Robust Structural Estimation

Jiaming Mao and Jingzhi Xu

Xiamen University

December 7, 2019

Motivation

Two paradigms of causal inference in applied economic research:

- **Structural**: estimation of structural models
- **Reduced-form**: estimation of causal effects using statistical models

Structural Estimation

Structural model: causal models based on economic theory.

- Complete models may specify *preferences*, *technology*, *information* available to agents, *constraints* under which they operate, and the *rules of interaction* in market and social settings.
- More generally: any causal models that use economic theory to specify the functional form of causal relationships.

Structural Estimation

- Rooted in the Cowles commission tradition of econometrics as "a branch of economics in which *economic theory* and *statistical method* are fused in the analysis of numerical and institutional data" (Hood and Koopmans, 1953).

- Ragnar Frisch, founding member of the Econometric Society, first editor-in-chief of Econometrica:

  Intermediate between mathematics, statistics, and economics, we find a new discipline ... so far, we have been unable to find any better word than "econometrics". We are aware of the fact that in the beginning somebody might misinterpret this word to mean economic statistics only. But ... we believe that it will soon become clear to everybody that the society is interested in economic theory just as much as in anything else.

Structural Estimation

Pros:

- External validity
- Learns "deep" parameters that are *policy-invariant*, allowing counterfactual simulations
- Can do welfare analysis
- Rosenzweig and Wolpin (2000), Sims (2010), Keane (2010), Wolpin (2013), Heckman (2015), Heckman and Vytlacil (2007), Deaton and Cartwright (2018)

Cons:

- Strong often unrealistic assumptions; Identification by functional form

Reduced-Form

- Historically: alternative representation of structural models. Given a structural model $g(x, y, \epsilon) = 0$, where x is exogenous, y is endogenous, ϵ is unobserved, write y as a function of x and ϵ: $y = f(x, \epsilon)$. f is the reduced-form of g.

- Today: any methods that focus on the estimation of treatment effects, using theory to guide research design in order to achieve *identification*, but not to specify functional forms.

  - Rust (2014): "At the risk of oversimplifying, empirical work that takes theory "seriously" is referred to as *structural econometrics* whereas empirical work that avoids a tight integration of theory and empirical work is referred to as *reduced form econometrics*."

Reduced-Form

Pros:

- Good designs allow for credible identification of causal effects
- Internal validity
- Angrist and Pischke (2008, 2010), Panhans and Singleton (2017)

Cons:

- Focused on learning population-specific, often local, treatment effects; lack of external validity; often unclear about assumptions relied upon to draw causal inference

Competing Paradigms

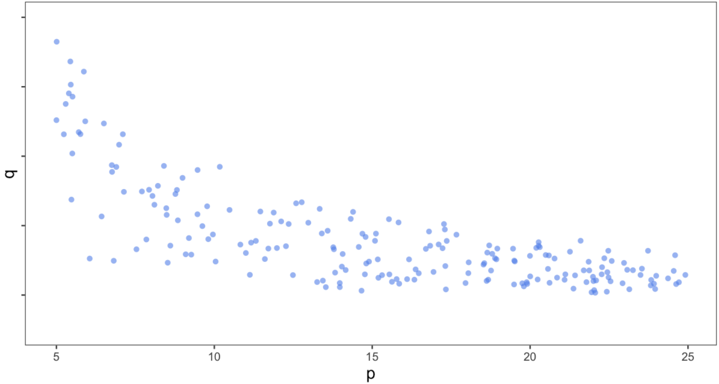
- Today, “the division between structural and reduced-form methods has split the economics profession into two camps whose research programs have evolved almost independently despite focusing on similar questions.” (Chetty, 2009)

This Paper

- This paper provides a unifying perspective of structural and reduced-form estimation as generative and discriminative causal inference methods, and proposes a set of **robust structural estimators** that combine the two approaches to estimate causal effects of interest:

  1. **Doubly robust RF-ST estimator**: consistent as long as either the structural or the reduced-form model is correctly specified.
  2. **Ensemble RF-ST estimator**: can out-perform either model when both are mis-specified.

This Paper

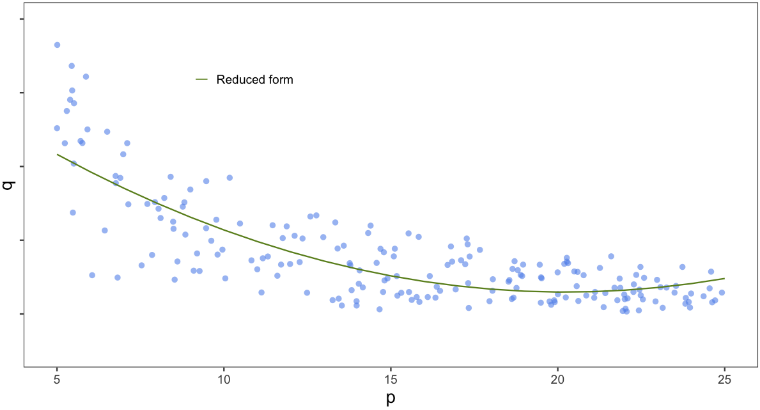


Example: Consumption Data

This Paper

Assuming changes in p are exogenous, we can fit the following demand curve:

$$q = \beta_0 + \beta_1 p + \beta_2 p^2 + e \tag{1}$$



This Paper

However, the observed consumption data is generated by the following model:

