AGT and Data Science

Jamie Morgenstern, University of Pennsylvania
Vasilis Syrgkanis, Microsoft Research

AGT and Data Science
Part 2

Econometric Theory for Games

Vasilis Syrgkanis, Microsoft Research

Comparison with Part (1)

- Optimization vs Estimation
  - Part 1: find revenue maximizing mechanism from data
  - Part 2: interested in inference of private parameters of structural model

- Truthful vs Strategic Data
  - Part 1: data set were i.i.d. samples of player valuations
  - Part 2: data are observed outcomes of strategic interaction (e.g. bids in FPA)

- Technical Exposition vs Overview
  - Part 1: in-depth exposition of basic tools
  - Part 2: overview of econometric theory for games literature with some in-depth drill downs

A Primer on Econometric Theory

Basic Tools and Terminology

Econometric Theory

- Given a sequence of i.i.d. data points $Z_1, \ldots, Z_n$
- Each Z_i is the outcome of some structural model

$$Z_i \sim D(\theta_0), \text{with } \theta_0 \in \Theta$$

- Parameter space Θ can be:
  - Finite dimensional (e.g. R^d): parametric model
  - Infinite dimensional (e.g. function): non-parametric model
  - Mixture of finite and infinite:
    - If we are interested only in parametric part: Semi-parametric
    - If we are interested in both: Semi-nonparametric

Main Goals

- Identification: If we new "population distribution" $D(\theta_0)$ then can we pin-point θ_0?

- Estimation: Devise an algorithm that outputs an estimate $\hat{\theta}_n$ of θ_0 when having n samples

Estimator Properties of Interest

- Finite Sample Properties of Estimators:
  - Bias $= E[\hat{\theta}_n] - \theta_0 = 0$?
  - Variance: $\mathrm{Var}(\hat{\theta}_n)$?
  - Mean-Squared-Error (MSE): $E\left[\left(\hat{\theta}_n - \theta_0\right)^2\right] = \text{Variance} + \text{Bias}^2$
- Large Sample Properties: $n \to \infty$
  - Consistency: $\hat{\theta}_n \to \theta_0$?
  - Asymptotic Normality: $a_n\left(\hat{\theta}_n - \theta_0\right) \to N(0, V)$?
  - $\sqrt{n}$-consistency: $a_n = \sqrt{n}$?
  - Efficiency: is limit variance V information theoretically optimal? (typically achieved by MLE estimator)

General Classes of Estimators

- Extremum Estimator

$$\theta_0 = \operatorname{argmax}_{\theta \in \Theta} Q_n(\theta)$$

- Examples
  - MLE: $Q_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \ln f(z_i; \theta)$
  - GMM Estimator: suppose in population $E[m(z, \theta)] = 0$. Empirical analogue: for some W positive definite

$$Q_n(\theta) = \left[\frac{1}{n} \sum_i m(z_i, \theta)\right]' W \left[\frac{1}{n} \sum_i m(z_i, \theta)\right]$$

Consistency of Extremum Estimators

Consistency Theorem. If there is a function $Q_0(\theta)$ s.t.:

1. $Q_0(\theta)$ is uniquely maximized at θ_0
2. $Q_0(\theta)$ is continuous
3. $Q_n(\theta)$ converges uniformly in probability to $Q_0(\theta)$, i.e. $\sup_\theta |Q_n(\theta) - Q_0(\theta)| \to_p 0$

Then $\hat{\theta} \to_p \theta_0$

- If $Q_n(\theta) = \frac{1}{n}\sum_i g(z_i, \theta)$ and $Q_0(\theta) = E[g(z,\theta)]$, then (2.,3.) will be satisfied if
  - $g(z,\theta)$ is continuous
  - $g(z,\theta) \leq d(z)$ with $E[d(z)] \leq \infty$
- Typically referred to as “regularity conditions”

Asymptotic Normality

- Under "regularity conditions" asymptotic normality of extremum estimators follows by ULLN, CLT, Slutzky thm and consistency
- Roughly: consider case $Q_n(\theta) = \frac{1}{n}\sum_i g(z_i, \theta)$
  - Take first order condition

$$\frac{1}{n}\sum_i \nabla_\theta g(z_i, \hat{\theta}) = 0$$

  - Linearize around θ_0 by mean value theorem

$$\frac{1}{n}\sum_i \nabla_\theta g(z_i, \theta_0) + \left[\frac{1}{n}\sum_i \nabla_{\theta\theta} g\left(z_i, \bar{\theta}\right)\right]\left(\hat{\theta} - \theta_0\right) = 0$$

  - Re-arrange:

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = \underbrace{\left[\frac{1}{n}\sum_i \nabla_{\theta\theta} g\left(z_i, \bar{\theta}\right)\right]^{-1}}_{\to_p E[\nabla_{\theta\theta} g(z, \theta_0)]} \cdot \underbrace{\frac{1}{\sqrt{n}}\sum_i \nabla_\theta g(z_i, \theta_0)}_{\to_d N\left(0, Var\left(\nabla_\theta g(z, \theta_0)\right)\right)} \quad \to_d N(0, U)$$

In practice, typically variance is computed via Bootstrap [Efron'79]: Re-sample from your samples with replacement and compute empirical variance

Econometric Theory for Games

Econometric Theory for Games

- Z_i are observable quantities from a game being played
- θ_0: unobserved parameters of the game

- Address identification and estimation in a variety of game theoretic models assuming players are playing according to some equilibrium notion

Why useful?

- Scientific: economically meaningful quantities
- Perform counter-factual analysis: what would happen if we change the game?
- Performance measures: welfare, revenue
- Testing game-theoretic models: if theory on estimated quantities predicts different behavior, then in trouble

Outline of the rest of the talk

- Complete information games
  - Multiplicity of equilibria: partial identification and set inference
- Discrete Static and Dynamic Games of Incomplete Information
  - Two-stage estimators
- Auction games
  - Identification and estimation in first price auctions with independent private values
- Algorithmic game theory and econometrics
  - Mechanism design for data science
  - Econometrics for learning agents

A Seminal Example

Entry Games [Bresnahan-Reiss’90,91] and [Berry’92]

Entry Game

- Two firms deciding whether to enter a market
- Entry decision $y_i \in \{0,1\}$
- Profits from entry:

$$\pi_1 = x \cdot \beta_1 + y_2 \delta_1 + \epsilon_1$$
$$\pi_2 = x \cdot \beta_2 + y_1 \delta_2 + \epsilon_2$$

- Equilibrium:

