Econometrics of Incomplete Models

Shuowen Chen

Fall 2020: BU Econometrics Reading Group

Outline

- Notion of incompleteness
- One type of incompleteness: multiple equilibria
  - Complete information entry game
- Challenges due to incompleteness
  - Likelihood–based estimation
  - Assumptions to complete the model affects inference
  - Assumptions on data sampling affect inference
- Work in progress: robust score test
  - Inference on existence and sign of interaction effect
  - Robust to not knowing which equilibrium is played
  - Why not likelihood ratio test? Nuisance parameters

Model Coherence and Completeness (Lewbel, 2019)

Consider a proposed model of the form $Y = H(Y, V)$

- Y: a vector of endogenous outcomes (prices, agent choices, etc.)
- V: a set of (un)observables that determine outcomes (parameters of interest, exogenous covariates, error terms, etc.)

The model is **coherent** if

- $\forall v \in \Omega_V, \exists y \in \Omega_Y$ s.t. $y = H(y, v)$

The model is **complete** if

- $\forall v \in \Omega_V$, $\exists$ *at most one* $y \in \Omega_Y$ s.t. $y = H(y, v)$

Model Incoherence and Incompleteness

The **reduced form** of the model expresses Y solely in terms of V

$$y = G(v)$$

Remarks:

- Coherenence and completeness feature a unique reduced form

$$G(v) = H(G(v), v)$$

- An incoherent model has no solution for some values of v
  - A game with no Nash Equilibrium
- A coherent and incomplete model has multiple solutions for some values of v
  - A game with multiple Nash Equilibria
  - Reduced form $G(\cdot)$ is not unique

Outline

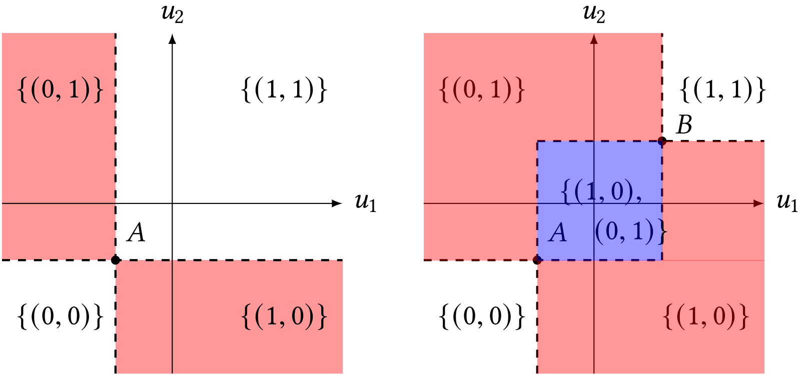
- One type of incompleteness: multiple equilibria
  - Complete information entry game

Binary Complete Info Static Entry–Exit Game

	Out	In
Out	$0, 0$	$0, x^{(2)\prime}\delta^{(2)} + u^{(2)}$
In	$x^{(1)\prime}\delta^{(1)} + u^{(1)}, 0$	$x^{(1)\prime}\delta^{(1)} + \beta^{(1)} + u^{(1)}, x^{(2)\prime}\delta^{(2)} + \beta^{(2)} + u^{(2)}$

- Competition effect: $\beta^{(1)} < 0, \beta^{(2)} < 0$
- $u := (u^{(1)}, u^{(2)}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$
  - Complete info: realizations are perfectly observed by both players
- $x^{(i)}$: exogenous covariates
- $\delta^{(i)}$: nuisance parameters
- **Inference:** $H_0 : \beta^{(i)} = 0,\ \delta \in \Theta_\delta$ vs $H_1 : \beta^{(i)} < 0,\ \delta \in \Theta_\delta$
- Solution concept: pure strategy Nash Equilibrium
  1. $(0, 0)$ is a NE when $u^{(i)} < -x^{(i)\prime}\delta^{(i)}$
  2. $(1, 1)$ is a NE when $u^{(i)} > -x^{(i)\prime}\delta^{(i)} - \beta^{(i)}$
  3. $(1, 0)$ is a NE when $u^{(1)} > -x^{(1)\prime}\delta^{(1)}$ and $u^{(2)} < -x^{(2)\prime}\delta^{(2)} - \beta^{(2)}$
  4. $(0, 1)$ is a NE when $u^{(2)} > -x^{(2)\prime}\delta^{(2)}$ and $u^{(1)} < -x^{(1)\prime}\delta^{(1)} - \beta^{(1)}$
- (3) and (4) intersects: $-x^{(i)\prime}\delta^{(i)} < u^{(i)} < -x^{(i)\prime}\delta^{(i)} - \beta^{(i)}$