$$\max u\left(q, q^{o}\right) \text{ s.t. } pq + p^{o} q^{o} \leq I \tag{2}$$

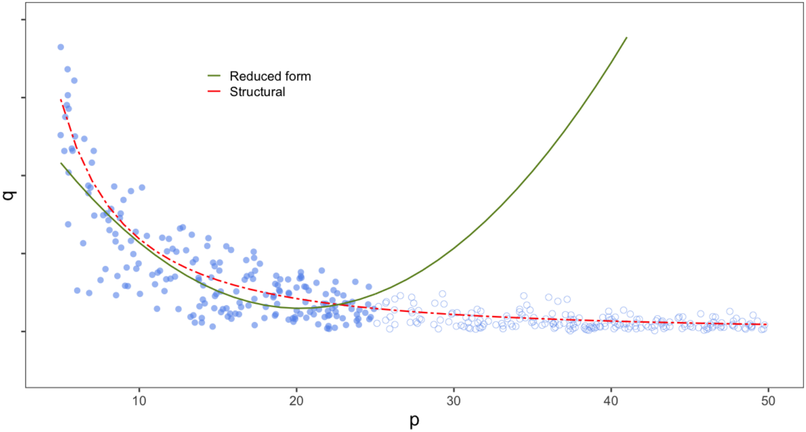
, where o stands for an outside good, I is income, and

$$u\left(q, q^{o}\right) = \left[0.5 q^{\rho} + 0.5 \left(q^{o}\right)^{\rho}\right]^{\frac{1}{\rho}}$$

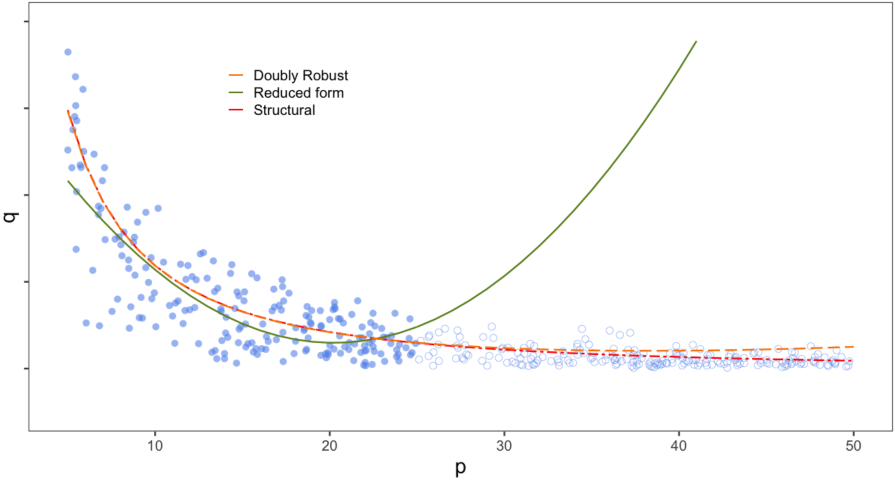
, with $\rho = 0.5$, implying an elasticity of substitution $\sigma = 2$.

This Paper

The reduced-form fit (1) does not look good when more (a larger range of) data are observed...



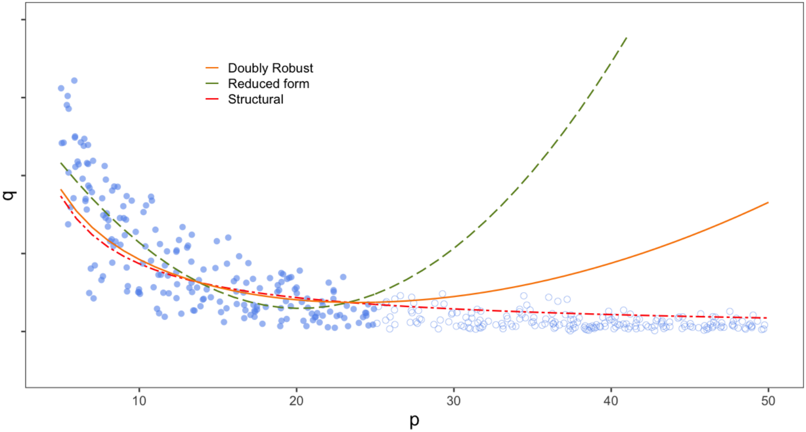
This Paper



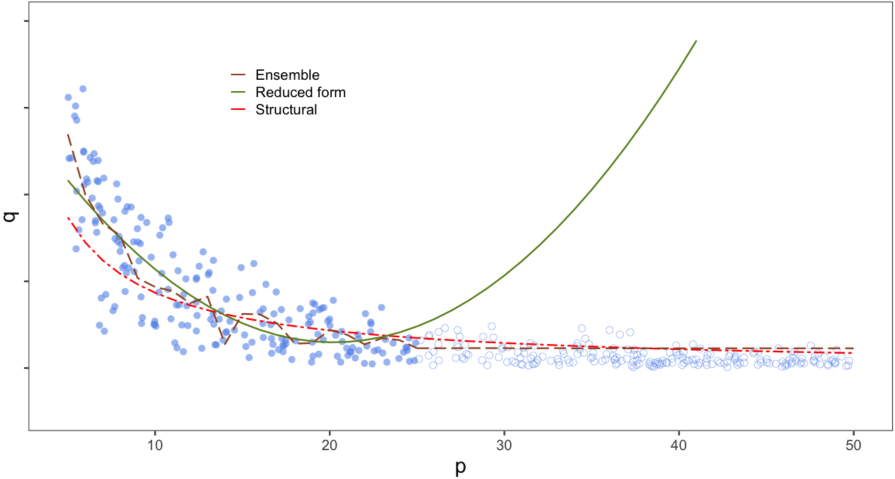
Doubly Robust RF-ST

This Paper

What if we estimate an incorrect structural model that assumes the utility function is Cobb-Douglas?



This Paper



Ensemble RF-ST

Related Literature

- Bound estimation: Manski (1990, 1997, 2013), Chernozhukov et al. (2007), Chernozhukov et al. (2013)
- Robust control: Hansen and Sargent (2001, 2010, 2019)
- Doubly robust estimation: Robins et al. (2000), Funk et al. (2010), Okui et al. (2012). Benkeser et al. (2017), Lewbel et al. (2019)
- Ensemble methods: van der Laan et al. (2007)
- Combining reduced-form and structural estimates for welfare analysis: Chetty (2009)

Outline

1. Discriminative vs. generative causal inference
2. The doubly robust RF-ST estimator
3. The ensemble RF-ST estimator
4. Simulations
    1. First-price auction
    2. Dynamic entry-exit
    3. Simultaneous equations model

Discriminative vs. Generative Causal Inference

Definition

A **causal model** $\mathcal{M}$ for a set of random variables $X = \{x_1, \ldots, x_n\}$ is $\mathcal{M} = (\mathcal{H}, \mathcal{G})$, where $\mathcal{H}$ is a generative statistical model for the joint distribution $p(x_1, \ldots, x_n)$, and $\mathcal{G}$ is the **causal structure** governing $\{x_1, \ldots, x_n\}$, which describes the causal relationships among the variables and can be represented by a causal graph.