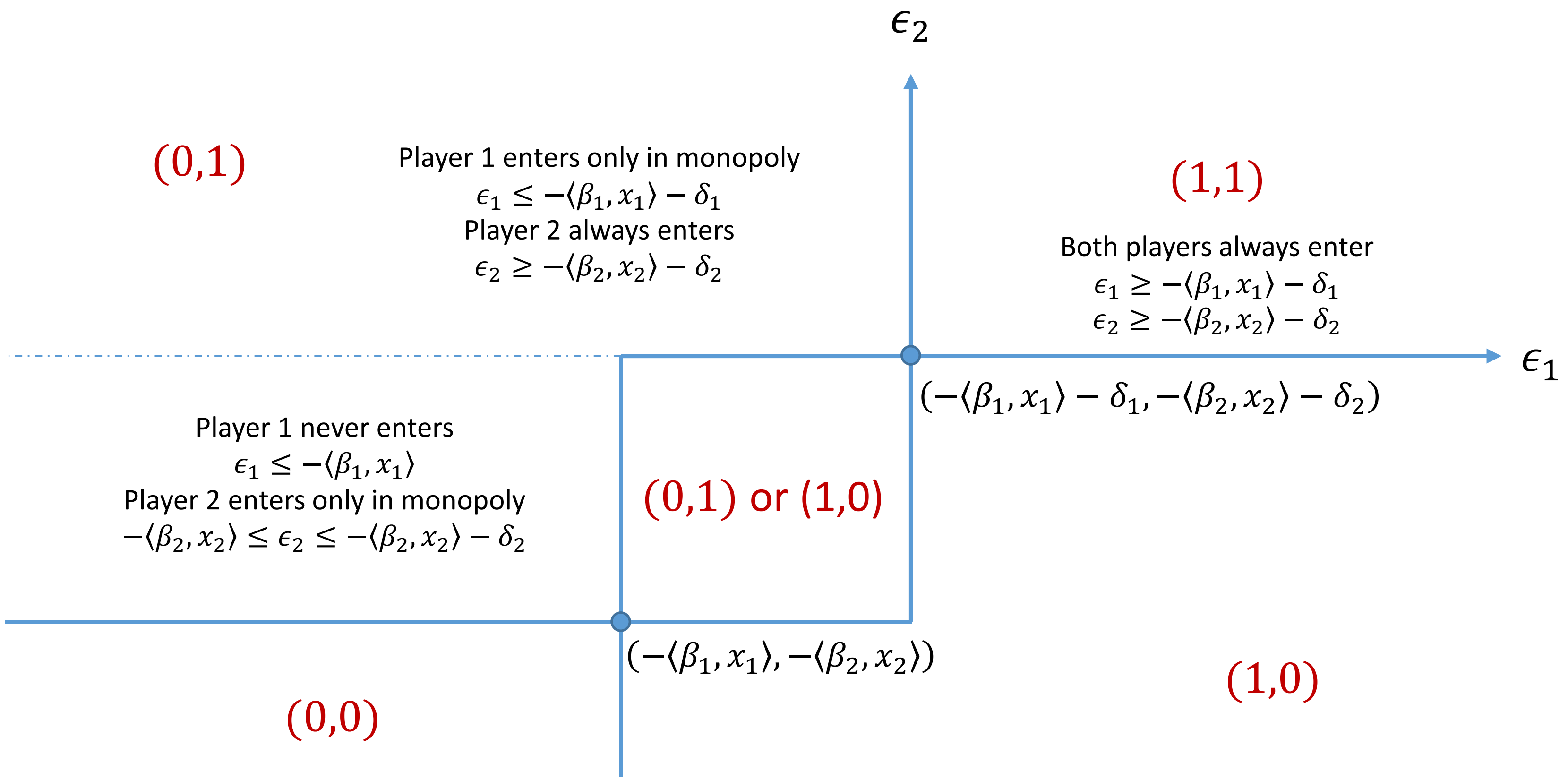
$$y_i = 1\{\pi_i \geq 0\}$$

- $\epsilon_i \sim F_i$: at each market i.i.d. from known distribution
- x: observable characteristics of each market
- β_i, δ_i: constants across markets

Assume $\delta_1, \delta_2 < 0$

[Bresnahan-Reiss'90,91], [Berry'92]

$$\pi_1 = x \cdot \beta_1 + y_2\delta_1 + \epsilon_1$$
$$\pi_2 = x \cdot \beta_2 + y_1\delta_2 + \epsilon_2$$



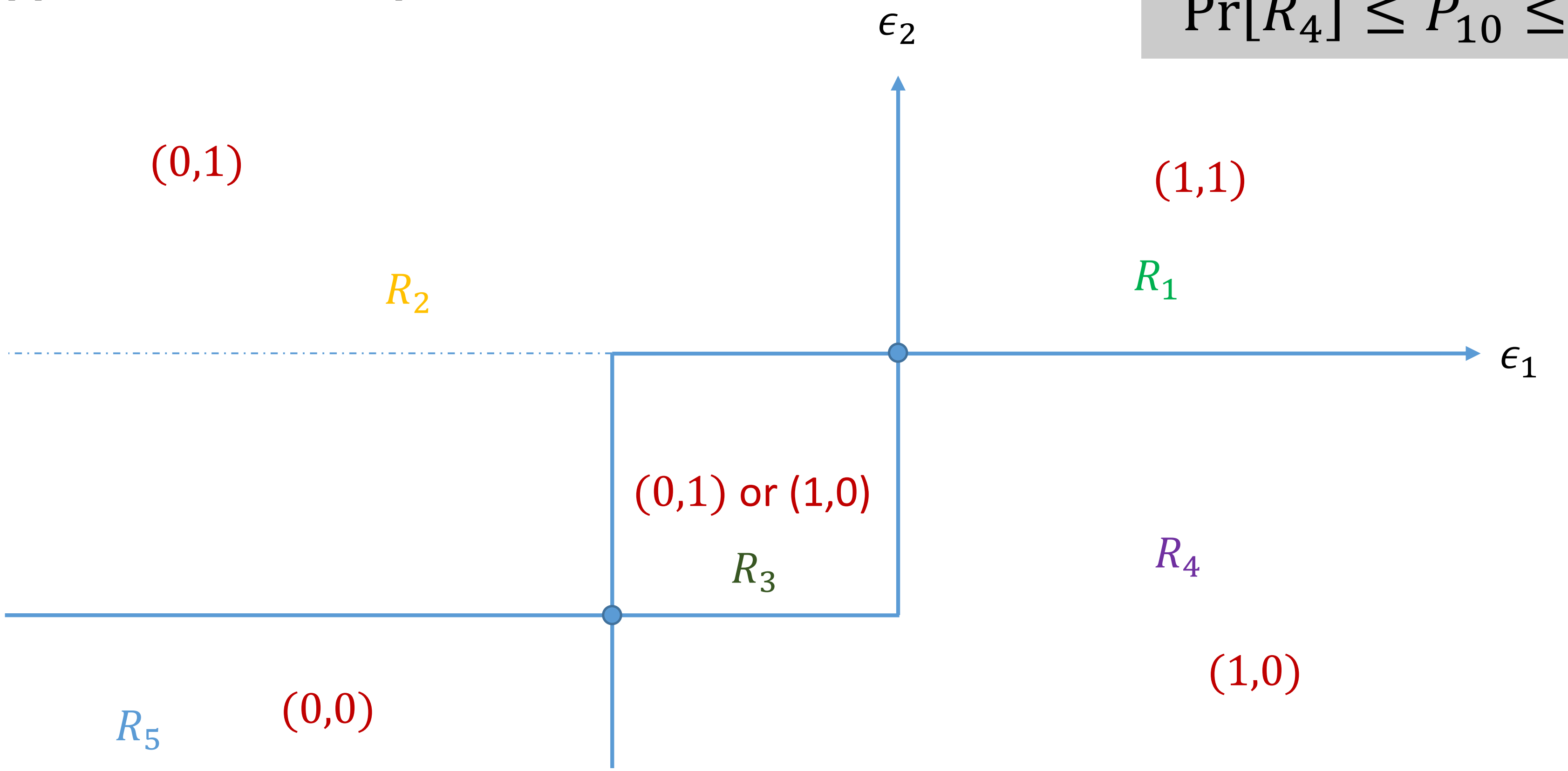
- In all regions: equilibrium number of entrants $N = y_1 + y_2$ is unique
- Can perform MLE estimation using N as observation