Visualization of the Equilibria



- $A \equiv (-x^{(1)\prime}\delta^{(1)}, -x^{(2)\prime}\delta^{(2)})$,
  $B \equiv (-x^{(1)\prime}\delta^{(1)} - \beta^{(1)}, -x^{(2)\prime}\delta^{(2)} - \beta^{(2)})$
- **Incompleteness:** relationship from u, X, β and δ to y is a correspondence rather than a function (Tamer, 2003)
- Complete if either $\beta^{(1)} = \beta^{(2)} = 0$, or an equilibrium selection mechanism is imposed in the blue region

Model–predicted Distributions of Outcomes

$$\mathcal{P}_\theta = \left\{ P \in \Delta(S) : P = \int_U P_u dm_\theta(u), \text{ for some } P_u \in \Delta(G(u \mid \theta; X)) \right\}$$

- $\theta := (\beta, \delta)$
- S: set of potential outcomes $\{(0,0), (0,1), (1,0), (1,1)\}$
- $\Delta(\cdot)$: probabilty simplex
- P_u : equilibrium selection mechanism
- $m_\theta(\cdot)$: probability measures on U
- $G(u \mid \theta; X)$: set of model–predicted outcomes

Remarks:

- $G(u \mid \theta; X)$: math expression of graphs in the previous slide
- $\beta^{(1)} = \beta^{(2)} = 0$: unique distribution, set of distributions otherwise
- δ can enter nonlinearly

Outline

- Challenges due to incompleteness
  - Likelihood–based estimation
  - Assumptions to complete the model affects inference
  - Assumptions on data sampling affect inference

Data: A Cross Section of Markets

Consider a sequences of observed outcomes and latent variables

$$s^n = (s_1, \ldots, s_n), \quad u^n = (u_1, \ldots, u_n)$$

Assumption: For each $\theta \in \Theta$, $m_\theta^n \in \Delta(U^n)$ is a product measure: u_i's are i.i.d across markets

- Takes values in Cartesian product of sets of permissible outcomes

$$s^n \in G^n(u^n \mid \theta; X) = \prod_{i=1}^{n} G(u_i \mid \theta; X)$$

- Set of model–compatible distributions:

$$\mathcal{P}_\theta^n = \left\{ P \in \Delta(S^n) : P = \int_U P_u dm_\theta^n, \text{ for some } P_u \in \Delta(G^n(u^n|\theta; X)) \right\}$$

Does this assumption restrict selection mechanisms to be IID? NO

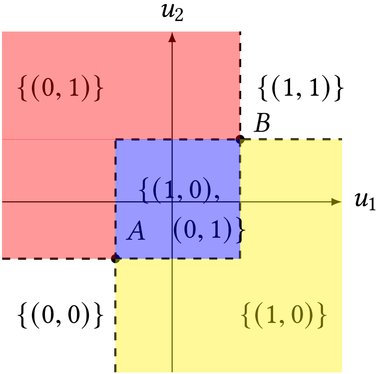
MLE is not Straightforward in Incomplete Models

- When $\beta = 0$, unique model prediction and hence likelihood
- When $\beta \neq 0$, if impose a selection in blue region (e.g., $(1, 0)$ is played for sure in each market), still unique likelihood
- But not knowing the selection? Non-unique likelihood, hence MLE not feasible

Remarks:

- Each specification of a selection leads to a different likelihood, hence MLE result
- Motivates an alternative approach that is agnostic about the selection mechanism