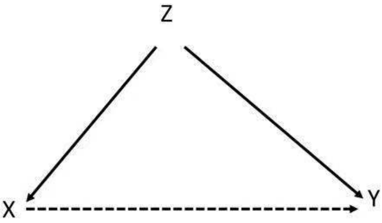
- A structural model is a causal model $\mathcal{M} = (\mathcal{H}, \mathcal{G})$: let $X^E \subset X$ be the variables that are *exogenous* to $\mathcal{G}$. The structural model specifies the statistical distributions of X^E as well as the functional forms governing $\mathcal{G}$, which together determine the joint distribution $p(X; \mathcal{H})$.

Discriminative vs. Generative Causal Inference

- The goal of reduced-form estimation is to learn a specific causal effect $p\left(x_j \middle| \mathrm{do}\left(x_i\right)\right)$ for $(x_i, x_j) \subset X$ (Pearl, 2009)[1].
- The reduced-form approach is a two-stage process:
  1. Choose a research design based on $\mathcal{G}$ that allows the identification of $p\left(x_j \middle| \mathrm{do}\left(x_i\right)\right)$.
     - Research design: regression, matching, IV, quasi-experimental, ...
  2. Use a statistical model $\mathcal{Q}$ to estimate $p\left(x_j \middle| \mathrm{do}\left(x_i\right)\right)$ based on the chosen design.
- We denote the target $p\left(x_j \middle| \mathrm{do}\left(x_i\right); \mathcal{G}, \mathcal{Q}\right)$

[1] Equivalently, using the potential outcomes notation from the Rubin causal model, $p\left(x_j \middle| \mathrm{do}\left(x_i = a\right)\right) = p\left(x_j^{x_i = a}\right)$.

Discriminative vs. Generative Causal Inference



$\mathcal{G}$

Structural

- Estimate a structural model $g(x, y, z; \theta) = 0$

Reduced-form (target: $E[y|\text{do}(x)]$)

1. Based on $\mathcal{G}$, $E[y|\text{do}(x)] = \int E[y|x, z]\, p(z)\, dz$
2. Estimate $E[y|x, z]$ using a statistical model, e.g., $y = \beta_0 + \beta_1 x + \beta_2 z + e$.

Discriminative vs. Generative Causal Inference

- Because structural estimation learns the entire causal model, once a model is estimated, we can use it to derive $p\left(x_j \middle| \operatorname{do}(x_i)\right)$ for any $\{x_i, x_j\} \subset X$.

- Analogy with the two paradigms of statistical modeling: structural estimation can be considered a generative approach to causal inference, while the reduced-form a discriminative approach.

Discriminative vs. Generative Causal Inference

Discriminative Model	Target
Discriminative statistical model	$p\left(x_j \mid x_i; \mathcal{Q}\right)$
Reduced-form estimation	$p\left(x_j \mid \mathrm{do}\left(x_i\right); \mathcal{G}, \mathcal{Q}\right)$

Generative Model	Target
Generative statistical model	$p\left(x_1, \ldots x_n; \mathcal{H}\right)$
Structural estimation	$p\left(x_1, \ldots x_n; \mathcal{M}\right)$

Towards a Robust Causal Estimator

Goal: construct a discriminative causal estimator for $E[y|\,\text{do}(x)]$.

- Since we can always derive a discriminative estimator from a generative one, $E[y|\,\text{do}(x)]$ can be estimated using both structural and reduced-form estimation.
- Idea: we can combine the two to have more external validity than the reduced-form and more robust against model mis-specification than the structural.

The Doubly Robust RF-ST Estimator

Classic Doubly Robust: outcome regression (OR) and inverse-probability weighting (IPW)

Assuming z satisfies the back-door criterion (Pearl, 2009)[2],

$$\widehat{E}\left[y\,|\mathrm{do}\,(x=a)\right] = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{y_i\mathcal{I}\,(x_i=a)}{\widehat{p}\,(x=a\,|z=z_i\,)} + \left(1-\frac{\mathcal{I}\,(x_i=a)}{\widehat{p}\,(x=a\,|z=z_i\,)}\right)\times\widehat{E}\left[y\,|x=a,z=z_i\,\right]\right]$$

, where $\frac{1}{N}\sum_{i=1}^{N}\widehat{E}\left[y\,|x=a,z=z_i\,\right]$ and $\frac{1}{N}\sum_{i=1}^{N}\left[\frac{y_i\mathcal{I}(x_i=a)}{\widehat{p}(x=a|z=z_i\,)}\right]$ are respectively the OR and IPW estimates for $\widehat{E}\left[y\,|\mathrm{do}\,(x=a)\right]$.

[2] Equivalently, $y^{x=a} \perp x|\, z$.

The Doubly Robust RF-ST Estimator

Generalized Doubly Robust (Lewbel et al., 2019): given data $\{d_i = (x_i, y_i, z_i)\}$, and two estimators g and f for causal effect τ with moment functions $m_g(d_i, \tau, \beta)$ and $m_f(d_i, \tau, \gamma)$ satisfying *either* $E[m_g(d_i, \tau_0, \beta_0)] = 0$ *or* $E[m_f(d_i, \tau_0, \gamma_0)] = 0$, define

$$Q_g = \kappa_g^{-1}\overline{m}_g(\tau, \beta)'\Omega_g\overline{m}_g(\tau, \beta),\ Q_f = \kappa_f^{-1}\overline{m}_f(\tau, \gamma)'\Omega_f\overline{m}_f(\tau, \gamma)$$

, where $\{\Omega_g, \Omega_f\}$ are positive definite matrices, and $\{\kappa_g, \kappa_f\}$ are d.o.f.

