Universal Inference for Incomplete Discrete Choice Models

Hiroaki Kaido Yi Zhang
Boston U. Jinan U.

August, 2024

Incomplete Models

Discrete choice models have been widely used.

Recent economic applications admit set-valued (or incomplete) predictions.

Incomplete Discrete Choice Models:

- Given exogenous variables (X, U), a set of outcome values is predicted for Y;

$$Y \in G(U|X;\theta),$$

It nests models with complete predictions: $Y = g(U|X;\theta)$.

Incomplete predictions

- Discrete games w/multiple equilibria (Tamer 03, Ciliberto/Tamer 09)
- Discrete choice w/heterogeneous choice sets (Barseghyan et al. 21)
- Discrete choice w/endogenous explanatory variables & IVs (Chesher/Rosen 17)
- Dynamic discrete choice models (Heckman 78; Honoré/Tamer 06; Torgovitsky 19; Chesher/Rosen/Zhang, 24)
- School choice models w/weak assump. on behavior (He 17; Agarwal/Somaini 20)
- Network formation models (de Paula/Richards-Shubik/Tamer 18; Sheng 20)
- Auctions (Haile/Tamer 03; Tamer et al. 18)

Inference Problem

This paper develops a procedure for testing composite hypotheses:

$$H_0 : \theta \in \Theta_0 \quad v.s. \quad H_1 : \theta \in \Theta_1,$$

e.g., $\Theta_0 = \{\theta : \varphi(\theta) = \varphi^*\}$ for some function $\varphi : \Theta \to \mathbb{R}$.

Challenges/Goals:

- Testing in the presence of incompleteness & nuisance parameters
- The asymptotic distributions of existing tests are often non-standard and require regularity conditions/tuning parameters
- We aim to provide a tractable method with finite-sample validity

Proposed test

Our proposal:

$$\phi_n = 1\left\{S_n > \frac{1}{\alpha}\right\}.$$

- S_n: A cross-fit version of a likelihood-ratio (LR) statistic $\mathcal{L}_0(\hat{\theta}_1)/\mathcal{L}_0(\hat{\theta}_0)$
- $\frac{1}{\alpha}$: fixed critical value

Universal validity:

For any n and over the class $\mathcal{P}_0^n$ of DGPs compatible with H_0:

$$\sup_{P^n \in \mathcal{P}_0^n} E_{P^n}\left[\phi_n\right] \leq \alpha.$$

Universal Inference

The test has the following properties.

(i) It has finite-sample validity without complex regularity conditions
(ii) It is tractable: no moment selection regularization, resampling, or simulations
(iii) It can yield confidence intervals for $\varphi(\theta)$ (e.g. counterfactual probabilities)
(iv) It can incorporate continuous and discrete covariates
(v) Nonparametric components are allowed (so long as we can calculate MLE)

Preliminary simulation results suggest it has nontrivial power in finite samples. Its power properties are comparable to those of existing tests in large samples.

Properties

We use a “tailor-made” likelihood using the model structure to ensure robustness to the model incompleteness

The finite sample validity builds on a sample-splitting argument and a Chernoff-style bound developed by Wassermann/Ramdas/Balakrishnan (20).

It can be viewed as a versatile (but not optimal) test that is effective in settings where

- $|\mathcal{Y}|$ is not too high
- models are rich enough to capture various heterogeneity through covariates
- θ contains nuisance components

Related Literature

We reference here only some closely related papers

- Identification in Incomplete Models:
  Jovanovic (89), Tamer (03), Galichon/Henry (11), Beresteanu/Molchanov/Molinari (11), Chesher/Rosen (17), Luo & Wang (18), Ponomarev (23)

- Inference in Partially Identified/Incomplete Models
  - Moment inequalities: Reviews by Canay/Shaikh (17), Molinari (20)
  - Likelihood-based:
    Chen/Tamer/Torgovitsky (11), Chen/Tamer/Christensen (18), Chen/Hansen/Hansen (21), Kaido/Molinari (24)
  - Monte Carlo-based: Li/Henry (23)

- Robust Statistics/Universal Inference:
  Huber/Stranssen (73,74), Wassermann/Ramdas/Balakrishnan (20)

Set-up

Set-up

- $Y \in \mathcal{Y}$: outcome, $\mathcal{Y}$: finite set
- $X \in \mathcal{X} \subseteq \mathbb{R}^{d_X}$: covariates
- $U \in \mathcal{U} \subseteq \mathbb{R}^{d_U}$: latent variables
- $F_\theta(\cdot|x)$: conditional law of $U|X$, which belongs to $F = \{F_\theta, \theta \in \Theta\}$.

The model's prediction is summarized by $G(\cdot|\cdot;\theta) : \mathcal{U} \times \mathcal{X} \twoheadrightarrow \mathcal{Y}$, a weakly measurable correspondence.

A sample $\{(Y_i, X_i), i = 1, \ldots, n\}$ is drawn. We assume independence of observations across i.

Examples

Entry game (Bresnahan & Reiss, 91, Ciliberto & Tamer 09)
Each player $j \in \{1, 2\}$ can choose to enter ($y^{(j)} = 1$) or to stay out of the market ($y^{(j)} = 0$).

