

Long-term causal effects via behavioral game theory

Panagiotis (Panos) Toulis

University of Chicago, Booth School

February 11, 2017

Better policies from Dem. or Rep. presidents?

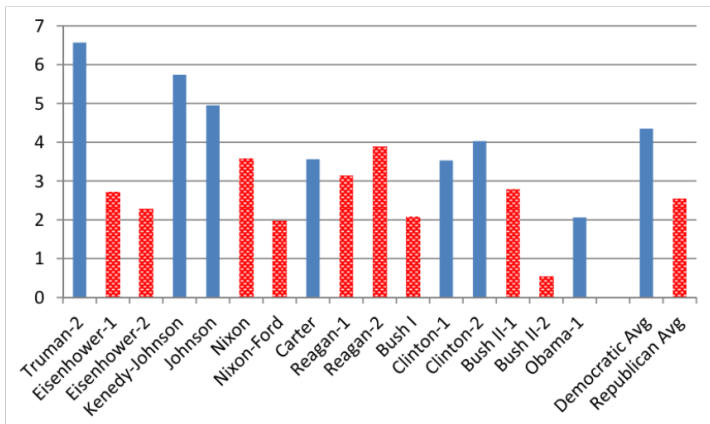


Figure: y-axis: %GDP growth; x-axis: incumbent president; color: party affiliation of incumbent president.

Better policies from Dem. or Rep. presidents?

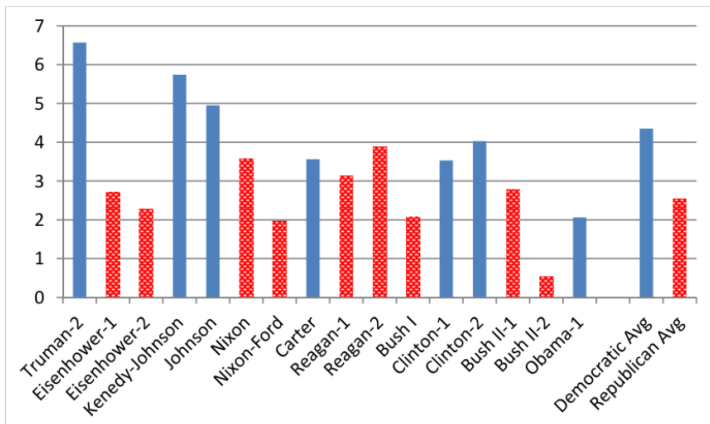


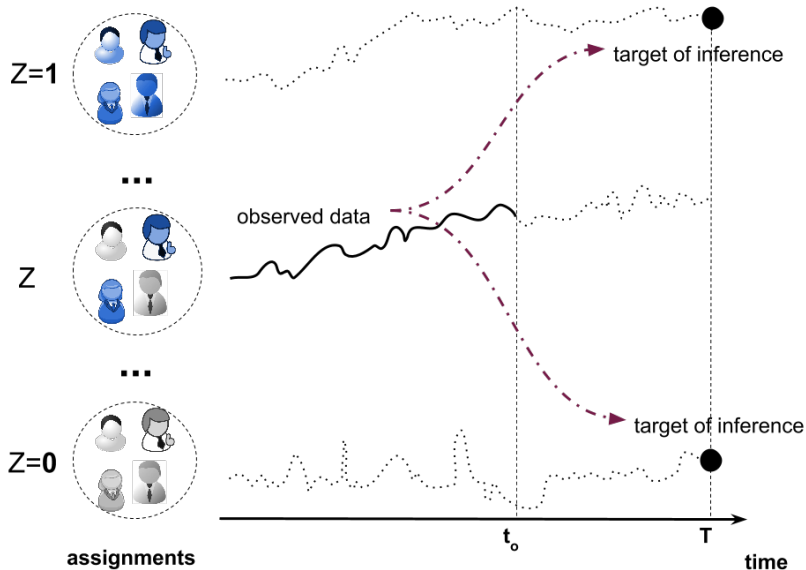
Figure: y-axis: %GDP growth; x-axis: incumbent president; color: party affiliation of incumbent president.

- $D - R = +1.8\%$ if policy effect lag=0 yrs; $D - R = +0.7\%$ if lag = 4 yrs; $D - R = -1.07\%$ if lag = 8 yrs.

Outline of problem

- We focus on simple multiagent systems (e.g., auctions).
 - Usually high-stakes! (e.g., multi-billion Yahoo! experiment, 2011)
- Two policies: policy 0 (baseline) and policy 1 (new).
- Agents **experimentally** assigned to policies.
- Goal is to compare policy 0 with policy 1 and decide which one is best using **short-term experimental data**.