Let

$$w_g = \frac{Q_f}{Q_f + Q_g}$$

Then

$$\widehat{\tau} = \widehat{w}_g\widehat{\tau}_g + (1 - \widehat{w}_g)\,\widehat{\tau}_f$$

The Doubly Robust RF-ST Estimator

- Based on causal structure $\mathcal{G}$, assume z satisfy the back-door criterion. Let g be the reduced-form model for $E[y|x,z]$. Let $\mathcal{M}$ be the structural model that describes the mechanisms governing these variables. Let $(\widehat{\tau}_g, \widehat{\tau}_{\mathcal{M}})$ be the estimated causal effect according to g and $\mathcal{M}$.
- We apply the method for constructing the generalized doubly robust estimator to combining g and $\mathcal{M}$.
- To avoid overfitting, use validation sets to compute moment functions.

The Doubly Robust RF-ST Estimator

Theorem (Lewbel et al., 2019)

$\widehat{\tau} \to^P \tau_0$ as long as $\widehat{\tau}_g \to^P \tau_0$ or $\widehat{\tau}_{\mathcal{M}} \to^P \tau_0$, in which case under regularity conditions,

$$\sqrt{n}\left(\widehat{\tau} - \tau_0\right) \to^d \mathcal{N}\left(0, V\right)$$

$$\frac{1}{N}\sum_i \widehat{\eta}_i \widehat{\eta}_i' \to^P V$$

$$\widehat{\eta}_i = \widehat{w}_g \widehat{\eta}_i^g + \left(1 - \widehat{w}_g\right) \widehat{\eta}_i^{\mathcal{M}}$$

, where $\left\{\widehat{\eta}_i^g, \widehat{\eta}_i^{\mathcal{M}}\right\}$ are estimated GMM influence functions.

The Ensemble RF-ST Estimator

Let $f(x) = E_{\mathcal{M}}[y|x,z]$ be the implied CEF according to $\mathcal{M}$. The ensemble RF-ST estimates

$$y = h\left(\widehat{g}(x,z), \widehat{f}(x,z); \theta\right) + \epsilon$$

$\Rightarrow$

$$\widehat{E}[y\,|\mathrm{do}(x)] = \int h\left(\widehat{g}(x,z), \widehat{f}(x,z); \widehat{\theta}\right) p(z)\, dz$$

The Ensemble RF-ST Estimator

- **Linear ensemble**:

$$h = \theta_0 + \theta_1 \widehat{g}(x,z) + \theta_2 \widehat{f}(x,z)$$

  - Stacking:

$$\widehat{\theta} = \arg\min_{\theta} \sum_{i=1}^{N} \left(y_i - \theta_0 - \theta_1 \widehat{g}^{-i}(x_i, z_i) - \theta_2 \widehat{f}^{-i}(x_i, z_i) \right)^2$$

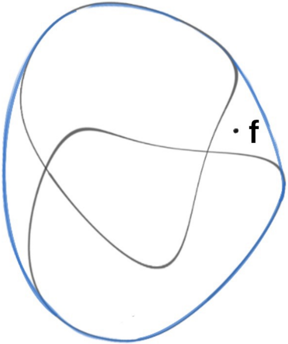
  , where $\widehat{g}^{-i}(x_i, z_i)$ and $\widehat{f}^{-i}(x_i, z_i)$ are trained minus the i^{th} obs.

- **Random forest ensemble**: h a random forest
  - Allows weighs to vary in each region of the predictor space, depending on which one – the reduced form or the structural – works better.

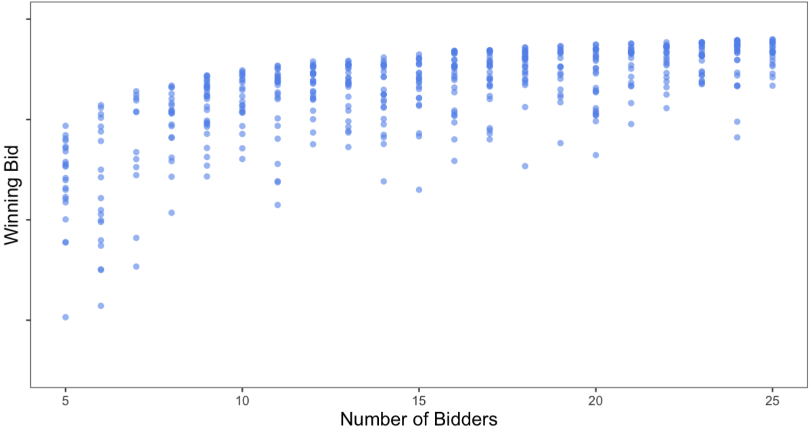
The Ensemble RF-ST Estimator: Intuition

- Ng and Jordan (2002):
  - Discriminative learning has lower asymptotic error.
  - Generative learning may approach asymptotic error faster.
  - There can be two distinct "regimes of performance" as sample size increases, one in which each does better.
- Dietterich (2000): Three reasons why ensemble methods work:
  1. Statistical: combining predictions lowers variance
  2. Computational: combining local optima produced by local search
  3. Representational: combined and enlarged hypothesis set capable of producing a better approximation to target function.