The players' payoffs are

$$\pi_j = Y^{(j)}(X^{(j)}\delta^{(j)} + \beta^{(j)} Y^{(3-j)} + U^{(j)}), \quad j = 1, 2,$$

which is common knowledge.

- $\mathcal{Y} = \{(0,0), (0,1), (1,0), (1,1)\}$
- $X = (X^{(1)}, X^{(2)})$: Observable payoff shifters
- $U = (U^{(1)}, U^{(2)})$: Unobservable payoff shifters

Suppose a pure strategy Nash equilibrium (PSNE) is played.

Example: 2 Player Entry Game

		player 2	
		$Y_2 = 0$	$Y_2 = 1$
player 1	$Y_1 = 0$	$(0, 0)$	$\left(0, X^{(2)}\delta^{(2)} + U^{(2)}\right)$
	$Y_1 = 1$	$\left(X^{(1)}\delta^{(1)} + U^{(1)}, 0\right)$	$\left(X^{(1)}\delta^{(1)} + \beta^{(1)} + U^{(1)}, X^{(2)}\delta^{(2)} + \beta^{(2)} + U^{(2)}\right)$

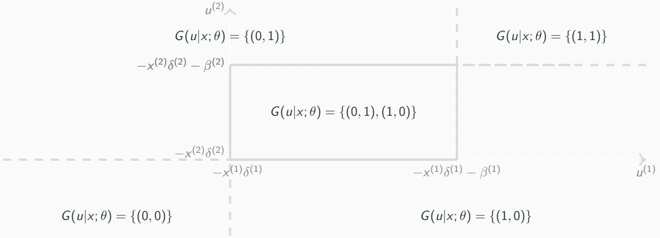
If $\beta_j < 0, j = 1, 2$

$$G(u|x;\theta) = \begin{cases} \{(0,0)\} & u \in a_\theta(x) \\ \{(0,1)\} & u \in b_\theta(x) \\ \{(1,0)\} & u \in c_\theta(x) \\ \{(1,1)\} & u \in d_\theta(x) \\ \{(1,0),(0,1)\} & u \in e_\theta(x) \end{cases}$$

Example: 2 Player Entry Game

		player 2	
		$Y_2 = 0$	$Y_2 = 1$
player 1	$Y_1 = 0$	$(0, 0)$	$\left(0, X^{(2)}\delta^{(2)} + U^{(2)}\right)$
	$Y_1 = 1$	$\left(X^{(1)}\delta^{(1)} + U^{(1)}, 0\right)$	$\left(X^{(1)}\delta^{(1)} + \beta^{(1)} + U^{(1)}, X^{(2)}\delta^{(2)} + \beta^{(2)} + U^{(2)}\right)$

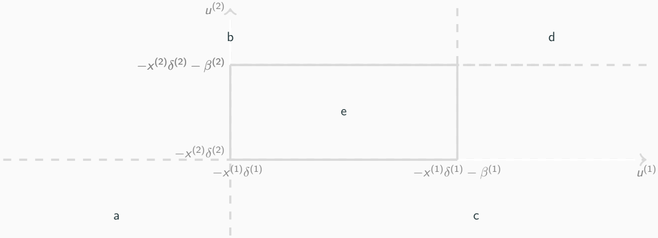
If $\beta_j < 0, j = 1, 2$



Example: 2 Player Entry Game

		player 2	
		$Y_2 = 0$	$Y_2 = 1$
player 1	$Y_1 = 0$	$(0, 0)$	$\left(0, X^{(2)}\delta^{(2)} + U^{(2)}\right)$
	$Y_1 = 1$	$\left(X^{(1)}\delta^{(1)} + U^{(1)}, 0\right)$	$\left(X^{(1)}\delta^{(1)} + \beta^{(1)} + U^{(1)}, X^{(2)}\delta^{(2)} + \beta^{(2)} + U^{(2)}\right)$

If $\beta_j < 0, j = 1, 2$



Procedure

Overview

Artstein's inequality:

Let $\mathcal{C}$ be the set of all closed subsets of $\mathcal{Y}$.

$$Y \in G(U|x;\theta),\ a.s. \quad \Leftrightarrow \quad \underbrace{P(A|x) \geq \nu_\theta(A|x)}_{\text{Sharp Identifying Restrictions}}, \ \forall A \in \mathcal{C}.$$

- The containment functional $\nu_\theta(A|x) \equiv \int 1\{G(u|x;\theta) \subseteq A\} dF_\theta(u|x)$ determines the distribution of G (Molchanov, 2017).

The set of conditional densities

Let

$$\mathfrak{q}_{\theta,x} \equiv \{q(\cdot|x) : \sum_{y \in A} q(y|x) \geq \nu_\theta(A|x),\ A \in \mathcal{C}\}.$$

- The conditional density of Y is restricted by linear inequalities.

Ex. Entry game
Any density in $\mathfrak{q}_{\theta,x}$ satisfies

$$q((0,0)|x) = F_\theta(a_\theta(x)|x)$$
$$q((1,1)|x) = F_\theta(d_\theta(x)|x)$$
$$q((1,0)|x) \leq F_\theta(c_\theta(x)|x) + F_\theta(e_\theta(x)|x)$$
$$q((1,0)|x) \geq F_\theta(c_\theta(x)|x)$$

Procedure

Split a sample $(Y_i, X_i), i = 1, \ldots, n$ into D_1 and D_0.