Illustration of problem



Long-Term Average Causal Effect (LACE)

- Z = binary assignment vector; $Z_i = 1$ agent i was assigned to policy 1; $Z_i = 0$ means agent i was assigned to policy 0.
- A = *action set* same for each agent; observed data = agent actions at $t = 0, 1, \dots, t_0$ for every policy g .
- T = long-term horizon.
- $R_{g,t}(Z)$ = value of actions in policy g at period t under assignment Z (e.g., revenue as function of bids)

Long-Term Average Causal Effect (LACE)

- Z = binary assignment vector; $Z_i = 1$ agent i was assigned to policy 1; $Z_i = 0$ means agent i was assigned to policy 0.
- A = *action set* same for each agent; observed data = agent actions at $t = 0, 1, \dots, t_0$ for every policy g .
- T = long-term horizon.
- $R_{g,t}(Z)$ = value of actions in policy g at period t under assignment Z (e.g., revenue as function of bids)

Definition

The *long-term causal effect* is defined as follows:

$$\tau = R_{1,T}(Z = \mathbf{1}) - R_{0,T}(Z = \mathbf{0}).$$

Methological challenges

- Need to extrapolate from Z to **1** and **0**; and from $[0, t_0]$ to T .

Methological challenges

- Need to extrapolate from Z to $\mathbf{1}$ and $\mathbf{0}$; and from $[0, t_0]$ to T .
- We need stability assumptions on both dimensions (c.f., policy invariance, SUTVA)
- A critical analysis choice:

Work on the (observed) action space or on a latent space?

- We argue in favor of working on **latent behavioral space**.

Behavioral Model

- A **behavior** $b \in B$ given a policy maps to a strategy (distribution over actions):

$$\mathcal{G} \times B \in \Delta^{|A|}$$

- Example: random behavior, risk averse,...

Behavioral Model

- A **behavior** $b \in B$ given a policy maps to a strategy (distribution over actions):

$$\mathcal{G} \times B \in \Delta^{|A|}$$

- Example: random behavior, risk averse,...
- $\beta_{g,t}(Z) \in \Delta^{|B|} =$ **population behavior**— fraction of agents adopting each behavior—in policy g , at period t , under assignment Z .

Assumption #1: Stability

Assumption [stability of initial behaviors]

Let ρ_Z be the proportion of agents assigned to new policy under assignment Z . Then, for every Z ,

$$\rho_Z \beta_{1,0}(Z) + (1 - \rho_Z) \beta_{0,0}(Z) = \beta^{(0)}, \quad (1)$$

where $\beta^{(0)}$ is population behavior invariant to Z .

- Agents “have made up their minds” to adopt $\beta^{(0)}$ before the experiment (no anticipation).

Assumption #1: Stability

Assumption [stability of initial behaviors]

Let ρ_Z be the proportion of agents assigned to new policy under assignment Z . Then, for every Z ,

$$\rho_Z \beta_{1,0}(Z) + (1 - \rho_Z) \beta_{0,0}(Z) = \beta^{(0)}, \quad (1)$$

where $\beta^{(0)}$ is population behavior invariant to Z .

- Agents “have made up their minds” to adopt $\beta^{(0)}$ before the experiment (no anticipation).
- **Invariant quantities** wrt to Z are necessary to extrapolate across assignments.

Assumption #1: Stability

Assumption [stability of initial behaviors]

Let ρ_Z be the proportion of agents assigned to new policy under assignment Z . Then, for every Z ,

$$\rho_Z \beta_{1,0}(Z) + (1 - \rho_Z) \beta_{0,0}(Z) = \beta^{(0)}, \quad (1)$$

where $\beta^{(0)}$ is population behavior invariant to Z .

- Agents “have made up their minds” to adopt $\beta^{(0)}$ before the experiment (no anticipation).
- **Invariant quantities** wrt to Z are necessary to extrapolate across assignments.
 - e.g. SUTVA (Cox, Rubin): if $Y(Z)$ = outcome under Z then

$$Y(Z) = Z_i \cdot Y^1 + (1 - Z_i) \cdot Y^0.$$

Assumption #2: Behavioral ignorability of treatment assignment

Assumption [behavioral ignorability]

Let ϕ, ψ denote vector parameters, then

$$\begin{aligned}\beta_{g,0}(Z) &\sim \pi_{\phi}, \\ \beta_{g,t}(Z) &\sim f_{\psi}(H_{g,<t}), \quad \forall g, t\end{aligned}\tag{2}$$

where π, f are known models and $H_{g,<t}$ denotes behavior history up to t .