The Ensemble RF-ST Estimator: Intuition



Experiment: First-Price Auction



First-price Sealed-bid Auctions for Identical Goods

Experiment: First-Price Auction

Structural Model

- N risk-neutral bidders.
- Independent private value $v_i \sim^{i.i.d.} F(.)$
- Each bidder knows her own v_i and the distribution F, but not the V_i of others.
- Observed bids are the Bayesian Nash Equilibrium outcome of the game.
- Equilibrium bidding strategy:

$$b_i = v_i - \frac{1}{F(v_i)^{N-1}} \int_0^{v_i} F(x)^{N-1} dx$$

Experiment: First-Price Auction

$$\text{Target: } E\left[\, b_{\max} \middle|\, N\right]$$

- Reduced-form:

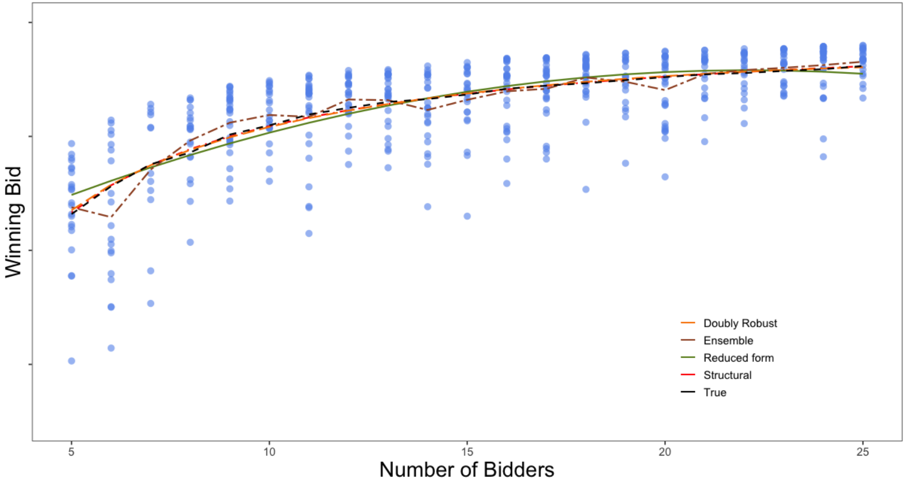
$$b_{\max} = g\left(N; \beta\right) + e$$

- Structural estimation: Guerre et al. (2000)

Experiment: First-Price Auction

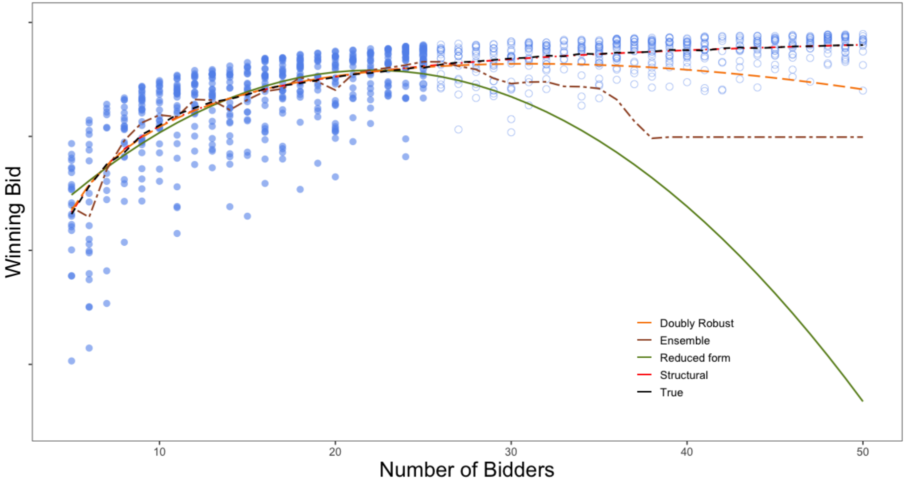
Experiment	True Mechanism	Structural Model
1	$v_i \sim^{i.i.d.} U(0,1)$	$v_i \sim^{i.i.d.} U(0,1)$
2	$v_i \sim^{i.i.d.} Beta(2,5)$	$v_i \sim^{i.i.d.} U(0,1)$
3	irrational bidding: $b_i = b_i^* + e_i$	rational bidding