Bounds Approach (Ciliberto and Tamer, 2009)



- $Pr(u \in \text{yellow}) \leq Pr((1,0)) \leq Pr(u \in \text{yellow}) + Pr(u \in \text{blue})$
- In vector form: $\boldsymbol{H_1(\theta; X)} \leq Pr(\boldsymbol{y} \mid \boldsymbol{X}) \leq \boldsymbol{H_2(\theta; X)}$
- Identified set: set of pars $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\delta})$ that satsifies inequalties
- Estimate identified set and construct confidence region

Why Incompleteness Can Affect Inference

- Inference on identified set imposes i.i.d. or stationarity and mixing assumptions on data (e.g., Chernozhukov, Hong and Tamer, 2007)
- Unknown selection mechanisms across markets can cause unobserved heterogeneity and dependence
- May lead to non–ergodic distribution of data
- Invalidates central limit theorem (Epstein, Kaido and Seo, 2016)

One Example of Non–ergodic Sequence

- Suppose the N markets can be partitioned into clusters
  - e.g.: Markets 1–4 form a cluster, 5–15 form a cluster, etc.
- Within each cluster k, a Bernoulli random variable picks $(1, 0)$ with some cluster–specific cutoff rule
- The sequence of Bernoulli r.v. is i.n.i.d, but the selection mechanisms within each cluster are dependent

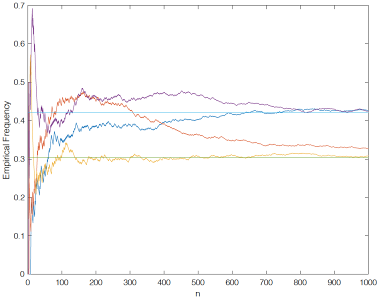


Figure 1.4: *Sample Path of* $n^{-1}\sum_{i=1}^{n} 1\{s_i = (1, 0)\}$ *(Incomplete Model)*

Outline

- Work in progress: robust score test
  - Inference on existence and sign of interaction effect $\boldsymbol{\beta}$
  - Robust to not knowing which equilibrium is played
    - Device: Least Favorable Pairs (Kaido and Zhang, 2019)
  - Why not likelihood ratio test? Nuisance parameters $\boldsymbol{\delta}$

Introduction of Least Favorable Pairs

Consider the simple null vs simple alternative testing in one market:

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta_0}, \quad H_1 : \boldsymbol{\theta} = \boldsymbol{\theta_1}$$

- Under H_0, set of model–compatible distributions $\mathcal{P}_{\boldsymbol{\theta_0}}$
- Under H_1, set of model–compatible distributions $\mathcal{P}_{\boldsymbol{\theta_1}}$

A test $\phi : S \rightarrow [0, 1]$ should

- Control the size under any distribution in $\mathcal{P}_{\boldsymbol{\theta_0}}$
- Have good power under any distribution in $\mathcal{P}_{\boldsymbol{\theta_1}}$
  - Lower power: power guaranteed regardless of unknown selection

Least favorable pair (LFP): $Q_0 \in \mathcal{P}_{\boldsymbol{\theta_0}}$ and $Q_1 \in \mathcal{P}_{\boldsymbol{\theta_1}}$

- Q_0 is least favorable for size control: among all distributions in $\mathcal{P}_{\boldsymbol{\theta_0}}$, largest type one error
- Q_1 is least favorable for lower power: among all distributions in $\mathcal{P}_{\boldsymbol{\theta_1}}$, smallest power

▸ Recap of size and power

LFP for Inference

Intuition:

- Given a simple hypothesis, find the pair of distributions that is the most difficult to distinguish from each other

Why can we do this?

- $\mathcal{P}_\theta$ has a structure
  - Characterize the set using lower probabilities
  - Entry–game: smallest probability that each outcome is played
- Tradeoff: $\mathcal{P}_\theta$ comes from model primitives

How do we do this?