More generally

[Tamer'03] [Cilliberto-Tamer'09]

Identified set Θ_I: β, δ s.t.:

$$P_{11} = \Pr[R_1]$$
$$P_{00} = \Pr[R_5]$$
$$\Pr[R_2] \leq P_{01} \leq \Pr[R_2] + \Pr[R_3]$$
$$\Pr[R_4] \leq P_{10} \leq \Pr[R_3] + \Pr[R_4]$$



- Equilibrium will be some selection of possible equilibria $S(\epsilon)$
- Imposes inequalities on probability of each action profile

Estimating the Identified set

[Cilliberto-Tamer'09]

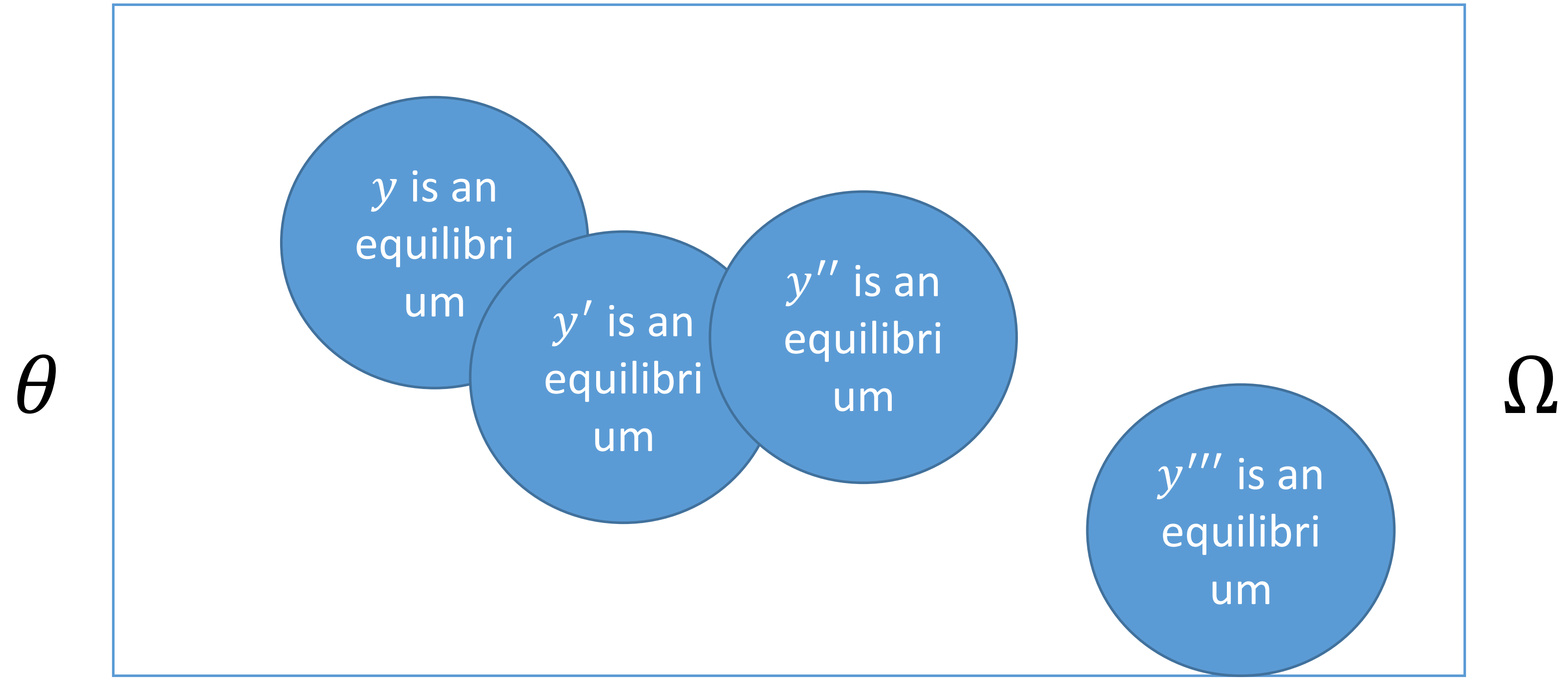
$$\Theta_I = \{\beta, \delta: P_{11} = \Pr[R_1], P_{00} = \Pr[R_5], \\ \Pr[R_2] \leq P_{01} \leq \Pr[R_2] + \Pr[R_3], \\ \Pr[R_4] \leq P_{10} \leq \Pr[R_3] + \Pr[R_4]\}$$

- Distribution of ϵ known: $\Pr[R_i]$ some known function $G_i(X; \beta, \delta)$ of parameters
- y_1, y_2, X: observed in the data
- Replace population probabilities with empirical: $P_{y_1 y_2 X} \rightarrow \hat{P}_{y_1 y_2 X}$
- Add slack to allow for error in empirical estimates:

$$\hat{P}_{y_1, y_2 X} \leq G_2(X; \beta, \delta) + G_3(X; \beta, \delta) + \frac{\nu_n}{n}$$

where $\nu_n \rightarrow \infty$ and $\frac{\nu_n}{n} \rightarrow 0$ (asymptotic properties [Chernozukhov-Hong-Tamer'07])

General Games



- Ω: probability space where unobserved randomness lives (e.g. ϵ)
- Each θ defines the set of equilibria for each $\omega \in \Omega$
- One of these equilibria will be selected
- We only observe distribution of outcomes y: $\Pr[y = k]$ for each possible equilibrium k
- Is θ admissible for a given population of outcomes?

