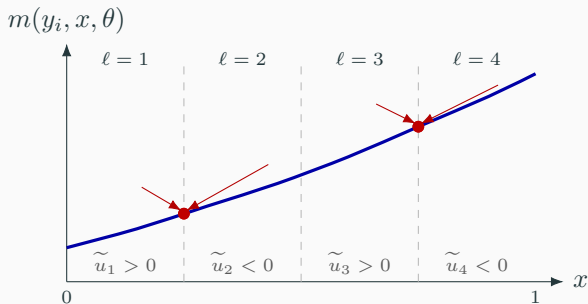
Checking whether a parameter value is in the identified set requires **solving a nested optimization problem**:

$$\inf_{u \in B_1^\infty} \mathbf{E} \left[\max_{(y,x) \in D(W^O)} \sum_{k \geq 1} u_k g_k(x) m(y, x, \theta) \right]$$

This is problematic, as the workflow usually involves a fine parameter grid $\Theta^{\text{candidate}}$, i.e. solving this for each candidate.

Monotone index models (probit, logit, linear, ...)

Suppose Y_i is observed and X_i is missing, so $X_i \in [0, 1]$. The refined cube partition divides the support of X into finitely many cells.

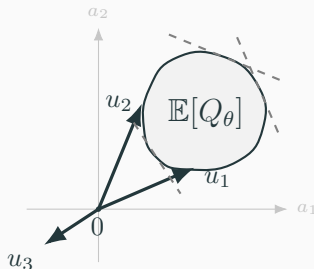


For these models we only need to check finitely many boundary values.

Collapses nested optimization to one linear program [Details](#)

Arbitrarily precise outer sets

We know $\theta \in \Theta_I^*$ iff $0 \in \mathbb{E}[Q_\theta]$. Fix θ :



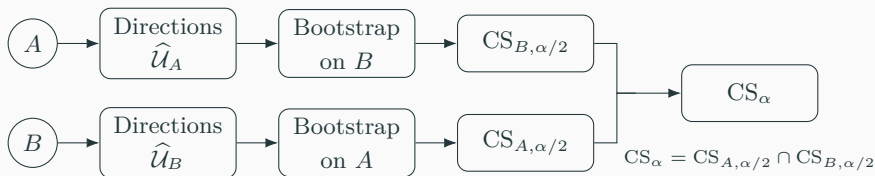
If only $\mathcal{U} = (u_1, u_2)$ are checked θ survives, but if u_3 is included θ is rejected. $\Rightarrow$ Every finite $\mathcal{U}$ defines an **outer set** ABH

Aucejo, Bugni, and Hotz (2017) can be recovered using a special set of directions $\mathcal{U}$.

IDEA: Instead of solving the problem for each parameter value, estimate a set of *informative* directions.

Finite split-sample implementation

To avoid overfitting, we implement the following split sample estimator:



Each directions set is independent of the sample on which it's used for the bootstrap. Thus, the individual confidence sets are valid and

$$P(\theta_0 \notin \text{CS}_A \cap \text{CS}_B) \leq P(\theta_0 \notin \text{CS}_A) + P(\theta_0 \notin \text{CS}_B) \leq \frac{\alpha}{2} + \frac{\alpha}{2} \leq \alpha$$

Simulations

Setting

Consider a simple regression model:

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

with true parameter $(\theta_0, \theta_1) = (0, 1)$.

For an interval-observed observation we set

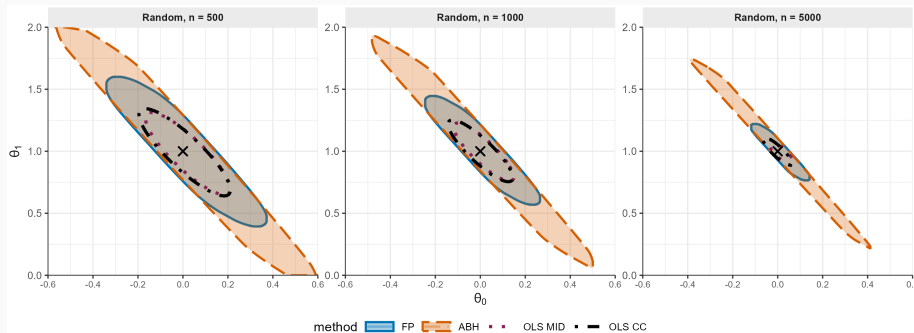
$$X_{i,l} = X_i - w s_i, \quad X_{i,u} = X_i,$$

where $s_i \sim U[0.5, 1.5]$. 500 Monte Carlo replications.