Procedure

Split a sample $(Y_i, X_i), i = 1, \ldots, n$ into D_1 and D_0.

1. **From D_1,**
   - Compute $\hat{\theta}_1$: any estimator of θ;
     (e.g., a minimizer of a criterion function)
   - Find $p(\cdot|x) \in \mathfrak{q}_{\hat{\theta}_1, x}$.

Procedure

Split a sample $(Y_i, X_i), i = 1, \ldots, n$ into D_1 and D_0.

1. From D_1,

- Compute $\hat{\theta}_1$: any estimator of θ;
  (e.g., a minimizer of a criterion function)

- Find $p(\cdot|x) \in \mathfrak{q}_{\hat{\theta}_1,x}$.

2. From D_0,

- Compute $\hat{\theta}_0$: *RMLE*

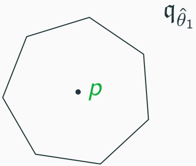
$$\hat{\theta}_0 \in \arg\max_{\theta \in \Theta_0} \prod_{i \in D_0} q_\theta(Y_i|X_i)$$

using

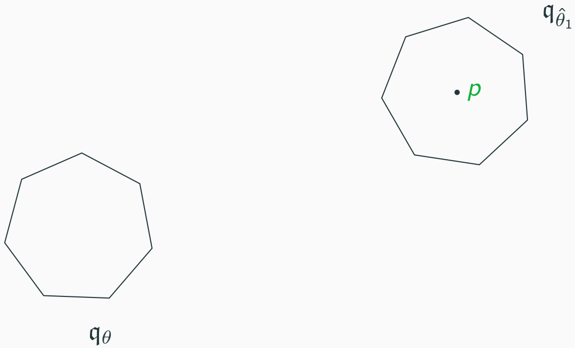
$$q_\theta(\cdot|x) = \arg\min_{q(\cdot|x) \in \mathfrak{q}_{\theta,x}} I_{KL}(q(\cdot|x) + p(\cdot|x) || q(\cdot|x)).$$

I_{KL}: Kullback-Leibler divergence

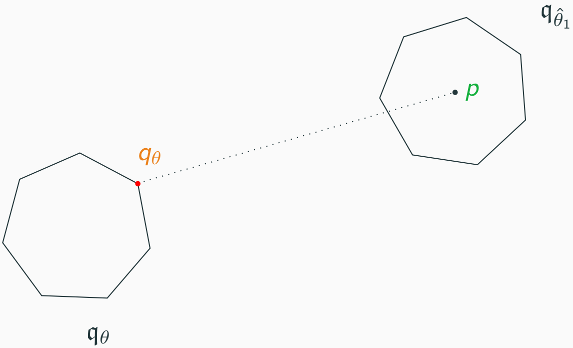
Intuition



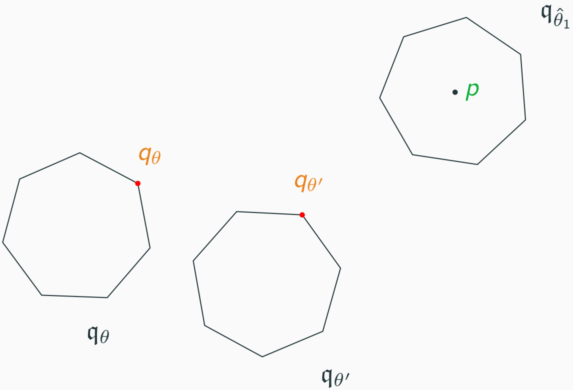
Intuition



Intuition

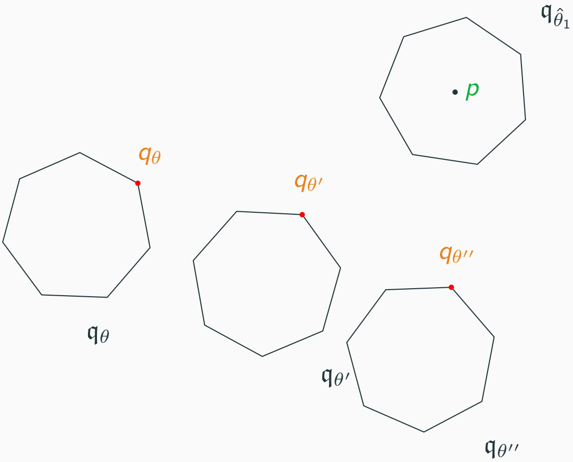


Intuition



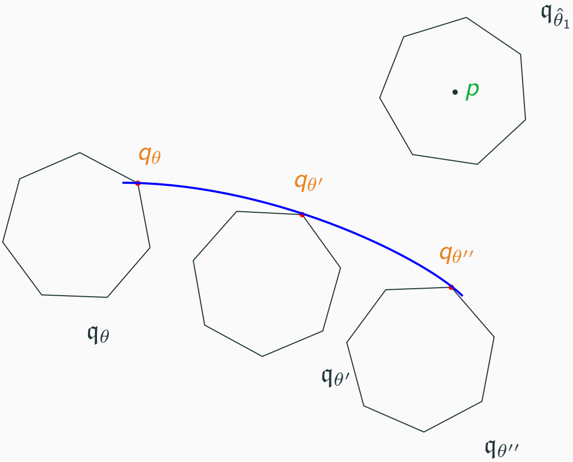
$\mathfrak{q}_{\hat{\theta}_1}$
p
q_θ
$q_{\theta'}$
$\mathfrak{q}_\theta$
$\mathfrak{q}_{\theta'}$