Characterization of the Identified Set

[Beresteanu-Molchanov-Mollinari'09]

Theorem [Artsein'83, Beresteanu-Molchanov-Mollinari'07]. Let Z_θ be a random set in 2^K and let y_θ be a random variable in K. Then y_θ is a selection of Z_θ (i.e. $y_\theta \in Z_\theta$ a.s.) if and only if:

$$\forall S \subseteq K\text{: } \Pr[y_\theta \in S] \leq \Pr[Z_\theta \cap S \neq \emptyset]$$

In games:

- K is the set of possible equilibria of a game
- Z_θ is the set of equilibria for a given realization of the unobserved ϵ,
- $\Pr[y_\theta \in S]$: population distribution of action profiles
- Thus: $\Theta_I = \{\theta\text{: } \forall S \subseteq K, \Pr[y_\theta \in S] \leq \Pr[Z_\theta \cap S \neq \emptyset]\}$
- Defined as a set of moment inequalities

Characterization of the Identified Set

[Beresteanu-Molchanov-Mollinari'09]

Theorem [Artsein'83, Beresteanu-Molchanov-Mollinari'07]. Let Z_θ be a random set in 2^K and let y_θ be a random variable in K. Then y_θ is a selection of Z_θ (i.e. $y_\theta \in Z_\theta$ a.s.) if and only if:

$$\forall S \subseteq K: \ \Pr[y_\theta \in S] \leq \Pr[Z_\theta \cap S \neq \emptyset]$$

- For the example latter is equivalent to Θ_I of [Cilliberto-Tamer'09]
- For more general settings it is strictly smaller and sharp
- Can perform estimation based on moment inequalities similar to [CT'09]

$$\widehat{\Theta}_I = \left\{\theta: \hat{P}[y_\theta \in S] \leq \Pr[Z_\theta \cap S] + \frac{\nu_n}{n}\right\}$$

where $\nu_n \to \infty$ and $\frac{\nu_n}{n} \to 0$

Main take-aways

- Games of complete information are typically partially identified
- Multiplicity of equilibrium is the main issue
- Leads to set-estimation strategies and machinery [Chernozhukov et al'09]
- Very interesting random set theory for estimating the sharp identifying set

Incomplete Information Games and Two-Stage Estimators

Static Games: [Bajari-Hong-Krainer-Nekipelov’12]

Dynamic Games: [Bajari-Benkard-Levin’07], [Pakes-Ostrovsky-Berry’07], [Aguirregabiria-Mira’07], [Ackerberg-Benkard-Berry-Pakes’07], [Bajari-Hong-Chernozhukov-Nekipelov’09]

High level idea

- At equilibrium agents have beliefs about other players actions and best respond
- If econometrician observes the same information about opponents as the player does then:
  - Estimate these beliefs from the data in first stage
  - Use best-response inequalities to these estimated beliefs in the second stage and infer parameters of utility

Static Entry Game with Private Shocks

- Two firms deciding whether to enter a market
- Entry decision $y_i \in \{0,1\}$
- Profits from entry:

$$\pi_1 = x \cdot \beta_1 + y_2\delta_1 + \epsilon_1$$
$$\pi_2 = x \cdot \beta_2 + y_1\delta_2 + \epsilon_2$$

- $\epsilon_i \sim F_i$: at each market i.i.d. from known distribution and **private to player**
- x: observable characteristics of each market
- β_i, δ_i: constants across markets

Static Entry Game with Private Shocks

- Firms best-respond only in expectation
- Expected profits from entry:

$$\Pi_1 = x \cdot \beta_1 + \Pr[y_2 = 1|x]\,\delta_1 + \epsilon_1$$
$$\Pi_2 = x \cdot \beta_2 + \Pr[y_1 = 1|x]\,\delta_2 + \epsilon_2$$

- Let $\sigma_i(x) = \Pr[y_i = 1|x]$
- Then:

$$\sigma_1(x) = \Pr[x \cdot \beta_1 + \sigma_2(x)\delta_1 + \epsilon_1 > 0]$$
$$\sigma_2(x) = \Pr[x \cdot \beta_2 + \sigma_1(x)\delta_2 + \epsilon_2 > 0]$$

Static Entry Game with Private Shocks

- If ϵ_i is distributed according to an extreme value distribution:
$$\sigma_1(x) \propto \exp[x \cdot \beta_1 + \sigma_2(x)\delta_1]$$
$$\sigma_2(x) \propto \exp[x \cdot \beta_2 + \sigma_1(x)\delta_2]$$
- Non-linear system of simultaneous equations
- Computing fixed point is computationally heavy and fixed-point might not be unique
- Idea [Hotz-Miller'93, Bajari-Benkard-Levin'07, Pakes-Ostrovsky-Berry'07, Aguirregabiria-Mira'07, Bajari-Hong-Chernozhukov-Nekipelov'09]: Use a two stage estimator
    1. Compute non-parametric estimate $\hat{\sigma}(x)$ of function $\sigma_i(x)$ from data
    2. Run parametric regressions for each agent individually from the condition that:
    $$\sigma_i(x) \propto \exp[x \cdot \beta_i + \hat{\sigma}_{-i}(x)\ \delta_i]$$
    3. The latter is a simple logistic regression for each player to estimate β_i, δ_i

Simple case: finite discrete states

- If there are d states, then σ_i are d-dimensional parameter vectors
- Easy $\sqrt{n}$-consistent first-stage estimators $\hat{\sigma} = (\hat{\sigma}_1, \hat{\sigma}_2)$ of $\sigma = (\sigma_1, \sigma_2)$, i.e.:
$$\sqrt{n}(\hat{\sigma}_i - \sigma) \to N(0, V)$$
- Suppose for second stage we do generalized method of moment estimator:
  - Let $\hat{\theta} = \left(\hat{\beta}_1, \hat{\beta}_2, \hat{\delta}_1, \hat{\delta}_2\right)$ and $\theta_0 = (\beta_1, \beta_2, \delta_2, \delta_2)$
  - Let $y_t = (y_{1t}, y_{2t})$ and $\Gamma(x, \sigma, \theta) = \left(\Gamma_1(x, \sigma, \theta), \Gamma_2(x, \sigma, \theta)\right)$ with $\Gamma_i(x, \sigma, \theta) = \frac{e^{x \cdot \beta_i + \sigma_{-i}\delta}}{1 + e^{x \cdot \beta_i + \sigma_{-i}\delta}}$
  - Then second stage estimator $\hat{\theta}$ is the solution to:
$$\frac{1}{n} \sum_{t=1}^{n} A(x_t) \cdot \left(y_t - \Gamma\left(x_t, \hat{\sigma}, \hat{\theta}\right)\right) = 0$$
- Does first stage error affect second stage variance and how?
- This is a general question about two stage estimators

Two-Stage GMM with $\sqrt{n}$-Consistent First Stage [Newey-McFadden'94: Large Sample Estimation and Hypothesis Testing]

- Run a first step estimator $\hat{\sigma}$ of σ, with $\sqrt{n}\,(\hat{\sigma}-\sigma) \to N(0,V)$
- Second stage is a GMM estimator based on moment conditions $E[m(z,\theta,\sigma)] = 0$, i.e. $\hat{\theta}$ satisfies:

$$\frac{1}{n}\sum_{t=1}^{n} m\left(z_t,\hat{\theta},\hat{\sigma}\right) = 0$$

- Linearize around θ:

$$\sqrt{n}\left(\hat{\theta}-\theta\right) = -\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial m\left(z_t,\bar{\theta},\hat{\sigma}\right)}{\partial\theta}\right]^{-1}\cdot\frac{1}{\sqrt{n}}\sum_{t=1}^{n} m(z_t,\theta,\hat{\sigma})$$