Two designs

- Random-Censoring with 10% interval-valued with $w = 0.2$
- Y-dependent-censoring with 30% interval-valued with $w = 0.5$

Random censoring: 10% interval X

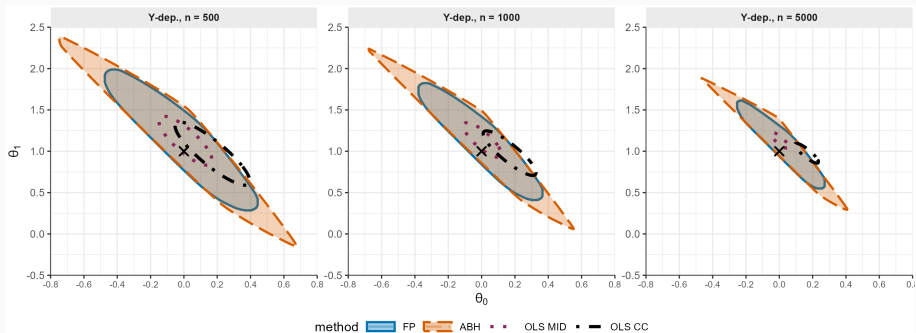


DGP	FP	ABH	OLS MID	OLS CC
Random, $n = 500$	0.958	0.952	0.894	0.890
Random, $n = 1000$	0.960	0.966	0.896	0.902
Random, $n = 5000$	0.974	0.964	0.838	0.890

Valid 90% confidence intervals should have ≥ 0.9

FP: Our method, ABH: Aucejo, Bugni, and Hotz (2017), OLS MID: midpoint, OLS CC: complete cases

Nonrandom censoring: 30% interval X



DGP	FP	ABH	OLS MID	OLS CC
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

Valid 90% confidence intervals should have ≥ 0.9

FP: Our method, ABH: Aucejo, Bugni, and Hotz (2017), OLS MID: midpoint, OLS CC: complete cases

Application

Castillo, Jordan, and Petrie (2018) Summary

Inspired by the famous *Stanford Marshmallow Experiment*, the authors ask:

Is patience among high-schoolers associated with on-time graduation?

- **Measure children's time preferences** (discount rates) *directly* using an incentivized experiment in 8th grade.
- Test whether experimentally measured discount rates predict *on-time* high school graduation five years later.

Discount rate elicitation via multiple price list (MPL)

In an MPL, respondents are faced with a sequence of questions:

Row	A (paid in 1 month)	B (paid in 7 months)
1	\$49	\$50.83
2	\$49	\$52.71
3	\$49	\$54.66
		$\vdots$
20	\$49	\$98.02

Table 1: MPL Design for Discount Rates

If a respondent chooses A for row 2 and B for row 3, we only know the *discount measure* is in between \$52.71 and \$54.66.

- All discount rates are interval valued
- 8% of students switch twice

inconsistent, see discussion in Yu, Zhang, and Zuo (2021)

- For 30% of students graduation status unknown

Missing and interval-data robust results

Estimate a **probit model** with outcome **graduation** and the main regressor the **discount rate**. Here we report the marginal effect at the mean.

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]

- **Point estimate:** +100pp higher discount rate $\Rightarrow \approx 8.6$ pp lower on-time graduation probability.
- The $[\cdot, \cdot]$ denote 90% confidence intervals

Aucejo, Bugni, and Hotz (2017) (ABH) provide an **outer set** for conditional moment models with missing values. Special case of our result. **ABH**

Conclusion

What we have seen:

- **Sharp identification:** The first usable characterization of the smallest set of parameter values consistent with the data and model without any auxiliary assumptions.
- **Valid inference:** Hausdorff-consistent estimation and uniformly valid confidence sets.
- **Practical implementation:** Fast linear programs, split-sample inference, and an **R package**.