Experiment: First-Price Auction



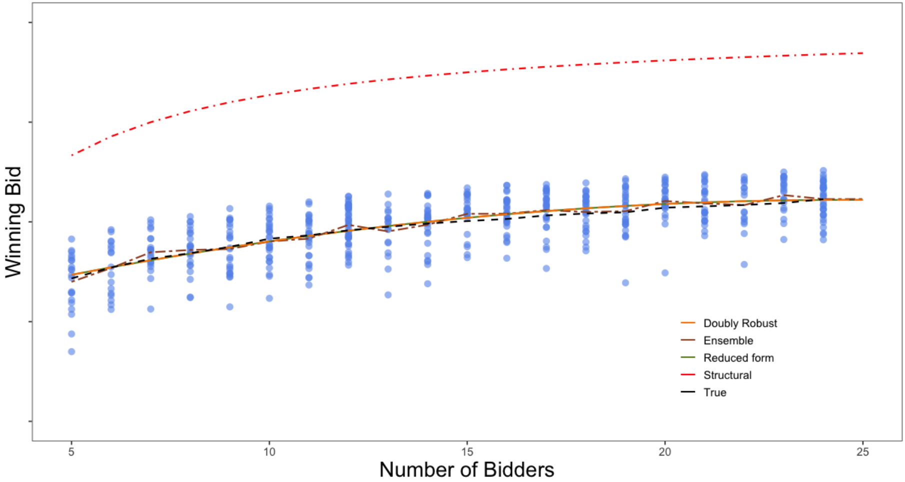
Experiment 1: in-sample

Experiment: First-Price Auction



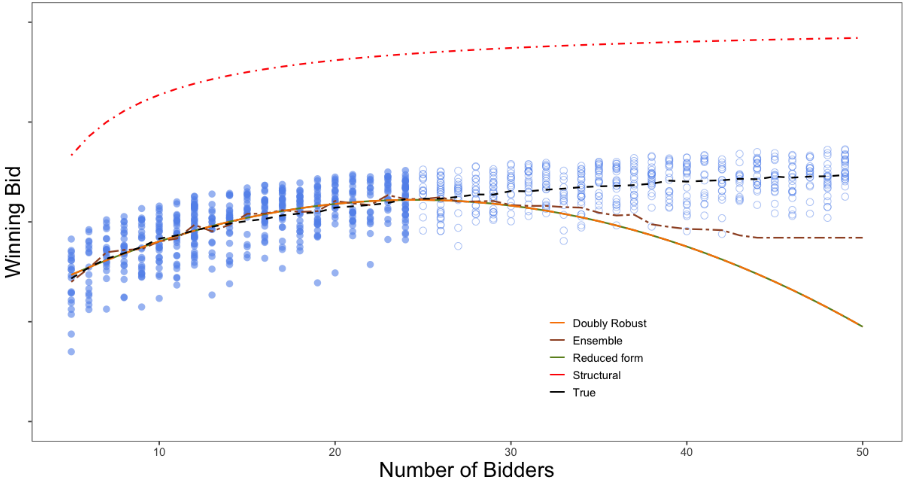
Experiment 1: out-of-sample

Experiment: First-Price Auction



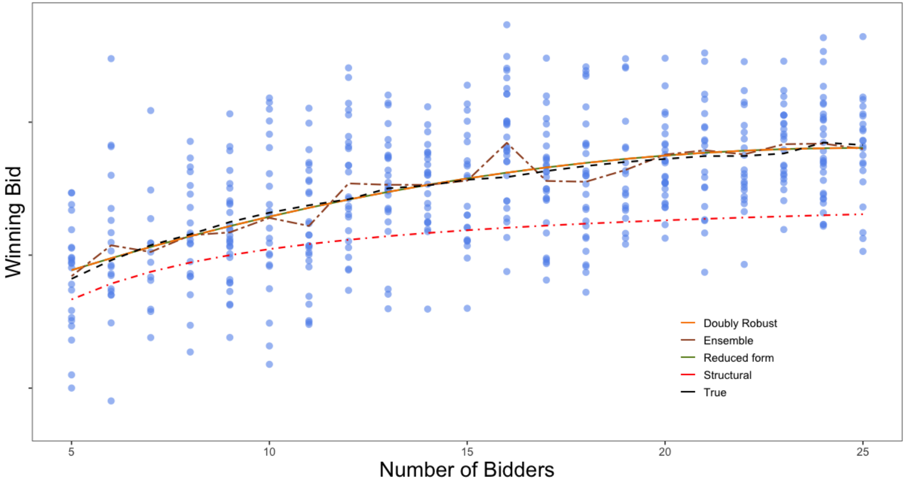
Experiment 2: in-sample

Experiment: First-Price Auction



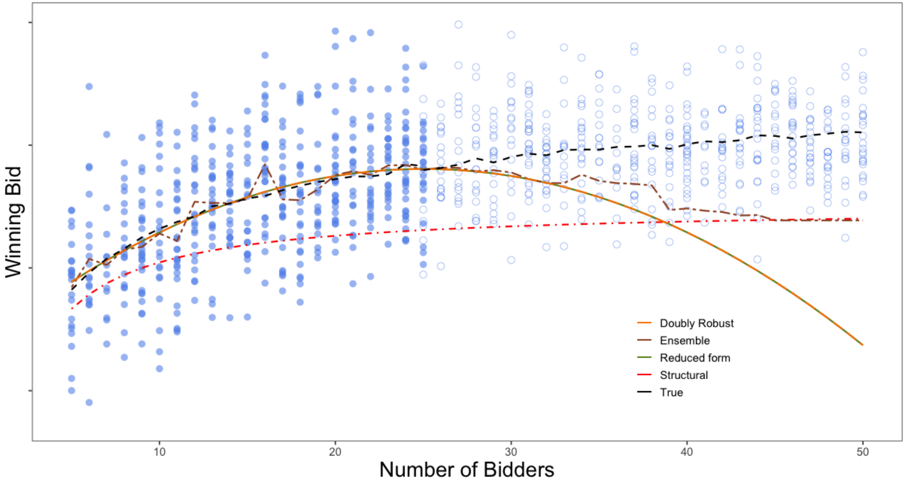
Experiment 2: out-of-sample

Experiment: First-Price Auction



Experiment 3: in-sample

Experiment: First-Price Auction

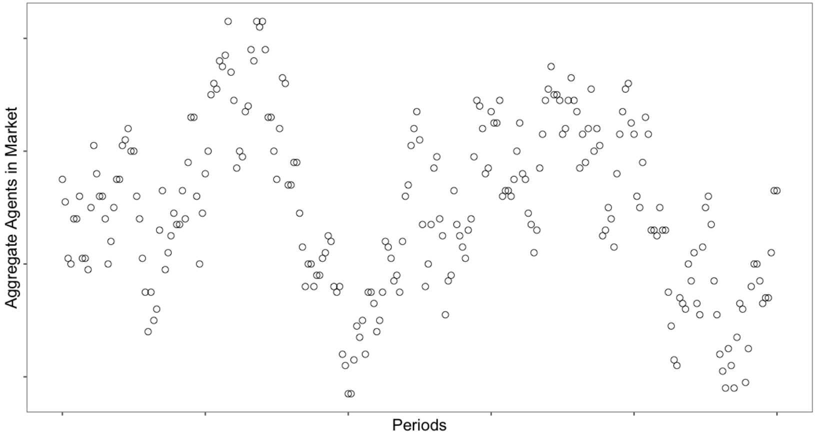


Experiment 3: out-of-sample

Experiment: First-Price Auction

Experiment	Error (ℓ_2)	ST	RF	DR	Ensemble
1	In	.00387	.00409	.00389	.00361
	Out	.00065	.09324	.00335	.01408
2	In	.13195	.00295	.00295	.00285
	Out	.12679	.03372	.03370	.01253
3	In	.05598	.03137	.03137	.03019
	Out	.13150	.13147	.07803	.04979

Experiment: Dynamic Entry and Exit



Experiment: Dynamic Entry and Exit

Structural Model

- N agents in a market. In each period, incumbents decide whether to remain in the market and potential entrants decide whether to enter. The return to operating in the market in time t is R_t (exogenous time-varying payoff).
- Let entry status be indicated by $(0, 1)$. The time-t flow utility of agent i who is in state j in time $t-1$ and state k in time t is:

$$u_{i,t}^{j,k} = \pi\left(R_t, c^{j,k}\right) + e_{i,t}^{k}$$

, where $c^{j,k}$ is the cost of transitioning from state j to k, and $e_{i,t} = (e_{i,t}^{0}, e_{i,t}^{1}) \sim^{i.i.d.} \text{Gumbel}\,(0, 1)$ are idiosyncratic shocks.