Intuition



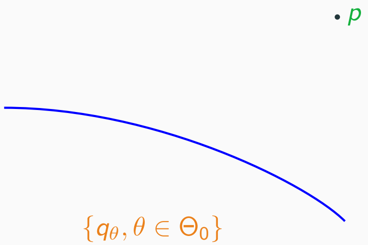
$\mathfrak{q}_{\hat{\theta}_1}$
p
q_θ
$q_{\theta'}$
$q_{\theta''}$
$\mathfrak{q}_\theta$
$\mathfrak{q}_{\theta'}$
$\mathfrak{q}_{\theta''}$

Intuition

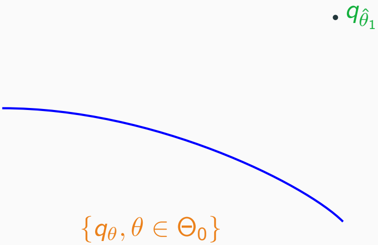


$\mathfrak{q}_{\hat{\theta}_1}$
p
q_θ
$q_{\theta'}$
$q_{\theta''}$
$\mathfrak{q}_\theta$
$\mathfrak{q}_{\theta'}$
$\mathfrak{q}_{\theta''}$

Intuition



Intuition



Procedure:

Conditional on D_1, we construct a parametric model:

$$\{q_\theta(\cdot|x), \theta \in \Theta_0 \cup \{\hat{\theta}_1\}\}$$

- $q_{\hat{\theta}_1}$: a density in the unrestricted model;
- $\{q_\theta, \theta \in \Theta_0\}$: a collection of least-favorable densities under H_0.

Any incomplete model admits the existence of such a parametric model.

This is due to ν_θ belonging to a class of 2-monotone capacities and results from the robust statistics lit. (Huber/Strassen 73, Kaido/Zhang, 19).

Cross-fitting

3. Compute S_n

Split LR statistic:

$$T_n = \frac{\prod_{i \in D_0} q_{\hat{\theta}_1}(Y_i|X_i)}{\prod_{i \in D_0} q_{\hat{\theta}_0}(Y_i|X_i)} = \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)}$$

Cross-fit LR statistic:

$$S_n = \frac{T_n + T_n^{\text{swap}}}{2}$$

T_n^{swap} is calculated in the same way as T_n after swapping the roles of D_0 and D_1.

4. Reject H_0 if $S_n > 1/\alpha$.

LFP-based density

Ex. Entry game

$\mathcal{Y} = \{(0,0),(0,1),(1,0),(1,1)\}$.

Any density in $\mathfrak{q}_{\theta,x}$ satisfies

$$q((0,0)|x) = F_\theta(a_\theta(x)|x) \tag{1}$$
$$q((1,1)|x) = F_\theta(d_\theta(x)|x) \tag{2}$$
$$q((1,0)|x) \leq F_\theta(c_\theta(x)|x) + F_\theta(e_\theta(x)|x) \tag{3}$$
$$q((1,0)|x) \geq F_\theta(c_\theta(x)|x) \tag{4}$$

LFP-based density

More details on Step 2.

Ex. Entry game

Minimizing I_{KL} is equivalent to solving

$$\min_{q(\cdot|x)\in\Delta^{\mathcal{Y}}} \sum_{y\in\mathcal{Y}} \ln\left(\frac{q(y|x)+p(y|x)}{q(y|x)}\right)(q(y|x)+p(y|x))$$

$$s.t. \quad q \text{ satisfies (1)-(4)}$$

Can obtain q_θ in closed form.

Properties

Size Control

Suppose $(Y_i, X_i, U_i)_{i=1}^n$ are independently distributed, and $(X_i, U_i)_{i=1}^n$ are identically distributed across i.

Let

$$\mathcal{P}_\theta^n = \Big\{ P^n = \bigotimes_{i=1}^n P_i,\ P_i(A|x) \geq \nu_\theta(A|x),\ \forall A \in \mathcal{C}, x \in \mathcal{X}, \\ P_{i,X} = P_X,\ P_X \in \Delta(\mathcal{X}) \Big\}.$$

Theorem:
For any $n \in \mathbb{N}$,

$$\sup_{P^n \in \mathcal{P}_0^n} P^n\big(S_n > \frac{1}{\alpha}\big) \leq \alpha$$

where $\mathcal{P}_0^n = \{P^n \in \mathcal{P}_\theta^n : \theta \in \Theta_0\}$.