Extensions:

Current planned extensions allow researchers to impose additional credible restrictions to progressively tighten the identified set.

How can we make credible use of incomplete, imprecise, or external information in empirical work?

- **Robust Bayesian Inference under Partially Specified Priors** Details

Above we derived Θ_I , all $\theta \in \Theta_I$ are consistent with the data. What if we have some partial knowledge about how to rank θ within the set? I show how to do Bayesian inference without committing to a precise fully specified prior.

- **Power Priors for VARs** (with Marcellino, Tornese)

Systematic framework to exploit external information to sharpen inference in VAR models.

- **Conformal Quantile Estimation for Nowcasting**

Fast and reliable quantile forecasting method.

- **Model Confidence Set for Stata** (with Nespoli, Przewoski)

Software implementation of a widely applied inference procedure.

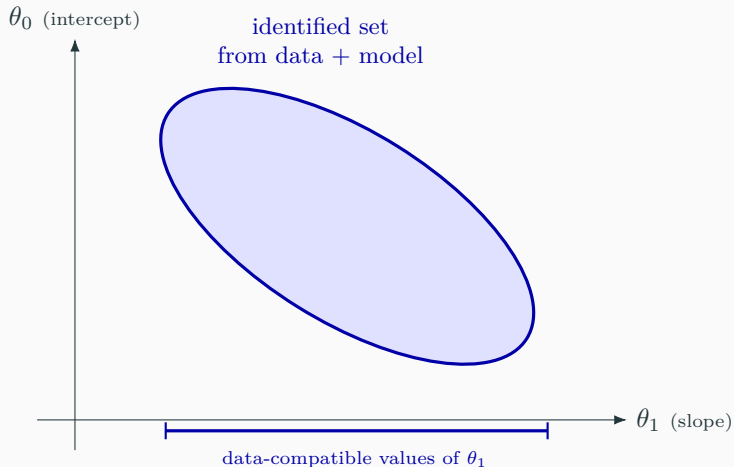
Thanks a lot!

`martin.fankhauser@unibocconi.it`

`www.martin-fankhauser.at`

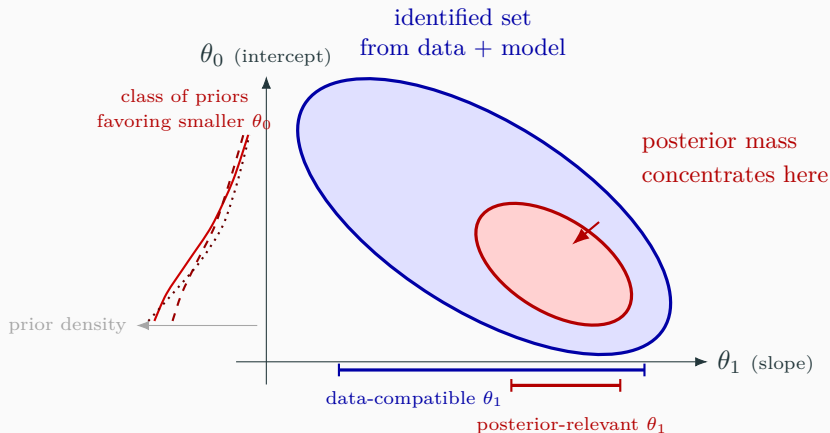
Appendix

Robust Bayes under Partially-Specified Priors



Incomplete data leave both the intercept and slope set identified.

Robust Bayes under Partially-Specified Priors



A class of partially specified priors on θ_0 can sharply restrict posterior beliefs about θ_1 . [Back](#)

Health and Retirement Study

Q020

About how much wage and salary income did you receive in [LAST CALENDAR YEAR], before taxes and other deductions?

[INSTR: DO NOT PROBE DK/RF.]

About how much wage and salary income did you receive in [LAST CALENDAR YEAR], before taxes and other deductions?