- A convex algorithm based on Huber and Strassen (1973)

How to Compute an LFP in a Market

Conditional on $\boldsymbol{X}$ and $\boldsymbol{\delta}$, under the null $\boldsymbol{\beta_0} = \mathbf{0}$:

- Unique distribution of outcomes $(0,0),(0,1),(1,0),(1,1)$:

$$Q_0 = \Big((1-\Phi_1)(1-\Phi_2), (1-\Phi_1)\Phi_2, \Phi_1(1-\Phi_2), \Phi_1\Phi_2\Big),$$

  where $\Phi_i := \Phi(x^{(i)\prime}\delta^{(i)})$

Under an alternative $\boldsymbol{\beta_1} \in B := (-\infty, 0) \times (-\infty, 0)$:

- The algorithm partitions B into three regions whose boundaries depend on $\boldsymbol{\beta_1}, \boldsymbol{X}$ and $\boldsymbol{\delta}$
  - Economic interpretation: in the region of multiplicity, (1, 0) is played; (0, 1) is played; a mixture is played
- Each region has a unique distribution dependent on $\boldsymbol{\beta_1}, \boldsymbol{X}$ and $\boldsymbol{\delta}$

Given a specific alternative, algorithm determines the region it's in and hence distribution Q_1

An example of LFP form

LFPs in Cross Section

If latent variables are i.i.d. across markets, LFP for s^n is a Cartesian product of LFP for each market outcome (Kaido and Zhang, 2019)

- Even though selection mechanisms across markets can be intertwined in unknown forms, it doesn't matter for product LFP
- Each LFP is a pair of likelihood, product LFP is likelihood
- A unique likelihood under the null, another under the alternative
- Implication? Likelihood ratio test for simple hypothesis testing

Robust Likelihood Ratio Test (Kaido and Zhang, 2019)

Cross section without exogenous covariates, inference is

$$H_0 : \beta^{(i)} = 0, \quad H_1 : \beta^{(i)} = \boldsymbol{\beta_1}$$

Likelihood ratio test: $\phi_n : S^n \mapsto [0,1]$ such that

$$\phi_n(s^n) = \begin{cases} 1 & \Lambda_n(s^n) > C_n \\ \gamma_n & \Lambda_n(s^n) = C_n \\ 0 & \Lambda_n(s^n) < C_n \end{cases}$$

1. Given the alternative, compute LFP Q_0 and Q_1 for each market
2. Form ratio of using likelihood pairs: $\Lambda_n(s^n) = \prod_{i=1}^{n} \frac{dQ_1(s_i)}{dQ_0(s_i)}$
3. Compute $\mu_{Q_0} = \mathbb{E}_{Q_0}(\ln \Lambda_n(s^n))$ and $\sigma^2_{Q_0} = Var_{Q_0}(\Lambda_n(s^n))$. Denote z_α as $1-\alpha$ quantile of $\mathcal{N}(0,1)$, construct the critical value

$$C_n = \exp(n\mu_{Q_0} + \sqrt{n} z_\alpha \sigma_{Q_0})$$

Why Robust Score Test?

Motivation:

- Subvector inference:

$$H_0 : \beta^{(i)} = 0,\ \boldsymbol{\delta} \in \Theta_\delta \quad H_1 : \beta^{(i)} < 0,\ \boldsymbol{\delta} \in \Theta_\delta$$

- Composite null and composite alternative
- Kaido and Zhang (2019) provide a likelihood-ratio test that
  - controls size uniformly over $\Theta_0 \equiv \{\theta := (\boldsymbol{\beta}, \boldsymbol{\delta}) : \beta^{(i)} = 0, \delta \in \Theta_\delta\}$
  - maximizes the weighted average **lower** power
  - becomes computationally intensive for moderately high dimensional $\boldsymbol{\delta}$

Advantages of score test:

- Local power analysis
  - Under the null, can consistently estimate $\boldsymbol{\delta}$
- Relatively easy to implement

Key Ingredients of Robust Score Test

Purpose:

- Conduct inference on $\boldsymbol{\beta} = \mathbf{0}$ in the presence of unknown selection mechanisms and coefficients of exogenous covariates $\boldsymbol{\delta}$