Assumption #2: Behavioral ignorability of treatment assignment

Assumption [behavioral ignorability]

Let ϕ, ψ denote vector parameters, then

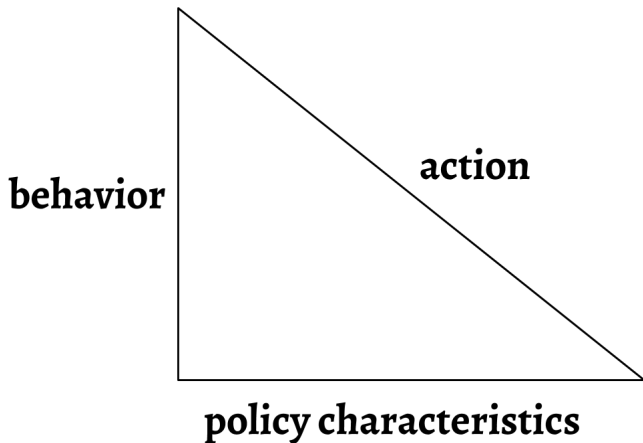
$$\begin{aligned}\beta_{g,0}(Z) &\sim \pi_{\phi}, \\ \beta_{g,t}(Z) &\sim f_{\psi}(H_{g,<t}), \quad \forall g, t\end{aligned}\tag{2}$$

where π, f are known models and $H_{g,<t}$ denotes behavior history up to t .

- ϕ, ψ may depend on Z **only through** ρ_Z ; also may depend on policy g .
- Adaptation of β in policy g is independent of Z conditional on history (Markovian assumption).

More on space of assumptions

- Stability assumptions on behavioral space are **plausible** because behavior does not depend on policy. Not plausible on actions.

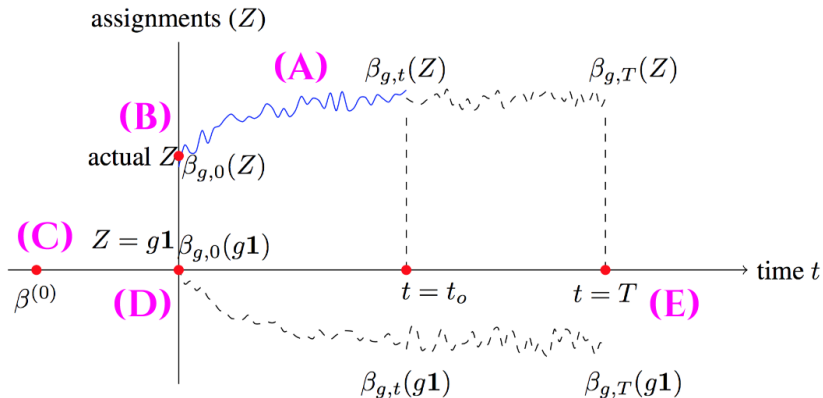


Main Result

Theorem [estimation of long-term effect]

Suppose that assumptions of no-anticipation and behavioral ignorability hold. Then, the long-term average causal effect (LACE) is identifiable and can be consistently estimated.

Illustration of estimation method



- Assumption 1 (Stability) is crucial in $(B) \rightarrow (C) \rightarrow (D)$.
- Assumption 2 (Ignorability) is crucial in $(A) \rightarrow (B)$ and $(D) \rightarrow (E)$.

In practice: QL_k and VAR(1)

For the behavioral model we adopt QL_3 model (Stahl and Wilson, 1984) with parameters $(\lambda_1, \lambda_{1(2)}, \lambda_2)$:

In practice: QL_k and VAR(1)

For the behavioral model we adopt QL_3 model (Stahl and Wilson, 1984) with parameters $(\lambda_1, \lambda_{1(2)}, \lambda_2)$:

- **Level-0** agent cannot compute expected utilities, and

plays actions w.p. $\propto 1$;

- **level-1** agent computes expected utilities u_1 assuming play against Level-0, and

plays actions w.p. $\propto e^{\lambda_1 u_1}$;

- **level-2** agent computes expected utilities u_2 assuming play against Level-1 agent with precision $\lambda_{1(2)}$, and

plays actions w.p. $\propto e^{\lambda_2 u_2}$.

In practice: QL_k and VAR(1)

We choose a lag-one autoregressive model, $VAR(1)$, for the evolution of population behavior:

$$w_{g,t} = \psi_0 + \psi_1 \cdot w_{g,t-1} + \psi_2 \cdot \epsilon_{g,t},$$

where

- $\epsilon_{g,t} \sim N(0, \sigma^2 I)$ iid;
- temporal parameters (ψ_0, ψ_1, ψ_2) may depend on policy g ;
- and w is the logit transform of population behavior β .