\$_____.00 GO TO Q024

DK

RF

Q021-

Q023

Question text: (Thinking about your wage and salary income in [LAST CALENDAR YEAR]):

Did it amount to less than \$____, more than \$____, or what?

PROCEDURES: 3Up, 2Up1Down, 1Up2Down

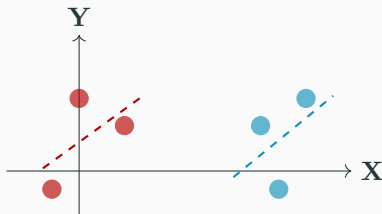
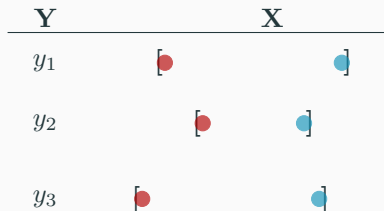
BREAKPOINTS: \$5,000, \$25,000, \$50,000, \$100,000

RANDOM ENTRY POINT ASSIGNMENT [1 (\$5,000)] or [2 (\$25,000)] or

[{NOT 1 and NOT 2} (\$50,000)] AT X042

Plugging in extreme points does not work

Finding all parameter values consistent with the data and model is not trivial:



Different regressions from lower- and upper-endpoint selections
($y_1 > y_2 > y_3$).

For instance, the OLS estimator is non-linear in X . [Back](#)

Regularity conditions on the parametric model

Assumption 2

The parameter space Θ is a nonempty, compact subset of $\mathbb{R}^d$.

Assumption 3

For every $\theta \in \Theta$, the map $(y, x) \mapsto m(y, x, \theta)$ is continuous. The moment map is uniformly bounded: for some $C_1 < \infty$,

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

[Back: Assumption 1](#)

[Back: Theorem 1](#)

Weighted hypercubes

For example, we can define

$$\mathcal{G}_{\text{c-cube}} \equiv \{g_{a,r} : g_{a,r}(x) := w(r)\mathbb{I}\{x \in C_{a,r}\} \quad \text{for } C_{a,r} \in \mathcal{C}_{\text{c-cube}}\},$$

with $w(r) = (100 + r)^{-\beta}$ and $\beta > 1$, where

$$\begin{aligned} \mathcal{C}_{\text{c-cube}} \equiv \{C_{a,r} := \prod_{j=1}^{d_x} ((a_j - 1)/(2r), a_j/(2r)] \in [0, 1]^{d_x} : \\ a = (a_1, \dots, a_{d_x})', a_j \in \{1, 2, \dots, 2r\} \\ \text{for } j = 1, \dots, d_x \text{ and } r = r_0, r_0 + 1, \dots\} \end{aligned}$$

Example:

For $\mathcal{X} = [0, 1]$ and $r = 1$ we have

$$\mathcal{C}_{\text{c-cube}} = \{(0, 0.5], (0.5, 1]\}$$

Regularity conditions on the cubes

Assumption 4

Let $\mathcal{G} = \{g_k : k \geq 1\}$ be a countable class of bounded Borel functions on $\mathcal{X}$, and let $\omega_k > 0$ for every k . Assume

1. the weighted series has a uniformly negligible tail:

$$\sup_{x \in \mathcal{X}} \sum_{k \geq 1} \omega_k |g_k(x)| < \infty, \quad \lim_{K \rightarrow \infty} \sup_{x \in \mathcal{X}} \sum_{k > K} \omega_k |g_k(x)| = 0;$$

2. for every $f \in C(\mathcal{X})$ and every $\varepsilon > 0$, there is a $v_\varepsilon \in \text{span}(\mathcal{G})$ such that $\|f - v_\varepsilon\|_\infty < \varepsilon$. Here v_ε is a finite linear combination of elements of $\mathcal{G}$.

The first point ensures that the set lives in a convenient ambient space (ℓ_1). The second says that the instruments are *comprehensively revealing* (Stinchcombe and White (1998)), i.e. rich enough to detect any violation of the moment restriction.