Confidence intervals

Let $\varphi : \Theta \to \mathbb{R}$:

- Counterfactual probability:
  $\varphi(\theta) = F_\theta(\{u : x_1{}'\beta_1 + \Delta_1 \geq -U_1\}) = \Phi(x_1{}'\beta_1 + \Delta_1)$
  e.g., probability of Player 1 entering when y_2 is set to 1;
  $U_1 \sim N(0, 1)$

- A component of θ or a linear combination of θ:
  $\varphi(\theta) = \theta_1, \;\; \varphi(\theta) = l'\theta.$

- (Average, distributional, quantile) structural function:
  $\varphi(\theta) = E_{F_\theta}[\mu(d, X, U; \theta)]$ (e.g., ASF)

- Treatment effects:
  $\varphi(\theta) = E_{F_\theta}[\mu(1, X, U; \theta) - \mu(0, X, U; \theta)]$ (e.g., ATE)

Confidence intervals

Let $\Theta_0(\varphi^*) \equiv \{\theta \in \Theta : \varphi(\theta) = \varphi^*\}$, and let $S_n(\varphi^*)$ be the corresponding cross-fit LR statistic.

A confidence interval for $\varphi(\theta)$ is

$$CI_n \equiv \left\{\varphi^* \in \mathbb{R} : S_n(\varphi^*) \leq \frac{1}{\alpha}\right\}.$$

Comments:

- The confidence interval covers $\varphi(\theta)$ with prob. at least $1 - \alpha$ in any finite sample.
- Only the denominator of T_n (or T_n^{swap}) needs to be re-calculated across φ^*.

Empirical Illustration

Empirical Illustration (In Progress)

Thornton (08): "The demand for, and impact of, learning HIV status"

Respondents in rural Malawi were offered a free door-to-door HIV test and were given randomly assigned vouchers (up to $3), redeemable upon obtaining their results at a nearby test center.

The study examined the effect of learning the HIV status on condom purchases.

- Y: Condom purchase (0, 3, or 6 in this exercise);
- D: Learning HIV test result (0 or 1);
- X: HIV status, other individual characteristics;
- Z: Voucher amount, distance to test center.

$n \simeq 1000$, 10-15 parameters, continuous/discrete covariates

Empirical Illustration

A model of ordered choice:

$$Y = \begin{cases} 0 & \text{if } \mu(D, X) + U \leq c_L \\ 3 & \text{if } c_L < \mu(D, X) + U \leq c_U \\ 6 & \text{if } \mu(D, X) + U > c_U. \end{cases}$$

$$D = 1\{V'Z + \psi(X, Z)'\delta \geq 0\}.$$

We work with a control function (CF) assumption (Han & Kaido, 2024)

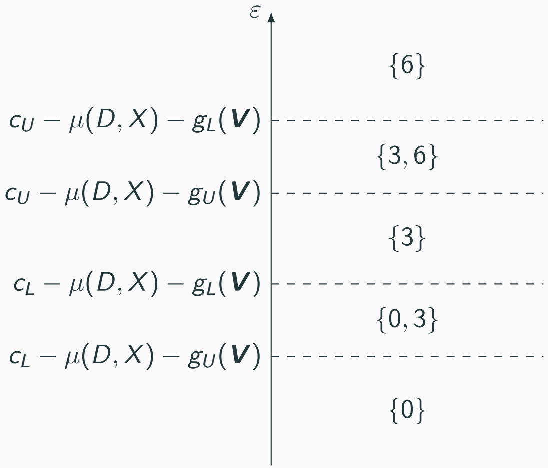
- $U|D, X, V \sim U|X, V$ for some CF $V \in \boldsymbol{V}$, where

$$\boldsymbol{V} = \begin{cases} \{v : v'z + \psi(x, z)'\delta \geq 0\} & \text{if } d = 1 \\ \{v : v'z + \psi(x, z)'\delta < 0\} & \text{if } d = 0. \end{cases}$$

This model yields a set-valued prediction. Right now, we use a fixed-coefficient model with an additive heterogeneity term v.

Prediction

$Y \in G(\epsilon | X; \theta),\ \ U = g(V) + \epsilon$



$g_L(\boldsymbol{V}) = \inf_{V \in \mathrm{Sel}(\boldsymbol{V})} g(V),\ g_U(\boldsymbol{V}) = \sup_{V \in \mathrm{Sel}(\boldsymbol{V})} g(V)$

Target objects

The model is characterized by 4 inequality restrictions and yields a closed-form LFP-based density q_θ.

We consider two objects

- Average Structural Function

$$\varphi(\theta) = \text{ASF}(d, x_{HIV})$$

- Policy Relevant Structural Function
  Consider giving the maximum voucher amount (z^*_{amt} =\$3) to everyone to encourage them to learn about their HIV status.

$$\varphi(\theta) = \text{PRSF}(\$3, x_{HIV})$$

Confidence Intervals

	HIV+	HIV-
ASF(1)	[0.090,5.216]	[0.332, 2.623]
ASF(0)	[0.030,4.251]	[0.211, 2.985]

	HIV+	HIV-
PRSF($3)	[0.091, 4.824]	[0.331, 2.502]

Findings

- We cannot determine the sign of the ATE. Thornton also finds very mild effects (2SLS estimands).
- The ASFs may be heterogeneous across different HIV status groups.
- One can attain effects similar to setting $d = 1$ by providing the financial incentive ($3).

Summary

This paper explores a universal inference method that has finite-sample validity for incomplete models.

- We use a "tailor-made" likelihood using the model structure to ensure robustness to the model incompleteness
- It can be viewed as a versatile (but not optimal) test that remains effective in settings with nuisance parameters, continuous and discrete covariates, and limited sample sizes.
- Preliminary codes are available at `https://github.com/hkaido0718/IncompleteDiscreteChoice`.