- Now the second term can be linearized around σ:

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} m(z_t,\theta,\hat{\sigma}) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} m(z_t,\theta,\sigma) + \frac{1}{n}\sum_{t=1}^{n}\frac{\partial m(z_t,\theta,\bar{\sigma})}{\partial\sigma}\cdot\sqrt{n}(\hat{\sigma}-\sigma)$$

Continuous State Space: $x \in R^d$

[Bajari-Hong-Kranier-Nekipelov'12]

- Then there is no $\sqrt{n}$-consistent first stage non-parametric estimator $\hat{\sigma}(\cdot)$ for function $\sigma(\cdot) = E[y|x]$
- Remarkably: still $\sqrt{n}$-consistency for second stage estimate $\hat{\theta}$!!
- For instance:
  - Kernel estimator for the first stage (tune bandwidth, "undersmoothing")
  - GMM for second stage
- Intuition (my rough take on it):
  - Kernel estimators have tunable "bias"-"variance" tradeoffs
  - Close to true θ: first stage bias and variance affect linearly second stage estimate
  - If variance and bias decay at $n^{-\frac{1}{2}}$ rates we
  - Requires at least $n^{-\frac{1}{4}}$-consistency of first
  - Typically we have wiggle room to get var decay at $n^{-\frac{1}{2}}$ rate (e.g. decrease the banc

For detailed exposition see:
- [Newey94, Ai-Chen'03]
- Section 8.3 of survey of [Newey-McFadden'94]
- Han Hong's Lecture notes on semi-parametric efficiency [ECO276 Stanford]

Dynamic Games

- Similar ideas extend to dynamic games with discounted payoffs
- Discrete state space $s_{\mathrm{t}} \in S$, private shock space $\epsilon_i \in V_i$, discrete or continuous actions $A_1, \dots, A_N$
- Steady state and at Markov-Perfect-Equilibria: mapping from states and shocks to actions.

$$V_i(s;\sigma,\theta) = E\left[\sum_{t=0}^{T} \beta^t \pi_i(\sigma(s_t,\nu_t), s_t, \epsilon_{it}) \,\middle|\, s_0 = s;\theta\right]$$

- Action specific i.i.d. profit shock and π_i is additively separable:

$$\pi_i(a,s,\epsilon_i) = \tilde{\pi}_i(a,s) + \epsilon_i(a_i)$$

- Define $v_i(a_i,s)$: "shockless" discounted expected equilibrium payoff.
- Player chooses action a_i if:

$$v_i(a_i,s) + \epsilon_i(a_i) \geq v_i(a_i',s) + \epsilon_i(a_i')$$

Dynamic Games: First Stage

[Bajari-Benkard-Levin'07]

- Suppose ϵ_i are extreme value and $v_i(0,s) = 0$, then
$$\log P_i(a_i|s) - \log P_i(0|s) = v_i(a_i, s)$$
- Non-parametrically estimate $\widehat{P}_i(a_i|s)$
- Invert and get estimate $\hat{v}_i(a_i, s)$

- We have a non-parametric first-stage estimate of the policy function:
$$\hat{\sigma}_i(s, \epsilon_i) = \underset{a_i \in A_i}{\operatorname{argmax}} \, \hat{v}_i(a_i, s) - \epsilon_i(a_i)$$
- Combine with non-parametric estimate of state transition probabilities
- Compute a non-parametric estimate of discounted payoff for each policy, state, parameter tuple: $\hat{V}_i(\sigma, s; \theta)$, by forward simulation

Dynamic Games: First Stage

[Bajari-Benkard-Levin'07]

- If payoff is linear in parameters:

$$\pi_i(a, s, \epsilon_i; \theta) = \Psi_{\mathrm{i}}(a, s, \epsilon_i) \cdot \theta$$

- Then:

$$V_i(\sigma, s; \theta) = W_i(\sigma, s) \cdot \theta$$

- Suffices to do only simulation for each (policy, state) pair and not for each parameter, to get first stage estimates $\widehat{W}_i(\sigma, s)$

Dynamic Games: Second Stage

[Bajari-Benkard-Levin'07]

- We know by equilibrium:

$$g(i,s,\sigma_i';\theta) = V_i(\sigma,s;\theta) - V_i(\sigma_i',\sigma_{-i};\theta) \geq 0$$

- Can use an extremum estimator:
  - Definite a probability distribution over (player, state, deviation) triplets
  - Compute expected gain from [deviation]₋ under the latter distribution

$$Q(\theta) = E[\min\{g(i,s,\sigma_i';\theta),0\}]$$

  - By Equilibrium $Q(\theta_0) = 0 = \min_\theta Q(\theta)$
- Do empirical analogue with estimate $\hat{g}$:

$$\hat{g}(i,s,\sigma_i';\theta) = \hat{V}_i(\hat{\sigma},s;\theta) - \hat{V}_i(\sigma_i',\hat{\sigma}_{-i};\theta)$$

coming from first stage estimates

- Two sources of error:
  - Error of $\hat{\sigma}$ and $\widehat{\mathrm{P}}(s'|s,a)$: $\sqrt{n}$-consistent, asymptotically normal, for discrete actions/states
  - Simulation error: can be made arbitrarily small by taking as many sample paths as you want

Notable Literature

- [Pakes-Ostrovsky-Berry’07], [Aguirregabiria-Mira’07], [Ackerberg-Benkard-Berry-Pakes’07], [Bajari-Hong-Chernozhukov-Nekipelov’09]
  - Provide similar but alternative two stage estimation approaches for dynamic games
  - [BHCN’09] can handle continuous states and semi-parametric estimation
  - All of them based on the inversion strategy proposed by [Hotz-Miller’93] for estimating single agent MDPs

Main take-aways

- When econometrician’s information is the same as each individuals (i.e. shocks are private to the players)
- Model can be viewed as fixed point of distribution over actions of players over the unobserved heterogeneity
- Can apply two-stage simulation approaches to avoid solving the fixed-point
- Data “designates” which equilibrium is selected
- Needs main assumption of “unique equilibrium in the data”

Auction Games: Identification and Estimation

FPA IPV: [Guerre-Perrigne-Vuong’00],

Beyond IPV: [Athey-Haile’02]

Partial Identification: [Haile-Tamer’03]

Comprehensive survey of structural estimation in auctions: [Paarsch-Hong’06]

First Price Auction: Non-Parametric Identification [Guerre-Perrigne-Vuong’00]

- Sealed bid first price auction
- Symmetric bidders: value $v_i \sim F$
- Observe all submitted bids
- Bids come from symmetric Bayes-Nash equilibrium

Non-parametric identification: Can we identify F from the distribution of bids G?