[Back](#)

Smoothness in θ

Assumption 5

$m(\cdot, \theta)$ has bounded first derivatives with respect to θ , i.e.

$$\sup_{(y, x, \theta) \in \mathcal{W} \times \Theta} \|\nabla_{\theta} m(y, x, \theta)\| \leq L < \infty$$

This can be relaxed to Lipschitz continuity.

[Back](#)

Expectation and sums of random sets

Take two sets K and L the **Minkowski sum** denoted $K \oplus L$ is defined as

$$K \oplus L \equiv \{k + l : k \in K, l \in L\}$$

Let X be a random set then the **Aumann expectations** is defined as:

$$\mathbb{E}[X] \equiv cl(\{\mathbf{E}[x] : x \in Sel(X), \mathbf{E}[x] < \infty\})$$

Under mild assumptions, even if X is not convex, $\mathbb{E}[X]$ is a convex set!

Tracing out Aumann expectation

Because $\mathbb{E}[X]$ is convex, we can trace it out using properties of the **support function**:

- **Exp. of support fct. = support fct. of Exp.:**

$$\mathbf{E}(h(X, u)) = h(\mathbb{E}(X), u) \quad \forall u \in B^\infty \text{ a.s.}$$

- **Verify membership through support fct.:**

$$\text{if } u'a \leq h(A, u) \quad \forall u \in B^\infty \text{ then } a \in A$$

- **Linearization property:**

$$h(cK + dL, u) = ch(K, u) + dh(L, u) \quad \forall u \in B^\infty$$

Why is the random-set characterization sharp?

Bluntly, because we constructed Q_θ such that it fulfills the relevant constraints.

However, note that this is not automatic. For instance, we used $Q_\theta \equiv \text{cl}_{\ell_1} M_\theta(\cdot)$ because M_θ is not continuous.

The proof therefore establishes:

- Link between $\text{Sel}(Q_\theta)$ and Γ
- Link $\mathbb{E}[Q_\theta(W^o)]$ and $\text{cl}_{\ell_1} I_\theta = \text{cl}_{\ell_1} \int M_\theta(y, x) \eta(dy, dx)$
- Taking closure does not affect identification

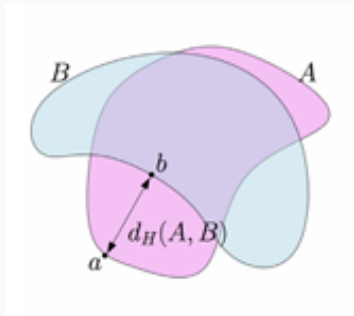
$$0 \in \text{cl}_{\ell_1} I_\theta \iff 0 \in I_\theta$$

[link to Schennach \(2014\) closure formulation](#)

Beresteanu, Molchanov, and Molinari (2011) do not need most of these arguments as they are resolved by assumption and/or mitigate in finite dimensional spaces.

[Back](#)

Hausdorff distance



For two sets A and B ,

$$\rho_H(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} \|a - b\|, \sup_{b \in B} \inf_{a \in A} \|a - b\| \right\}.$$

**Largest distance needed to
match the two sets.**

In our setting,

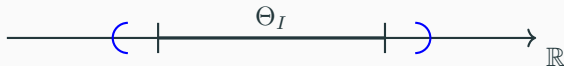
$$\rho_H(\hat{\Theta}_n, \Theta_I^*)$$

measures how far the estimated
identified set is from the true one.

Back

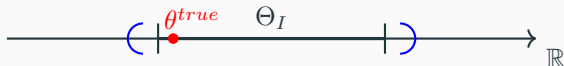
Confidence intervals (C_n) for sets or points?

C_n for the identified set:



$()$ covers the entire set Θ_I with $\text{prob} \geq 1 - \alpha$

C_n for the true parameter:



$()$ covers true parameter θ^{true} with $\text{prob} \geq 1 - \alpha$. IDEA: θ^{true} is either close to one boundary or the other, not both $\Rightarrow$ Shorter!