Procedures:

1. Given Q_0 and an alternative $\beta^{(i)} = 0 + h_i/\sqrt{n}$, compute Q_1
2. Compute the score (derivative) of the log likelihood $\ln Q_1$
3. Estimate $\boldsymbol{\delta}$ by restricted MLE (under $\boldsymbol{\beta} = \mathbf{0}$)
4. Compute the test statistic

Score Functions

For one observation, takes the following general form

$$\dot{\ell}(s;x) = \begin{bmatrix} \dot{\ell}_\beta(s;x) \\ \dot{\ell}_\delta(s;x) \end{bmatrix} = \sum_{\bar{x}\in X} \sum_{\bar{s}\in S} 1\{x=\bar{x}, s=\bar{s}\} \begin{bmatrix} z_\beta(\bar{s};\bar{x}) \\ z_\delta(\bar{s};\bar{x}) \end{bmatrix},$$

where for each $\bar{s} \in S$ and $\bar{x} \in X$,

$$z_\beta(\bar{s};\bar{x}) = \begin{bmatrix} z_{\beta^{(1)}}(\bar{s};\bar{x}) \\ z_{\beta^{(2)}}(\bar{s};\bar{x}) \end{bmatrix} = \frac{\partial}{\partial \beta} \ln q_1(\bar{s};\bar{x})$$

$$z_\delta(\bar{s};\bar{x}) = \begin{bmatrix} z_{\delta^{(1)}}(\bar{s};\bar{x}) \\ z_{\delta^{(2)}}(\bar{s};\bar{x}) \end{bmatrix} = \frac{\partial}{\partial \delta} \ln q_1(\bar{s};\bar{x})$$

For a sequence of observations:

$$\sum_{i=1}^{n} \dot{\ell}(s;x) = \begin{bmatrix} \sum_{i=1}^{n} \dot{\ell}_\beta(s;x) \\ \sum_{i=1}^{n} \dot{\ell}_\delta(s;x) \end{bmatrix} = \sum_{\bar{s}\in S} \sum_{\bar{x}\in X} \#\{(\bar{s},\bar{x})\} \begin{bmatrix} z_\beta(\bar{s};\bar{x}) \\ z_\delta(\bar{s};\bar{x}) \end{bmatrix}$$

$\#\{(\bar{s},\bar{x})\}$: number of occurrences of event $\bar{s}$ and covariate $\bar{x}$

Hypothesis and Neyman's Orthogonality

- Inference on β in the presence of nuisance parameter δ:

$$H_0 : \beta_0 = (\beta_0^{(1)}, \beta_0^{(2)}) = (0,0),\ \delta \in \Theta_\delta$$
$$H_1 : \beta_1 = (\beta_1^{(1)}, \beta_1^{(2)}) < (0,0),\ \delta \in \Theta_\delta$$

- Precursor: Neyman's $C(\alpha)$ test

$$C_{\beta,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{\ell}_\beta(s;x), \quad C_{\delta,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{\ell}_\delta(s;x)$$

- Replace unknown δ with consistently estimated $\widehat{\delta}$
- To guard against estimation error, orthogonalize the score using

$$\mathbb{E}_{Q(\beta,\delta)} \begin{bmatrix} \sum_{i=1}^n \dot{\ell}_{\beta^{(1)}}(s;x) \\ \sum_{i=1}^n \dot{\ell}_{\beta^{(2)}}(s;x) \\ \sum_{i=1}^n \dot{\ell}_{\delta^{(1)}}(s;x) \\ \sum_{i=1}^n \dot{\ell}_{\delta^{(2)}}(s;x) \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n \dot{\ell}_{\beta^{(1)}}(s;x) \\ \sum_{i=1}^n \dot{\ell}_{\beta^{(2)}}(s;x) \\ \sum_{i=1}^n \dot{\ell}_{\delta^{(1)}}(s;x) \\ \sum_{i=1}^n \dot{\ell}_{\delta^{(2)}}(s;x) \end{bmatrix}' = \begin{bmatrix} \mathcal{I}_{\beta\beta}, \mathcal{I}_{\beta\delta} \\ \mathcal{I}_{\delta\beta}, \mathcal{I}_{\delta\delta} \end{bmatrix}$$