First Price Auction: Non-Parametric Identification [Guerre-Perrigne-Vuong'00]

- At symmetric equilibrium $s(\cdot)$:

$$v = \underset{z}{\operatorname{argmax}} \big(v - s(z)\big) F^{n-1}(z)$$

- First-order-condition:

$$\big(v - s(v)\big)(n-1) f(v) F^{n-2}(v) = s'(v) F^{n-1}(v) \Rightarrow v = s(v) + \frac{s'(v)F(v)}{(n-1)f(v)}$$

- By setting $b = s(v)$*:*

$$G(b) = \Pr\big[\tilde{b} \leq b\big] = \Pr[\tilde{v} \leq s^{-1}(b)] = F\big(s^{-1}(b)\big)$$

$$g(b) = F\big(s^{-1}(b)\big)' = \frac{f(s^{-1}(b))}{s'\big(s^{-1}(b)\big)}$$

- Change variables $v = s^{-1}(b)$ in FOC:

$$s^{-1}(b) = b + \frac{G(b)}{(n-1)g(b)}$$

First Price Auction: Non-Parametric Identification [Guerre-Perrigne-Vuong'00]

$$\text{hidden value } v = s^{-1}(b) = b + \frac{G(b)}{(n-1)g(b)} = \xi(b, G)$$

- If G strictly increasing continuous and with continuous density then:
$$F(v) = G\left(\xi^{-1}(v, G)\right)$$

- F can be identified when having access to G!

First Price Auction: Non-Parametric Estimation [Guerre-Perrigne-Vuong'00]

- Sequence of bid samples from each player $\left\{(B_{it})_{i=1}^{N}\right\}_{t=1}^{n}$
- Estimate G and g non-parametrically via standard approaches
- $\hat{G}$ is empirical CDF:

$$\hat{G}(\mathrm{b}) = \frac{1}{\mathrm{n} \cdot N} \sum_{i,t} 1\{B_{it} \leq b\}$$

- $\hat{g}$ is a kernel-based estimator:

$$\hat{g}(\mathrm{b}) = \frac{1}{\mathrm{n} \cdot N} \sum_{i,t} \frac{1}{h_n} K\left(\frac{B_{it} - b}{h_n}\right)$$

- K is any density function with zero moments up to m and bounded m-th moment

First Price Auction: Non-Parametric Estimation

[Guerre-Perrigne-Vuong'00]

- Given $\hat{G}$ and $\hat{g}$ we can now find the pseudo-inverse value of the player
- Use empirical version of identification formula

$$\hat{V}_{it} = B_{it} + \frac{\hat{G}(\mathrm{B_{it}})}{(\mathrm{n}-1)\,\hat{\mathrm{g}}(\mathrm{B_{it}})}$$

- Similarly define second-stage estimators of $\hat{F}$ and $\hat{f}$:**

$$\hat{F}(\mathrm{v}) = \frac{1}{\mathrm{n}\cdot N}\sum_{i,t} 1\{\hat{V}_{it} \leq v\}$$

$$\hat{f}(\mathrm{v}) = \frac{1}{\mathrm{n}\cdot N}\sum_{i,t} \frac{1}{h_n} K\left(\frac{\hat{V}_{it} - b}{h_n}\right)$$

** Need some modifications if one wants unbiasedness

Uniform Rates of Convergence

- Suppose f has uniformly bounded continuous first derivative
- If we observed values then uniform convergence rate of $\left(\frac{n}{\log(n)}\right)^{-1/3}$
  - From classic results in non-parametric regression [Stone'82]

- Now that only bids are observed, [GPV'00] show that best achievable is: $\left(\frac{n}{\log(n)}\right)^{-\frac{1}{5}}$
  - The density f depends on the derivative of g

What if only winning bid is observed?

- For instance in a Dutch auction
- CDF of winning bid is simply:

$$G_W(b) = G(b)^N \Rightarrow G(b) = \left(G_W(b)\right)^{\frac{1}{N}}$$

- Hence, densities are related as:

$$g(b) = \frac{1}{N} g_W(b)\left(G_W(b)\right)^{\frac{1}{N}-1}$$

- Thus G and g are identified from G_W and g_W
- Hence, can apply previous argument and identify F and f

What if only winning bid is observed?

- Alternatively, we can identify value of winner as:

$$v_W = b_W + \frac{1}{N-1}\frac{G(b_W)}{g(b_W)} = b_W + \frac{N}{N-1}\frac{G_W(b_W)}{g_W(b_W)}$$

- Thus we can identify distribution of highest value F_W and f_W
- Subsequently, use $\mathrm{F}(v) = \left(F_W(v)\right)^{\frac{1}{N}}$ and $f(v) = \frac{1}{N} f_W(v)\left(F_W(v)\right)^{\frac{1}{N}-1}$ to identify F and f
- This also gives an estimation strategy (two-stage estimator, similar to case when all bids observed)

Notable Literature

- [Athey-Haile'02]
  - Identification in more complex than independent private values setting.
  - Primarily second price and ascending auctions
  - Mostly, winning price and bidder is observed
  - Most results in IPV or Common Value model
- [Haile-Tamer'03]
  - Incomplete data and partial identification
  - Prime example: ascending auction with large bid increments
  - Provides upper and lower bounds on the value distribution from necessary equilibrium conditions
- [Paarsch-Hong'06]
  - Complete treatment of structural estimation in auctions and literature review
  - Mostly presented in the IPV model

Main Take-Aways

- Closed form solutions of equilibrium bid functions in auctions
- Allows for non-parametric identification of unobserved value distribution
- Easy two-stage estimation strategy (similar to discrete incomplete information games)
- Estimation and Identification robust to what information is observed (winning bid, winning price)
- Typically rates for estimating density of value distribution are very slow

Algorithmic Game Theory and Econometrics

Mechanism Design for Inference

Econometrics for Learning Agents

Mechanism Design for Data Science

[Chawla-Hartline-Nekipelov'14]

- Aim to identify a class of auctions such that:
  - By observing bids from the equilibrium of one auction
  - Inference on the equilibrium revenue on any other auction in the class is easy
  - Class contains auctions with high revenue as compared to optimal auction
- Class analyzed: Rank-Based Auctions
  - Position auction with weights $w_1 \geq \cdots \geq w_N \geq w_{N+1} = 0$
  - Bidders are allocated randomly to positions based only the relative rank of their bid
  - k-th highest bidder gets allocation x_k
  - Pays first price: $x_k b_k$
  - Feasibility: $\sum_{i=1}^{\tau} x_i \leq \sum_{i=1}^{\tau} w_i$
- For "regular" distributions, best rank-based auction is 2-approx. to optimal

Optimizing over Rank-Based Auctions

[Chawla-Hartline-Nekipelov'14]