Thank you!

Monte Carlo Experiments

Simulation designs

Design 1:

Entry game with

$$\pi^{(j)} = y^{(j)}\big(\theta^{(j)} y^{(-j)} + u^{(j)}\big),\ j = 1, 2$$

with $(u^{(1)}, u^{(2)}) \sim N(0, I_2)$.

- $H_0 : \theta^{(j)} = 0$ for both players against $H_1 : \theta^{(j)} < 0$ for some j.
- We set $\theta^{(j)} = -h, j = 1, 2,\ h \geq 0$ to evaluate power. $Y_i = (1, 0)$ is selected w/ prob 0.5.

Tests:

1. Cross-fit LR test with $\hat{\theta}_1$ maximizing a likelihood-based criterion function
2. Cross-fit LR test with $\hat{\theta}_1$ minimizing a moment-based criterion function.

Design 1

Table 1: Size and Power of the Cross-Fit Tests for testing $H_0 : \theta^{(j)} = 0, j = 1, 2$

	Size	Power (values of h below)													
		0.069	0.138	0.207	0.276	0.345	0.414	0.483	0.552	0.621	0.690	0.759	0.828	0.897	0.966
Panel A: ($n = 100$)															
LR-test (MLE $\hat{\theta}_1$)	0	0.001	0.011	0.073	0.196	0.370	0.576	0.760	0.885	0.948	0.973	0.992	0.996	1.000	1.000
LR-test (CHT $\hat{\theta}_1$)	0	0.002	0.016	0.081	0.209	0.383	0.582	0.762	0.877	0.942	0.976	0.988	0.993	0.998	1.000
Panel B: ($n = 200$)															
LR-test (MLE $\hat{\theta}_1$)	0.001	0.007	0.060	0.235	0.522	0.794	0.948	0.988	0.997	1	1	1	1	1	1
LR-test (CHT $\hat{\theta}_1$)	0.002	0.008	0.066	0.246	0.521	0.776	0.940	0.988	0.996	1	1	1	1	1	1

Monte Carlo experiments

Design 2:

Similar to Design 1, but

$$\pi^{(j)} = y^{(j)}\big(x^{(j)\prime}\delta^{(j)} + \beta^{(j)}y^{(-j)} + u^{(j)}\big)$$

- $x^{(j)} \in \{-2, -1, 0, 1, 2\}$: player specific covariates
- Test $H_0 : \delta^{(j)} = 0, j = 1, 2$ against $H_1 : \delta^{(j)} \neq 0$ for some j.

Tests:

1. Cross-fit LR test (for $n \in \{50, 100, 200, 300, 5000, 7500\}$)
2. Moment-based test by Bugni, Canay, and Shi (17) (for $n = 5000, 7500$).

Design 2

n	Size	Power (values of h below)													
		0.105	0.211	0.316	0.421	0.526	0.632	0.737	0.842	0.947	1.053	1.158	1.263	1.368	1.474
50	0	0.000	0.005	0.030	0.100	0.222	0.354	0.493	0.612	0.732	0.834	0.894	0.940	0.972	0.982
100	0	0.000	0.011	0.070	0.194	0.375	0.499	0.631	0.748	0.864	0.926	0.963	0.978	0.989	0.998
200	0	0.000	0.056	0.252	0.504	0.665	0.790	0.867	0.929	0.965	0.981	0.987	0.991	0.994	0.997
300	0	0.006	0.158	0.558	0.809	0.912	0.954	0.973	0.989	0.998	0.996	0.996	0.998	0.997	0.997

Note: The size and power are calculated based on $S = 1000$ simulations. DGP with covariates taking 25 different values. The average sample size in each bin is at most 12.

Design 2

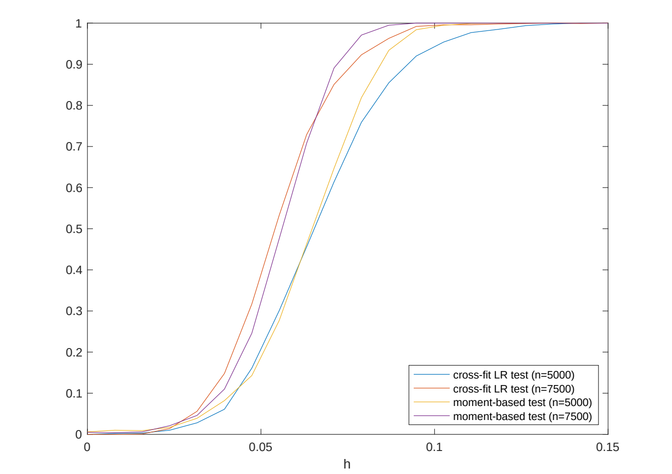


Figure 1: Power of the Cross-fit LR and Moment-based Tests: ($n \in \{5000, 7500\}$, $S = 1000$ replications)

Design 2

Cross-fit LR test	Moment-based test		
	4 cores	8 cores	16 cores
13.75	111.65	56.64	41.84

Table 2: Median computation time (in seconds)

Note: The median computation time is calculated based on $S = 1000$ simulations for the cross-fit LR test. For the moment-based test, we parallelized bootstrap replications ($B = 500$) with 4, 8, and 16 cores. The median computation time is calculated based on $S = 100$ simulation repetitions.