Sup Test Statistic

The orthogonalized score under $Q_{(\beta_0,\delta_0)}$:

$$g_n(\delta) = C_{\beta_0,n} - \mathcal{I}_{\beta_0\delta}\mathcal{I}_{\delta\delta}^{-1}C_{\delta,n}$$

which has variance

$$\mathcal{I}_{\beta:\delta} = \mathcal{I}_{\beta\beta} - \mathcal{I}_{\beta\delta}\mathcal{I}_{\delta\delta}^{-1}\mathcal{I}_{\delta\beta}$$

Define

$$Z_n = \begin{pmatrix} z_{1,n} \\ z_{2,n} \end{pmatrix} := \mathcal{I}_{\beta_0:\widehat{\delta}}^{-1/2} g_n(\widehat{\delta})$$

and consider the following test statistic

$$T_n := \max\left\{|z_{1,n}|, |z_{2,n}|\right\}$$

Remark: In simulations need regularization on $\mathcal{I}_{\delta\delta}$ or $\mathcal{I}_{\beta:\delta}$

Limiting Distribution and Critical Value

Limiting distribution:

$$T_n := \max\left\{|z_{1,n}|, |z_{2,n}|\right\} \overset{a}{\sim} \sup\{|w_1|, |w_2|\},$$

where $[w_1, w_2]' \sim \mathcal{N}(0, I_2)$. The critical value for size α is defined to be

$$c_\alpha = \inf\{x : Pr(\sup\{|w_1|, |w_2|\} \leq x) \geq 1 - \alpha\}$$

Procedures of getting critical values:

1. Draw a 2×5000 vector from standard normal distribution
2. Take max of absolute value for each row
3. Compute the $(1 - \alpha)$th quantile

Monte Carlo Simulation Design

Parameters

Fix $\delta_0 = [2, 2.5]$, $n = [200, 500, 1000, 1500, 3000, 5000]$

Size: $\beta_0 = [0, 0]$,

Power: $\beta_0^{(1)} = \beta_0^{(2)} = -h/\sqrt{n}, \quad h = -[eps : 0.5 : 15]$

DGP Construction Procedures

1. Draw x from the uniform discrete distribution $U\{-1, 1\}^2$. Four possible configurations: $(1, 1), (1, -1), (-1, 1)$ and $(-1, -1)$.
2. Draw (u_1, u_2) from the bivariate standard normal distribution.
3. For each draw of (u_1, u_2), determine $G(u \mid \beta; X, \delta)$ based on the analytical form.
4. Repeat procedures 1–3 for $S = 5000$ times

Remark: When $\beta_0 \neq 0$, multiple equilibria exist for some draws of (u_1, u_2), select according to one of the three selection mechanisms:

- IID; Non IID; LFP

▸ Non IID Details ▸ LFP Illustration

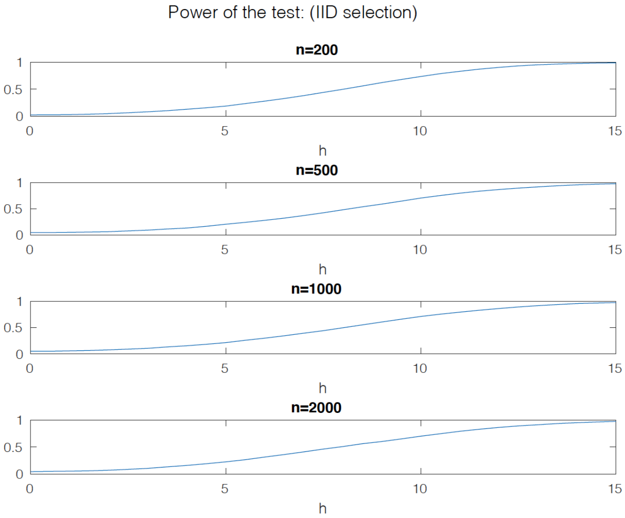
Size Properties

Table: Empirical Finite–Sample Size of Sup Statstic Test ($\alpha = 0.05$)