- Every rank-based auction can be viewed as a new position auction with weights: $\bar{w}_i$ satisfying $\sum_{i=1}^{\tau} \bar{w}_i \leq \sum_{i=1}^{\tau} w_i$
- Thus auctioneer's optimization is over such modifications to the setting
- Each of these auctions is equivalent to running a mixture of k-unit auctions, where k-th unit auction run w.p. $p_k = \bar{w}_k - \bar{w}_{k+1}$
- To calculate revenue of any rank based auction, suffices to calculate expected revenue R_k of each k-th unit auction

Main question. Estimation rates of quantity R_k when observing bids from a given rank-based auction

Estimation analysis

[Chawla-Hartline-Nekipelov'14]

- Similar to the FPA equilibrium characterization used by [GPV'00]
- As always, write everything in quantile space
- With a twist: $q = F(v)$
- At symmetric equilibrium $s(\cdot)$:

$$b(q) = \underset{z}{\operatorname{argmax}}(v(q) - z)x\left(b^{-1}(z)\right)$$

- FOC:

$$v(q) = b(q) + \frac{b'(q)x(q)}{x'(q)}$$

- $x(q)$ and $x'(q)$ are known from the rules of the auction

Estimation

[Chawla-Hartline-Nekipelov'14]

- Need to estimate $b(q)$ and $b'(q)$ if we want to estimate $v(q)$
- Compared to [GPV'00]:
  - $v(q) = F^{-1}(q)$
  - $b(q) = G^{-1}(q), b'(q) = \frac{1}{g(G^{-1}(q))}$
  - Estimating $v(q), b(q), b'(q)$ the same as estimating F, G, g

- Main message. The quantity R_k for any k depends only on $b(q)$ and not on $b'(q)$! Leads to much faster rates.

Fast Convergence for Counterfactual Revenue [Chawla-Hartline-Nekipelov'14]

- The per agent revenue of a k-unit auction can be written as:
$$E[R(q)x_k'(q)]$$
- $R(q) = v(q)(1-q)$: single buyer revenue from price $v(q)$
- $x_k(q)$: probability player with quantile q is among k-highest
- Remember $v(q) = b(q) + \frac{b'(q)x(q)}{x\prime(q)}$
- Dependence on $b'(q)$ is of the form:
$$E\left[b'(q)\frac{x(q)(1-q)x_k'(q)}{x'(q)}\right]$$
- Integrating by parts:
$$E\left[b(q)\left(\frac{x(q)(1-q)x_k'(q)}{x'(q)}\right)'\right]$$
which depends only on $b(q)$ and on "exactly" known quantities

Yields $O\left(\frac{1}{\sqrt{N}}\right)$ convergence* of MSE, since $b(q)$ is essentially a CDF inverted

*Exact convergence depends inversely on $x'(q)$
Need to restrict to rank-based auctions where $x'(q) > \epsilon$ (e.g. mixing each k-unit auction with probability ϵ/n)

Take-away points

[Chawla-Hartline-Nekipelov'14]

- By isolating mechanism design to rank based auctions, we achieve:
  - Constant approximation to the optimal revenue within the class
  - Estimation rates of revenue of each auction in the class of $O\left(\sqrt{N}\right)$

- Allows for easy adaptation of mechanism to past history of bids

- [Chawla et al. EC'16]: allows for A/B testing among auctions and for a universal B test! (+improved rates)

Econometrics for Learning Agents

[Nekipelov-Syrgkanis-Tardos'15]

- Analyze repeated strategic interactions
- Finite horizon $t \in \{1, \dots, T\}$
- Players are learning over time
- Unlike stationary equilibrium, or stationary MPE, or static game

- Use no-regret notion of learning behavior:

$$\forall a_i': \sum_t \pi_i(a_i^t, a_{-i}^t; \theta) \geq \sum_t \pi_i\left(a_i', a_{-i}^t; \theta\right) - \epsilon$$

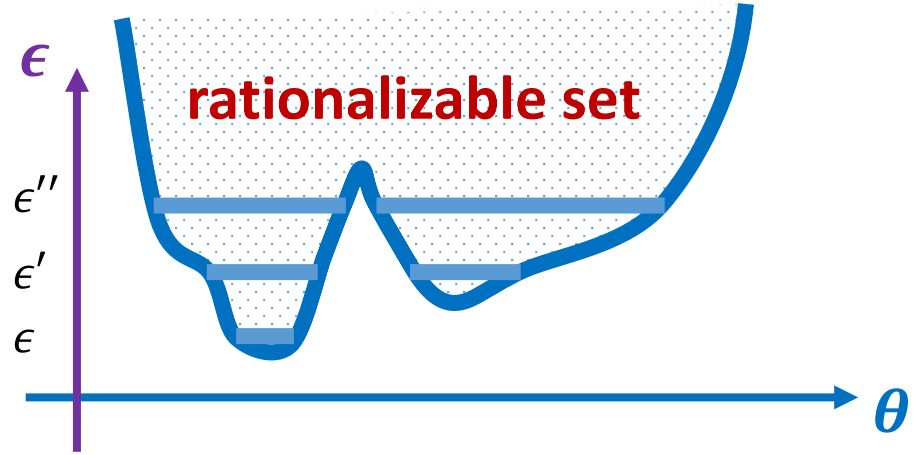
High-level approach

[Nekipelov-Syrgkanis-Tardos'15]

If we assume ϵ regret

$$\text{For all } a_i': \quad \underbrace{\frac{1}{T}\sum_t \pi_i(\boldsymbol{a}^{\boldsymbol{t}}; \boldsymbol{\theta})}_{\text{Current average utility}} \geq \underbrace{\frac{1}{T}\sum_t \pi_i\left(a_i', \boldsymbol{a}_{-i}^{\boldsymbol{t}}; \boldsymbol{\theta}\right)}_{\text{Average deviating utility from fixed action}} \underbrace{- \epsilon}_{\text{Regret}}$$

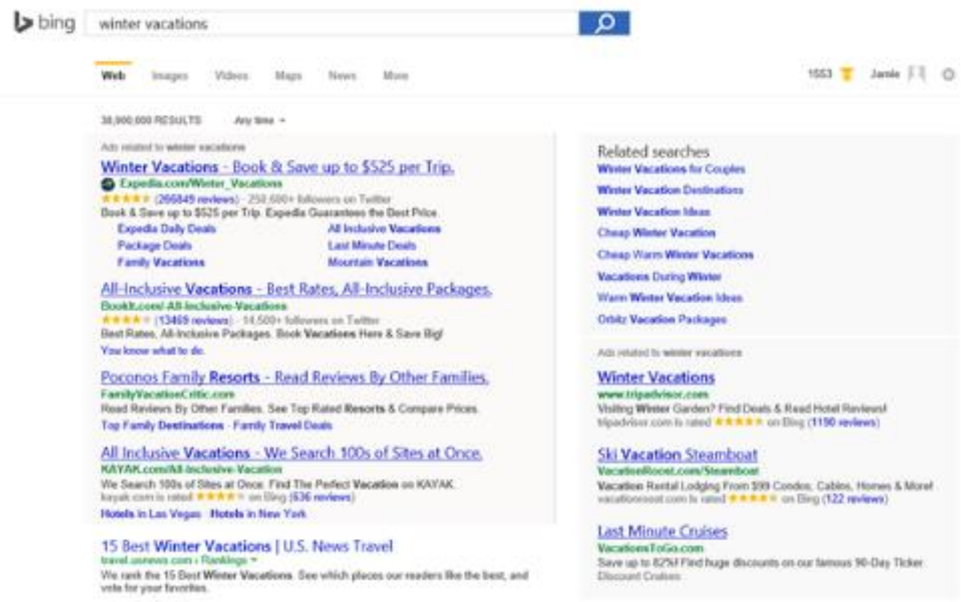
- Inequalities that unobserved $\boldsymbol{\theta}$ must satisfy
- Varying $\boldsymbol{\epsilon}$ we get the **rationalizable set of parameters**