Experiment: Dynamic Entry and Exit

Structural Model (cont.)

- At the beginning of each period t, each agent receives an idiosyncratic shock $e_{i,t}$ and chooses her entry status $d_{i,t}$ by solving the following problem:

$$d_{i,t} = \arg\max_k \left\{ \pi_t^{j,k} + e_{i,t} + \beta \mathbb{E}_t \overline{V}_{t+1}^k \right\}$$

, where

$$\overline{V}_t^j = \mathbb{E}_e \left[V_{i,t}^j \left(e_{i,t} \right) \right]$$

$$V_{i,t}^j \left(e_{i,t} \right) = \max_k \left\{ \pi_t^{j,k} + e_{i,t} + \beta \mathbb{E}_t \overline{V}_{t+1}^k \right\}$$

Experiment: Dynamic Entry and Exit

Let Q_t denote the total number of entries in period t.

$$\text{Target: } E\left[Q_t \middle| R_t\right]$$

- Reduced-form:

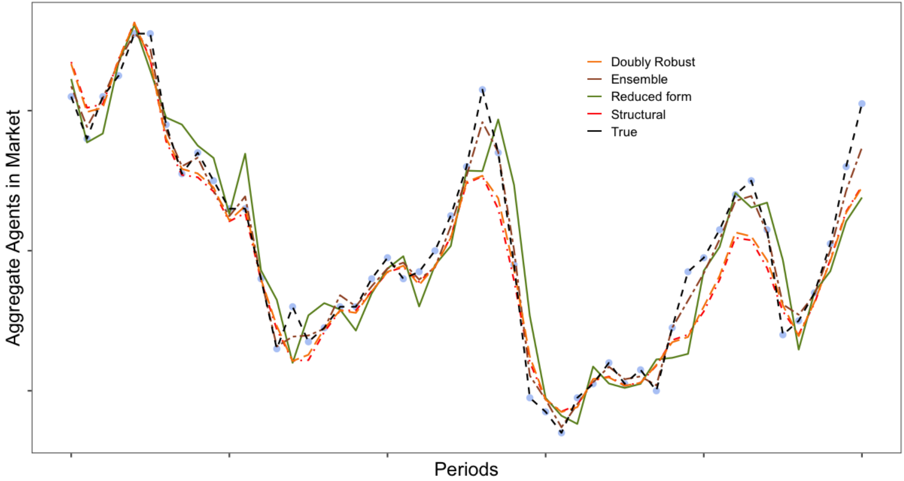
$$Q_t = \text{ARMAX}\left(R_t\right) + e$$

- Structural estimation: Hotz-Miller CCP (Hotz and Miller (1993), Arcidiacono and Miller (2011))

Experiment: Dynamic Entry and Exit

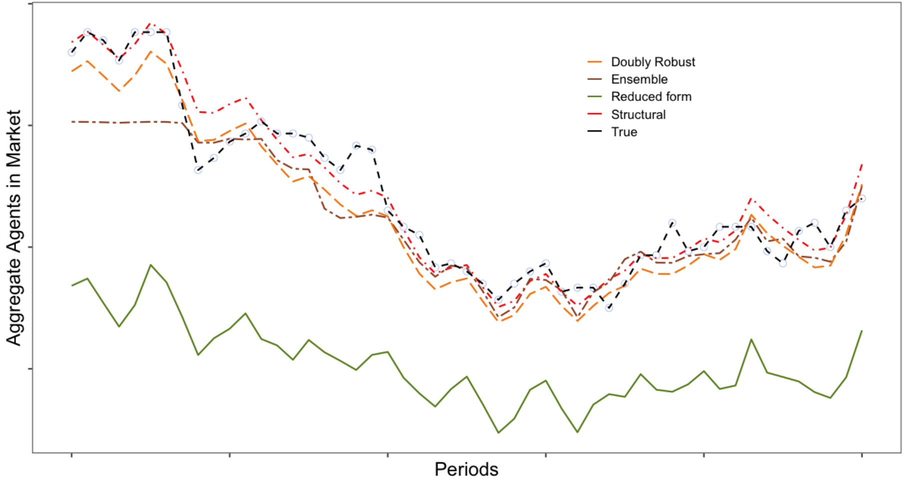
Experiment	True Mechanism	Structural Model
1	rational expectation	rational expectation
2	adaptive expectation	rational expectation
	(agents assume $R_{t'} = R_t \ \forall t' > t$)	
3	myopic	rational expectation