$N = 200$	$N = 500$	$N = 1000$	$N = 1500$	$N = 3000$	$N = 5000$
0.0306	0.0535	0.0568	0.0572	0.0492	0.0516

- In small sample, need regularization on test statstic
- Currently don't have a data–driven rule

Local Power Properties



▸ Plots for Non IID and LFP

Empirical Application

Data: 2nd quarter of 2010 Airline Origin and Destination Survey

- Source: Kline and Tamer (2016, QE)
- 7882 markets: trips between two airports irrespective of intermediate stops
- Players: LCC (low cost carriers); OA (other airlines)

Payoff of player $i = \{LCC, OA\}$ if it enters market m:

$$\delta_i^{cons} + \delta_i^{size} X_{m,size} + \delta_i^{pres} X_{i,m,pres} + \beta_i y_{-i,m} + \varepsilon_i$$

- $X_{m,size}$: size of market m; 1 if larger than median, 0 o.w.
- $X_{i,m,pres}$: market presence of i in m; 1 if larger than median, 0 o.w.
- $y_{-i,m}$: 1 if opponent enters, 0 o.w.
- ε_{LCC} and ε_{OA} are bivariate standard normal

Hypothesis: $\boldsymbol{\beta} = 0, \boldsymbol{\delta} \in \Theta_\delta$ vs $\boldsymbol{\beta} < 0, \boldsymbol{\delta} \in \Theta_\delta$

Implementation and Result

1. Under the null, estimate $\boldsymbol{\delta}_{LCC}$ and $\boldsymbol{\delta}_{OA}$ using RMLE with multiple starting points
2. Compute the sup test statistic

Test statistic: $2.9102 > crit_{0.99} = 2.7244$

- Reject the null, competition effect exists
- Confirms results in Kline and Tamer (2016)

Ongoing Work

- Testing on the sign of interaction effects: differentiation vs coordination in an incomplete information game[1]
  - Incomplete information: error realization is private knowledge
  - Application in mind: radio commercials (Sweeting, 2009)
  - *Challenge*: multiple equilibria affects outcome indirectly via equilibrium choice probabilities
- A null that features incompleteness
  - *Challenge*: might not be able to consistently estimate δ
- Combine with a test that has global power
  - Two–step testing approach
  - *Challenge*: how to account for first–step testing error?

[1] We thank Marc Rysman for suggesting this extension.

Conclusion

- An incomplete model makes set–valued predictions
- Assumptions on selection mechanisms and data sampling affect estimation and inference
- Robust score test for the existence of interaction effects

An Example of LFP in a Market

When $(1,0)$ is played for sure in the multiplicity region:

$$Q_1 = (q_1(0,0), q_1(0,1), q_1(1,0), q_1(1,1)),$$

where

$$q_1(0,0) = (1-\Phi_1)(1-\Phi_2)$$
$$q_1(0,1) = (1-\Phi_1)\Phi_2 + \Phi(x^{(2)\prime}\delta_2 + \beta^{(2)})[\Phi_1 - \Phi(x^{(1)\prime}\delta_1 + \beta^{(1)})]$$
$$q_1(1,0) = \Phi_1(1-\Phi(x^{(2)\prime}\delta^{(2)} + \beta^{(2)}))$$
$$q_1(1,1) = \Phi(x^{(1)\prime}\delta^{(1)} + \beta^{(1)})\Phi(x^{(2)\prime}\delta^{(2)} + \beta^{(2)})$$

◂ Back

Size and Power in Testing

- Two concepts in hypothesis testing: **size** and **power**
  - Size: given that H_0 is true, probability that the test rejects H_0
  - Power: given that H_1 is true, probability that the test rejects H_0
- For size properties, examine the distribution of the test under H_0
  - Implication? LM test has good size
  - In Monte Carlo, generate data under H_0 and compare empirical critical values with theoretical ones
- The alternative is $h(\theta) \neq 0$, rather broad
  - Depends on direction and magnitude of deviation from the null
    1. Direction: Tests are typically not omnibus
    2. Magnitude: local power analysis
  - Implication? Wald test has good power for specific alternatives[2]
  - In Monte Carlo, generate data under specific H_1
- Takeaway: Among tests that have well–controled size, the optimal test should have the highest power, which depends on the alternatives. ◂ Back

[2]However, one drawback of Wald is that it is not invariant to the way the hypothesis is written unless it is linear.