Application: Online Ad Auction setting

[Nekipelov-Syrgkanis-Tardos'15]

- Each player has value-per-click v_i
- Bidders ranked according to a scoring rule
- Number of clicks and cost depends on position
- Quasi-linear utility



Value-Per-Click   Expected Payment

$$\pi_i(\mathbf{b}; \boldsymbol{v}_i) = \underbrace{\boldsymbol{v}_i}_{} \cdot x_i(\mathbf{b}) - p_i(\mathbf{b})$$

Expected click probability

Main Take-Aways of Econometric Approach

[Nekipelov-Syrgkanis-Tardos'15]

- Rationalizable set is convex
- Support function representation of convex set depends on a one dimensional function
- Can apply one-dimensional non-parametric regression rates
- Avoids complicated set-inference approaches

Comparison with prior econometric approaches:

- Behavioral learning model computable in poly-time by players
- Models error in decision making as unknown parameter rather than profit shock with known distribution
- Much simpler estimation approach than prior repeated game results
- Can handle non-stationary behavior

Potential Points of Interaction with Econometric Theory

- Inference for objectives (e.g. welfare, revenue, etc.) + combine with approximation bounds (see e.g. Chawla et al'14-16, Hoy et al.'15, Liu-Nekipelov-Park'16,Coey et al.'16)
- Computational complexity of proposed econometric methods, computationally efficient alternative estimation approaches
- Game structures that we have studied exhaustively in theory (routing games, simple auctions)
- Game models with combinatorial flavor (e.g. combinatorial auctions)
- Computational learning theory and online learning theory techniques for econometrics
- Finite sample estimation error analysis

AGT+Data Science

- Large scale mechanism design and game theoretic analysis needs to be data-driven

- Learning good mechanisms from data
- Inferring game properties from data
- Designing mechanisms for good inference
- Testing our game theoretic models in practice (e.g. Nisan-Noti’16)

References

Primer on Econometric Theory

- Newey-McFadden, 1994: *Large sample estimation and hypothesis testing*, Chapter 36, Handbook of Econometrics
- Amemiya, 1985: *Advanced Econometrics*, Harvard University Press
- Hong, 2012: Stanford University, Dept. of Economics, course ECO276, *Limited Dependent Variables*

Surveys on Econometric Theory for Games

- Ackerberg-Benkard-Berry-Pakes , 2006: *Econometric tools for analyzing market outcomes*, Handbook of Econometrics
- Bajari-Hong-Nekipelov, 2010: *Game theory and econometrics: a survey of some recent research*, NBER 2010
- Berry-Tamer, 2006: *Identification in models of oligopoly entry*, Advances in Economics and Econometrics

Complete Information Games

- Bresnahan-Reiss, 1990: *Entry in monopoly markets*, Review of Economic Studies
- Bresnahan-Reiss, 1991: *Empirical models of discrete games*, Journal of Econometrics
- Berry, 1992: *Estimation of a model of entry in the airline industry*, Econometrica
- Tamer, 2003: *Incomplete simultaneous discrete response model with multiple equilibria*, Review of Economic Studies
- Ciliberto-Tamer, 2009: *Market Structure and Multiple Equilibria in Airline Markets*, Econometrica
- Beresteanu-Molchanov-Molinari, 2011: *Sharp identification regions in models with convex moment predictions*, Econometrica
- Chernozhukov-Hong-Tamer, 2007: *Estimation and confidence regions for parameter sets in econometrics models*, Econometrica
- Bajari-Hong-Ryan, 2010: *Identification and estimation of a discrete game of complete information*, Econometrica

References

Dynamic Games of Incomplete Information

- Bajari-Benkard-Levin, 2007: *Estimating dynamic models of imperfect competition*, Econometrica
- Aguirregabiria-Mira, 2007: *Sequential estimation of dynamic discrete games*, Econometrica
- Pakes-Ostrovsky-Berry, 2007: *Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)*, RAND Journal of Economics
- Pesendorfer-Schmidt-Dengler, 2003: *Identification and estimation of dynamic games*
- Bajari-Chernozhukov-Hong-Nekipelov, 2009: *Non-parametric and semi-parametric analysis of a dynamic game model*
- Hotz-Miller, 1993: *Conditional choice probabilities and the estimation of dynamic models,* Review of Economic Studies

Static Games of Incomplete Information

- Bajari-Hong-Krainer-Nekipelov, 2006: *Estimating static models of strategic interactions*, Journal of Business and Economic Statistics

Semi-Parametric two-stage estimation $\sqrt{\boldsymbol{n}}$-consistency

- Hong, 2012: ECO276, *Lecture 5: Basic asymptotic for $\sqrt{n}$ Consistent semiparametric estimation*
- Robinson, 1988: *Root-n-consistent semiparametric regression*, Econometrica
- Newey, 1990: *Semiparametric efficiency bounds*, Journal of Applied Econometrics
- Newey, 1994: *The asymptotic variance of semiparametric estimators*, Econometrica
- Ai-Chen, 2003: *Efficient estimation of models with conditional moment restrictions containing unknown functions*, Econometrica
- Chen, 2008: *Large sample sieve estimation of semi-nonparametric models* Chapter 76, Handbook of Econometrics

References

Auctions

- Guerre-Perrigne-Vuong, 2000: *Optimal non-parametric estimation of first-price auctions*, Econometrica
- Haile-Tamer, 2003: *Inference in an incomplete model of English auctions*, Journal of Political Economy
- Athey-Haile, 2007: *Non-parametric approaches to auctions*, Handbook of Econometrics
- Paarsch-Hong, 2006: *An introduction to the structural econometrics of auction data*, The MIT Press

Algorithmic Game Theory and Econometrics

- Chawla-Hartline-Nekipelov, 2014: *Mechanism design for data science*, ACM Conference on Economics and Computation
- Nekipelov-Syrgkanis-Tardos, 2015: *Econometrics for learning agents*, ACM Conference on Economics and Computation
- Chawla-Hartline-Nekipelov, 2016: *A/B testing in auctions*, ACM Conference on Economics and Computation
- Hoy-Nekipelov-Syrgkanis, 2015: *Robust data-driven guarantees in auctions*, Workshop on Algorithmic Game Theory and Data Science