Appendix

Comparison

Split LRT:

$$2\ln\left(\frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)}\right) > c_\alpha$$

- $c_\alpha = 2\ln(1/\alpha)$
- $\mathcal{L}_0(\hat{\theta}_1)$: out-of-sample likelihood
- $\mathcal{L}_0(\theta) = \prod_{i\in D_0} q_\theta(Z_i)$

$\rightarrow$ c_α is due to Markov's inequality applied to $\exp(t\ln\left(\frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)}\right))$ with $t = 1$ (poorman's Chernoff bound)

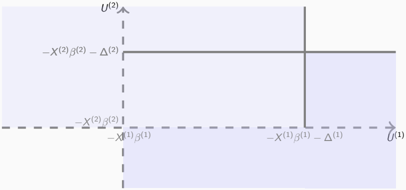
Standard LRT:

$$2\ln\left(\frac{\mathcal{L}_{full}(\hat{\theta}_1)}{\mathcal{L}_{full}(\hat{\theta}_0)}\right) > c_{d,\alpha}$$

- $c_{d,\alpha}$: $1-\alpha$ quantile of $\chi^2_{d,\alpha}$
- $\mathcal{L}_{full}(\hat{\theta}_1)$: full-sample likelihood
- $\mathcal{L}_{full}(\theta) = \prod_{i=1}^n q_\theta(Z_i)$

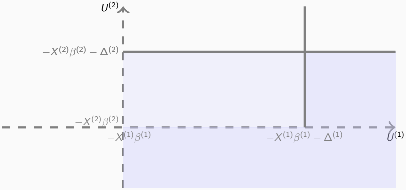
$\rightarrow$ $c_{d,\alpha}$ is due to Wilks' thereom.

LFP-based density



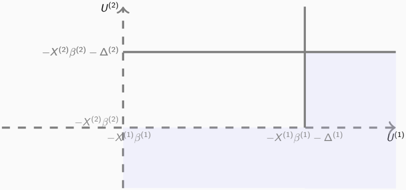
$$\eta_1(x;\theta) = 1 - F_\theta(a|x) - F_\theta(d|x)$$

LFP-based density



$$\eta_2(x;\theta) = F_\theta(c|x) + F_\theta(e|x)$$

LFP-based density



$$\eta_3(x;\theta) = F_\theta(c|x)$$

Examples

Panel Dynamic Discrete Choice (Heckman, 78, Honore & Tamer, 09)

Binary decisions across multiple periods, according to

$$Y_{it} = 1\{X_{it}'\lambda + Y_{it-1}\beta + \alpha_i + \epsilon_{it} \geq 0\},\ i = 1, \ldots, n,\ t = 1, \ldots, T.$$

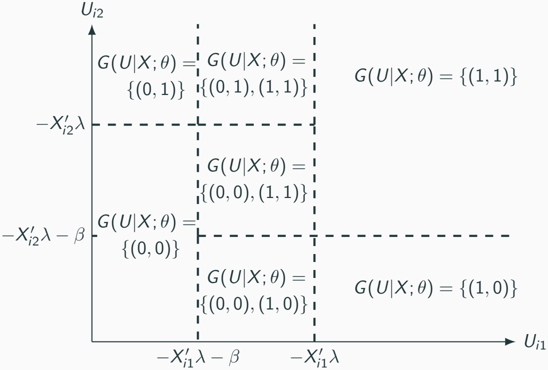
The model admits state dependence through $Y_{it-1}\beta$.

The initial value Y_{i0} is not observed.

Multiple outcome sequences $(Y_{i1}, \ldots, Y_{iT}) \in \{0, 1\}^T$ can satisfy the model restrictions.

Example: Panel Dynamic Discrete Choice ($T = 2$)

If $\beta > 0$,



Implementation

Option 1 (Test inversion)

For each φ^* in a grid over $\varphi(\Theta)$, compute $S_n(\varphi^*)$ and keep the ones that pass the test.

Option 2 (Bayesian optimization/Response-surface method)

The endpoints of CS_n are

$$\min / \max \varphi^* \quad s.t.\ S_n(\varphi^*) \leq \frac{1}{\alpha}$$

where $\varphi^* \mapsto S_n(\varphi^*)$ is a black-box function.

Can use global optimization algorithms for solving a problem w/ black-box constraints (Jones, et. al., 98).

Core-determining class

A sub-collection $\mathcal{A}$ of $\mathcal{C}$ is *core-determining* (Galichon & Henry, 11) if

$$\begin{aligned}\mathfrak{q}_{\theta,x} &= \{q(\cdot|x) : \sum_{y \in A} q(y|x) \geq \nu_\theta(A|x),\ A \in \mathcal{C}\} \\ &= \{q(\cdot|x) : \sum_{y \in A} q(y|x) \geq \nu_\theta(A|x),\ A \in \mathcal{A}\}\end{aligned}$$

There is a minimal core determining class $\mathcal{A}^*$, which is determined by the graph representation of G (Luo & Wang, 18, Ponomarev 23).

- For applications with relatively high $|\mathcal{Y}|$, we recommend reducing the number of constraints to a manageable size.