[logit(x) = ($\log(x_2/x_1), \log(x_3/x_1), \dots$)]

Application: Rapoport and Boebel (1992)

	a'_1	a'_2	a'_3	a'_4	a'_5
a_1	W	L	L	L	L
a_2	L	L	W	W	W
a_3	L	W	L	L	W
a_4	L	W	L	W	L
a_5	L	W	W	L	L

- RB randomly people to play row or column; 20 players in each game; four sessions, each for multiple rounds;

We **re-appropriate** the data for our needs:

What is the effect switching from $(W, L) = (\$10, \$6)$ to $(W, L) = (\$15, \$1)$?

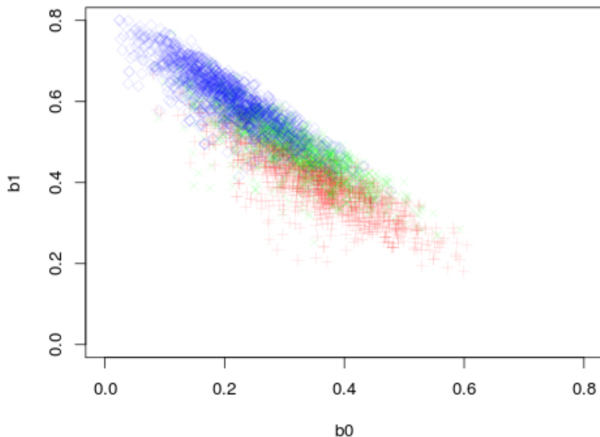
Data

Policy	Period	row agent				column agent			
		a_1	a_2	a_3	a_4	a'_1	a'_2	a'_3	a'_4
0	0	0.308	0.307	0.113	0.120	0.350	0.218	0.202	0.092
0	1	0.293	0.272	0.162	0.100	0.333	0.177	0.190	0.140
0	2	0.273	0.350	0.103	0.123	0.353	0.133	0.258	0.102
0	3	0.295	0.292	0.113	0.135	0.372	0.192	0.222	0.063
1	0	0.258	0.367	0.105	0.143	0.332	0.115	0.245	0.140
1	1	0.290	0.347	0.118	0.110	0.355	0.198	0.208	0.108
1	2	0.355	0.313	0.082	0.100	0.355	0.215	0.187	0.110
1	3	0.323	0.270	0.093	0.105	0.343	0.243	0.168	0.107

- We define hold-out set at $T = 3$;
- Revenue is linear combination of actions.

Adaptation of behavior

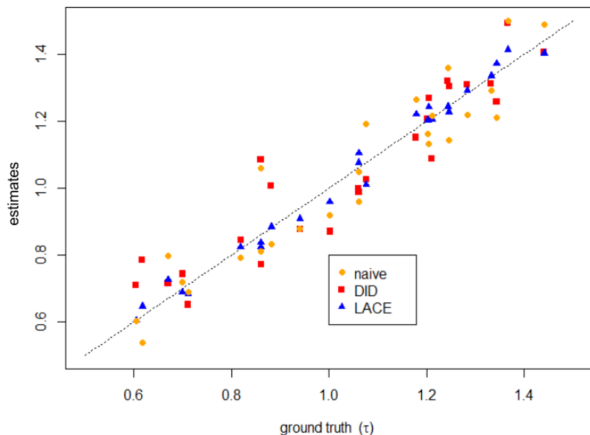
We estimate significant $\psi_1 \neq 0$, indicating temporal trend and learning—recall $w_{g,t} = \psi_0 + \psi_1 \cdot w_{g,t-1} + \psi_2 \cdot \epsilon_{g,t}$.



red: period 0, green: period 1, blue: period 2

Results: estimates of long-term causal effect

DID: mse = 0.361; Naive: mse=0.183; LACE: mse = 0.045.



- Naive: estimate causal effect as $R_{1,3} - R_{0,3}$.
- DID: estimate causal effect as $(R_{0,3} - R_{0,1}) - (R_{1,3} - R_{1,1})$.

Conclusion

- Leverage **behavioral game theory** for causal estimation of long-term effects.
- Effects are estimable under **stability assumption** on initial behaviors; **ignorability assumption** on behavior adaptation.

Conclusion

- Leverage **behavioral game theory** for causal estimation of long-term effects.
- Effects are estimable under **stability assumption** on initial behaviors; **ignorability assumption** on behavior adaptation.

Extensions:

- How to **define** long-term T ? Depends crucially on choice of temporal assumption and model.
- Combine with **richer temporal behavioral models** (e.g., agent-level learning).
- Combine with **payoff uncertainty**.
- **Apply** on large-scale real-world experiment.

Thank you!