Non IID Details

Let N_k^* be an increasing sequence of integers. For each i, let $h(i) = N_k^*$ where $N_{k-1}^* < i \leq N_k^*$, define

$$\tilde{\nu}_i = \begin{cases} 1 & \Psi_{h(i)}^G(u) > \Lambda_{h(i)} \\ 0 & \Psi_{h(i)}^G(u) \leq \Lambda_{h(i)} \end{cases}$$

where

$$\Psi_{h(i)}^G(u) = \frac{\sum_{i=1}^{h(i)} 1[G(u_i|\beta; X, \delta) = \{(1,0)\}]}{\sum_{i=1}^{h(i)} 1[G(u_i|\beta; X, \delta) = \{(1,0),(0,1)\}]}$$

Conditional on X, compute the conditional lower probabilities of $(1,0)$ and $(0,1)$, $\nu_{\beta,\delta|X}((1,0))$ and $\nu_{\beta,\delta|X}((0,1))$. Let N_c denote the number of occurrences of X's configuration c within N_k^*, i.e., $\sum_c N_c = h(i)$, calculate the empirical weighted sum of the two events and define $\Lambda_{h(i)}$ as follows, ◂ Back

$$\Lambda_{h(i)} = \frac{\sum_{c \in \{(1,1),(1,-1),(-1,1),(-1,-1)\}} N_c \nu_{\beta,\delta|c}((1,0))}{\sum_{c \in \{(1,1),(1,-1),(-1,1),(-1,-1)\}} N_c \Big(\nu_{\beta,\delta|c}((1,0)) + \nu_{\beta,\delta|c}((0,1))\Big)}$$

LFP Mechanism Illustration

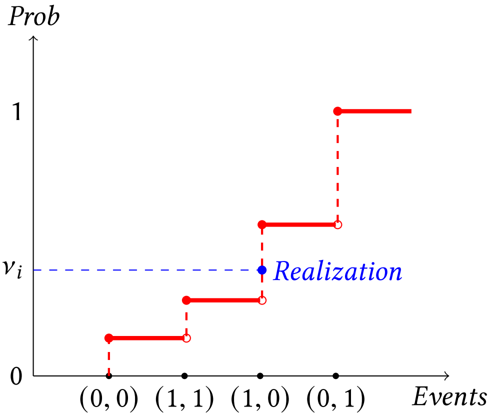
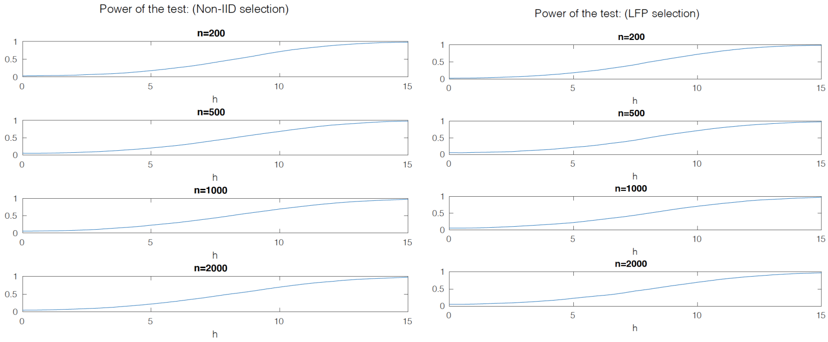


Figure: CDF of LFP of One Observation

◂ Back

Power Properties



- One possible reason for similarity: difference between different selection mechanism arises with small probabilties given the alternatives

Back