Back

Uniform vs pointwise consistency: definition

Consider a CI for θ , denoted C_n . C_n is **pointwise consistent** if

$$\forall \theta \in \Theta \quad \liminf_{n \rightarrow \infty} P(\theta \in C_n) \geq 1 - \alpha$$

which could be also written as

$$\inf_{\theta \in \Theta} \liminf_{n \rightarrow \infty} P(\theta \in C_n) \geq 1 - \alpha$$

C_n is **uniform consistent** if

$$\liminf_{n \rightarrow \infty} \inf_{\theta \in \Theta} P(\theta \in C_n) \geq 1 - \alpha$$

Missing-data robust confidence sets

For simplicity take $d_m = 1$. For each candidate θ and direction u , define

$$\tilde{m}(W_i^O, \theta, u) = h(Q_\theta(W_i^O), u), \quad \overline{m}_n(\theta, u) = \frac{1}{n} \sum_{i=1}^n \tilde{m}(W_i^O, \theta, u).$$

The null for a candidate θ is

$$H_0(\theta) : \quad \mathbf{E}[\tilde{m}(W_i^O, \theta, u)] \geq 0 \quad \forall u \in B_1^\infty.$$

We use an Andrews and Shi (2017)-type statistic:

$$T_n^{AS}(\theta) = \sup_{u \in B_1^\infty} \left[\frac{\sqrt{n} \overline{m}_n(\theta, u)}{\sqrt{\overline{\Sigma}_n(\theta, u)}} \right]_-^2.$$

A bootstrap critical value $c_{n,1-\alpha}^{AS}(\theta)$ is used to construct:

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

From corner values to an exact linear program

For each observation i and intersected refined cell ℓ , define

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

from the corner values.

Then

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

Maximum of finitely many linear functions of u .

This yields a linear epigraph reformulation:

$$\begin{aligned} \min_{u,t} \quad & \frac{1}{n} \sum_{i=1}^n t_i \quad \text{s.t.} \quad -1 \leq u \leq 1, \quad t_i \geq \tilde{u}_\ell m_{i\ell}^{\max}(\theta), \quad t_i \geq \tilde{u}_\ell m_{i\ell}^{\min}(\theta) \\ & \forall i, \ell \in S_i. \end{aligned}$$

The finite direction search is therefore one exact LP.

[Back](#)

Relation to Other Methods

Outer sets based on finitely many directions

The above also implies that

Every $\mathcal{U} \subset B_1^\infty$ will identify an **outer set!**

Aucejo, Bugni, and Hotz (2017) is a special case
using a restricted $\mathcal{U}$!

In particular, they use basis vectors and global directions. I.e. for $K(n) = 3$

$$\mathcal{U} = \begin{pmatrix} \pm 1 & \pm 1 & \pm 1 \\ \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & \pm 1 \end{pmatrix}$$

Interestingly, Aucejo, Bugni, and Hotz (2017) arrive there from a completely different approach.

[Back: Outer Sets](#)

[Back: Application](#)

Schennach (2014): Integrating Out the Latent Distribution

For known moments $g(\cdot)$ latent U and observed Z :

$$\underbrace{\inf_{\mu \in \mathcal{P}_{U|Z}} \|E_{\mu \times \pi_0}[g(U, Z, \theta)]\|}_{\infty\text{-dimensional over latent distrib.}} = 0 \quad \Longleftrightarrow \quad \underbrace{\inf_{\gamma \in \mathbb{R}^{d_g}} \|E_{\pi_0}[\tilde{g}(Z, \theta, \gamma)]\|}_{\text{finite-dimensional entropy dual}} = 0.$$

where

$$\tilde{g}(z, \theta, \gamma) = \frac{\int g(u, z, \theta) \exp(\gamma' g(u, z, \theta)) \, d\rho(u \mid z; \theta)}{\int \exp(\gamma' g(u, z, \theta)) \, d\rho(u \mid z; \theta)}.$$

The latent distribution search is replaced exactly by a finite-dimensional optimization over γ .

γ itself is ∞ -dim in cond. moment case.

Manski and Tamer (2002)

Manski and Tamer (2002) simplifies the problem a lot by introducing an **auxiliary assumption**:

$$\text{Mean independence } \mathbf{E}[Y|X, X_L, X_U] = \mathbf{E}[Y|X]$$

This links observables to unobservables. With this additional structure they obtain a simple solution estimator.