Experiment: Dynamic Entry and Exit



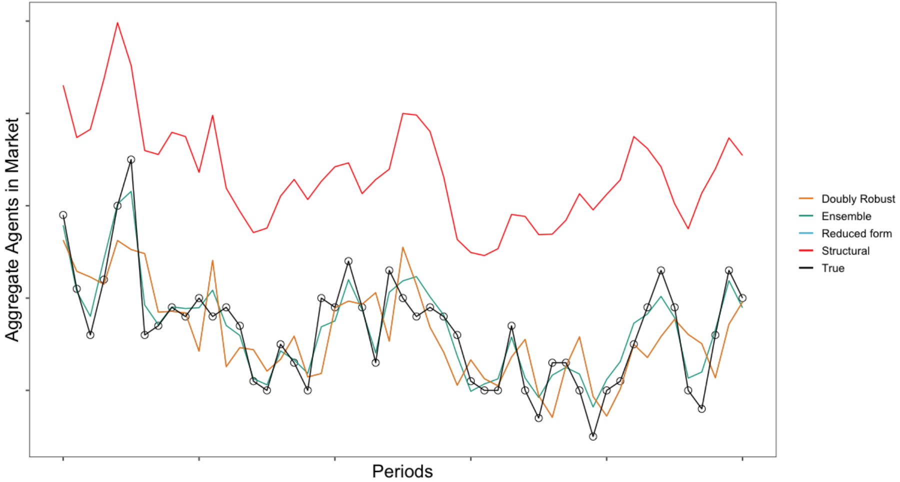
Experiment 1: in-sample

Experiment: Dynamic Entry and Exit



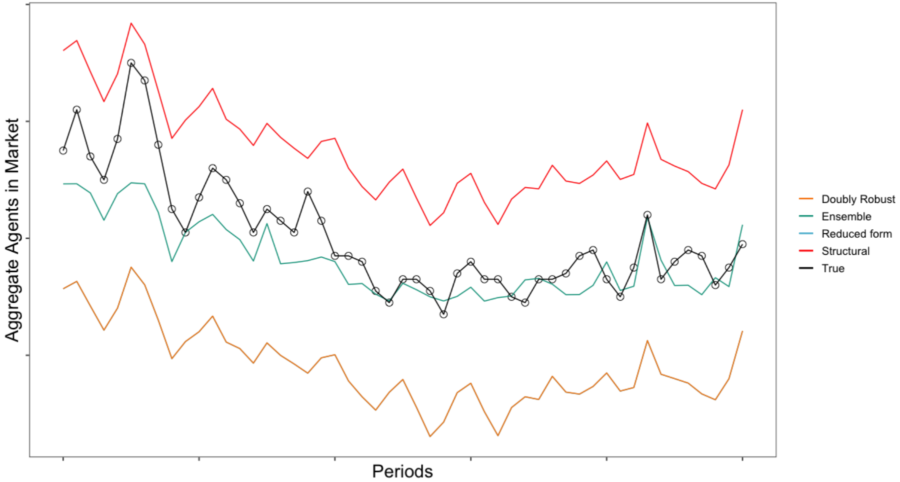
Experiment 1: out-of-sample

Experiment: Dynamic Entry and Exit



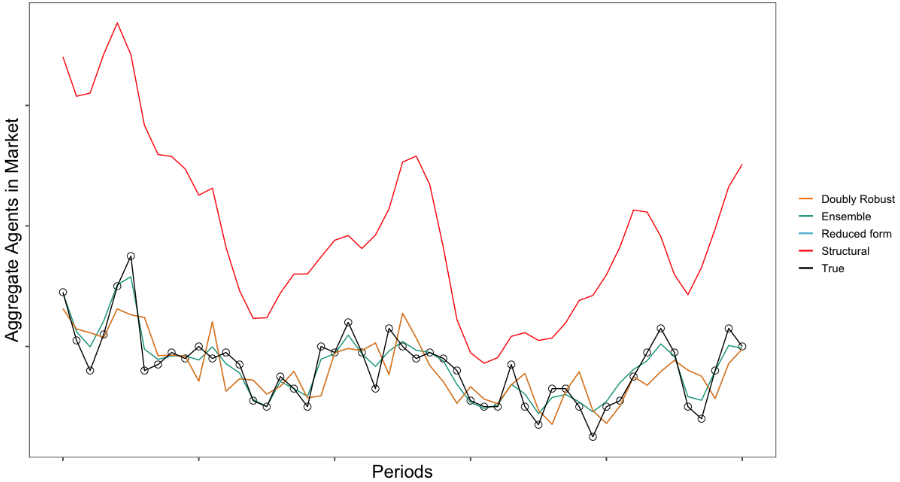
Experiment 2: in-sample

Experiment: Dynamic Entry and Exit



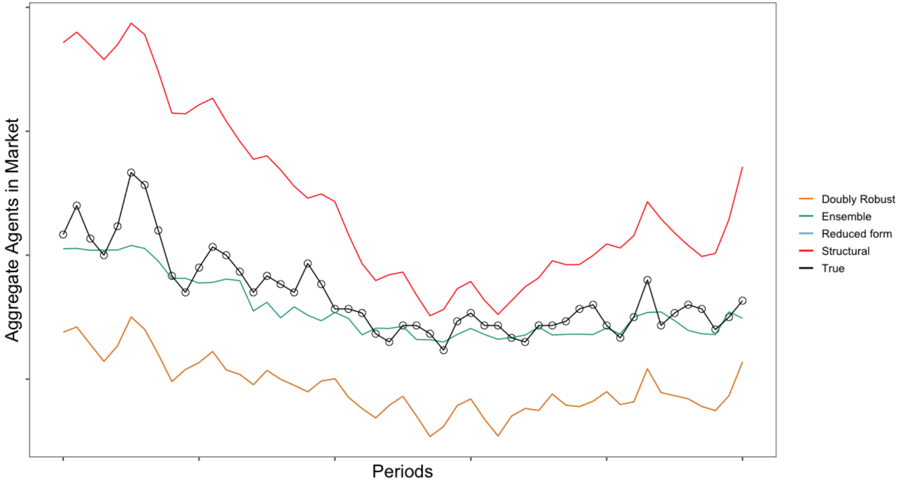
Experiment 2: out-of-sample

Experiment: Dynamic Entry and Exit



Experiment 3: in-sample

Experiment: Dynamic Entry and Exit



Experiment 3: out-of-sample

Experiment: Dynamic Entry and Exit

Experiment	Error (ℓ_2)	ST	RF	DR	Ensemble
1	In	23.61	18.84	20.54	4.65
	Out	23.31	318.88	817.47	71.33
2	In	247.21	15.57	20.71	3.79
	Out	238.08	123.71	488.02	51.73
3	In	916.406	16.542	20.717	3.71
	Out	945.65	133.63	488.00	38.10

Future Work

- Work in progress: construct robust structural estimators for counterfactual policy and welfare analyses.