Confidence intervals

Corollary:

For any n,

$$\inf_{(\varphi^*, P^n) \in \mathcal{F}^n} P^n\big(\varphi^* \in CS_n\big) \geq 1 - \alpha,$$

where $\mathcal{F}^n = \big\{(\varphi^*, P^n) : \varphi(\theta) = \varphi^*, P^n \in \mathcal{P}^n_\theta, \text{ for some } \theta \in \Theta\big\}$.

- The sharp identification region for $\varphi(\theta)$ under P^n is $\mathcal{H}_{P^n}[\varphi] = \{\varphi^* : \varphi(\theta) = \varphi^*, P^n \in \mathcal{P}^n_\theta, \text{ for some } \theta \in \Theta\}$.
- The result above ensures that CS_n covers elements of $\mathcal{H}_{P^n}[\varphi]$ across P^n.

Getting $\hat{\theta}_1$

A natural choice $\hat{\theta}_1$ is an extremum estimator that minimizes a sample criterion function $\theta \mapsto \hat{\mathsf{Q}}_1(\theta)$.

Examples:

- $\hat{\mathsf{Q}}_1(\theta) = \sup_{j,x} \frac{\{\nu_\theta(A_j|x) - \hat{P}_1(A_j|x)\}_+}{\hat{s}_{\theta,1}(A_j|x)}$, where $\hat{s}_{\theta,1}(A_j|x)$ is an estimator of the standard error of $\hat{P}_1(\cdot|x)$ (Chernozhukov, et. al., 07, 13)

- $\hat{\mathsf{Q}}_1(\theta) = \sum_{i \in D_1} - \ln p_\theta(Y_i|X_i; \hat{p}_n)$, where p_θ is the KLIC projection of a nonparametric estimator $\hat{p}_n$ of $p_0(\cdot|x) = P_0(Y = \cdot|x)$ (Kaido/Molinari, 24)

Finding p

One can obtain $p(\cdot|x)$ in $\mathfrak{q}_{\theta,x}$ by solving a linear feasibility problem.

$$\text{Find } p(\cdot|x) \in \Delta^{\mathcal{Y}}$$
$$s.t. \sum_{y \in A} p(y|x) \geq \nu_{\hat{\theta}_1}(A|x), \; A \in \mathcal{C}.$$

Any solution $p(\cdot|x)$ (positive density) can be used.

Proof sketch

Suppose, for the moment, the model is complete so that $\mathfrak{q}_\theta = \{q_\theta\}$. We proceed with Wasserman et. al.'s argument.

Let θ^* be the true value. It is straightforward to show $E_{q_{\theta^*}}\left[\frac{\mathcal{L}_0(\theta')}{\mathcal{L}_0(\theta^*)}\right] \leq 1$ for any $\theta' \in \Theta$.

By Markov's inequality

$$P_{\theta^*}(T_n > \frac{1}{\alpha}) \leq \alpha E_{q_{\theta^*}}\left[\frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)}\right] \leq \alpha E_{q_{\theta^*}}\left[\frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\theta^*)}\right].$$

Conditioning on D_1,

$$\alpha E_{q_{\theta^*}}\left[\frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\theta^*)}\right] = \alpha E_{q_{\theta^*}}\left(E_{q_{\theta^*}}\left[\frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\theta^*)}\middle| D_1\right]\right) \leq \alpha.$$

We extend this argument by replacing q_θ with the LFP-based parametric model.

Proof sketch

Let $\theta_1 \in \Theta$ and let $Q_{\theta_1} \in \mathcal{P}_{\theta_1}$.

For each $\theta \in \Theta_0$, consider distinguishing $\mathcal{P}_{\theta_1}$ against a singleton set $\{Q_{\theta_1}\}$.

There exists a least-favorable pair (LFP) $(Q_\theta, Q_{\theta_1}) \in \mathcal{P}_\theta \times \{Q_{\theta_1}\}$ such that for all $t \in \mathbb{R}$

$$\sup_{P \in \mathcal{P}_\theta} P(\Lambda > t) = Q_\theta(\Lambda > t)$$

$$\inf_{P \in \mathcal{P}_{\theta_1}} P(\Lambda > t) = Q_{\theta_1}(\Lambda > t),$$

where $\Lambda = dQ_{\theta_1}/dQ_\theta$ (due to $\mathcal{P}_\theta = core(\nu_\theta)$, Huber & Stranssen, 73).

Proof sketch

$$P^n(T_n(\theta_1) > \frac{1}{\alpha}) \leq \alpha E_{P^n}\Big[\frac{\mathcal{L}_0(\theta_1)}{\mathcal{L}_0(\hat{\theta}_0)}\Big] \leq \alpha \sup_{\tilde{P}^n \in \mathcal{P}_\theta^n} E_{\tilde{P}^n}\Big[\frac{\mathcal{L}_0(\theta_1)}{\mathcal{L}_0(\theta)}\Big],$$

The supremum is attained by the product of the least-favorable distribution Q_θ at θ.

Using this,

$$E_{Q_\theta^n}\Big[\frac{\mathcal{L}_0(\theta_1)}{\mathcal{L}_0(\theta)}\Big] \leq 1.$$

The rest of the argument is similar to the complete case.

Critical values