Implication:

If one observation is fully observed, the identified set is a singleton (point identification).

Interpreting Mean Independence

Assuming mean independence is rather strong¹. E.g. for missing data:
Let $Z = 0$ if X is observed, i.e. $X = X_L = X_U$ and $Z = 1$ otherwise.
Then

$$\begin{aligned}\mathbb{E}[Y|X = x] &= \mathbb{E}[Y|X = x, X_L = x, X_U = x] \\ &= \mathbb{E}[Y|X = x, X_L = X = X_U] \\ &= \mathbb{E}[Y|X = x, Z = 0]\end{aligned}$$

Assumes the data is missing at random.

[Back](#)

¹Manski and Tamer (2002) point out that it is equivalent to assume random censoring.

References

-  Andrews, D. and X. Shi (2017). “**Inference based on many conditional moment inequalities**”. In: *Journal of Econometrics* 196, pp. 275–287.
-  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.
-  Beresteanu, A. and F. Molchanov I. and Molinari (2012). “**Partial Identification Using Random Set Theory**”. In: *Journal of Econometrics* 166.1, pp. 17–32.
-  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.

-  Bretschneider, T. and Y. Kaddoura (2026). **“Factor-Based Imputation of Missing Values Using Cross-Sectional Averages”**. In.
-  Breunig, C. and P. Haan (2021). **“Nonparametric Regression with Selectively Missing Covariates”**. In: *Journal of Econometrics* 223.1, pp. 28–52.
-  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., H. Hong, and E. Tamer (2007). **“Parameter set inference in a class of econometric models”**. In: *Econometrica* 75.5, pp. 1243–1284.
-  Chesher, A. and A. Rosen (2017). **“Generalized Instrumental Variable Models”**. In: *Econometrica* 85.3, pp. 959–989.

-  Chesher, A. and A. Rosen (2020). “**Generalized Instrumental Variable Models, Methods, and Applications**”. In: *Handbook of Econometrics*. Vol. 7A. Elsevier, pp. 1–110.
-  Duan, J. and M. Pelger (2025). “**Imputation-powered inference for missing covariates**”. In: *NBER Working Paper*.
-  Fang, F., J. Zhao, and J. Shao (2018). “**Imputation-Based Adjusted Score Equations in Generalized Linear Models with Nonignorable Missing Covariate Values**”. In: *Statistica Sinica* 28.4, pp. 1677–1701.
-  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.
-  Horowitz, J. and C. Manski (2000). “**Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data**”. In: *Journal of the American Statistical Association*, pp. 77/84.

-  Hsiao, C. (1983). “**Regression Analysis with a Categorized Explanatory Variable**”. In: *Studies in Econometrics, Time Series and Multivariate Statistics*, pp. 93–129.
-  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.
-  Magnac, T. and E. Maurin (2008). “**Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data**”. In: *The Review of Economic Studies* 75, pp. 835–864.
-  Manski, C. and E. Tamer (2002). “**Inference on Regressions with Interval Data on a Regressor or Outcome**”. In: *Econometrica* 70, pp. 519–546.
-  Miao, W. and Tchetgen Tchetgen E. (2018). “**Identification and Inference with Nonignorable Missing Covariate Data**”. In: *Statistica Sinica* 28.4, pp. 2049–2067.
-  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.
-  Pacini, David (2018). “**The two-sample linear regression model with interval-censored covariates**”. In: *Journal of Applied Econometrics* 33, pp. 1–16.
-  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 (1994). “**Estimation of Regression Coefficients When Some Regressors Are Not Always Observed**”. In: *Journal of the American Statistical Association* 89.427, pp. 846–866.
-  Schennach, S. (2014). “**Entropic Latent Variable Integration via Simulation**”. In: *Econometrica* 82.1, pp. 345–385.



Stinchombe, M. and H. White (1998). “**Consistent Specification Testing with Nuisance Parameters Present Only Under The Alternative**”. In: *Econometric Theory* 14, pp. 295–325.



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.