Counterfactuals with Latent Information

Ryosei Okamoto, Hideyuki Sasaki

November 6th, 2023

Introduction

Counterfactual Prediction

- Main Topic
  - A method to characterize counterfactual predictions in incomplete information games
- Counterfactual Predictions
  - The analyst observes behavior assumed to be rationalized by a Bayesian model
  - What would have been true under different circumstances?

Comparison to Previous Studies

- Previous works
  - Most applied work relies on strong assumptions and undermines the credibility of the models
- Novelty of this work
  - Non-parametric approach to treat latent information structures
  - Concise description of counterfactual predictions

Organization of thesis

- The authors
  - proved 2 theorems that characterize counterfactual predictions
  - showed examples of counterfactual analysis using the theorems

Preliminaries

Base game

- $\theta \in \Theta$: finite state of the world
- $i = 1, .., N$: players
- A_i : finite set of actions
- $u_i : A \times \Theta \longrightarrow \mathbb{R}$: utility
- $\mathcal{G} = (A_i, u_i)_{i=1}^{N}$: base game

Bayesian game

- $\mu \in \Delta(\Theta)$: prior distribution over states
- S_i : measurable set of signals
- $\pi : \Theta \longrightarrow \Delta(S)$: distribution of signals
- $\mathcal{I} = \left((S_i)_{i=1}^{N}, \pi\right)$: information structure
- $(\mu, \mathcal{G}, \mathcal{I})$: Bayesian game

Nash equilibrium

- $\sigma_i : S_i \longrightarrow \Delta(A_i)$: strategy
  - $\sigma_i(a_i|s_i)$: probability of a_i given s_i
- $U_i(\sigma) = \sum_{\theta \in \Theta} \int_{s \in S} \sum_{a \in A} u_i(a, \theta)\sigma(a \mid s)\pi(ds \mid \theta)\mu(\theta)$
  : expected utility under σ

Def. 1: Nash equilibrium

σ is a Nash equilibrium if $U_i(\sigma) \geq U_i\left(\sigma_i', \sigma_{-i}\right)$ for all i and for all strategies σ_i'

Bayes Correlated equilibrium

- $\phi \in \Delta(A \times \Theta)$: outcome of $\mathcal{G}$
- ϕ is induced by $(\mu, \mathcal{I}, \sigma)$

$$\phi(a,\theta) = \int_{s \in S} \sigma(a \mid s) \pi(ds \mid \theta) \mu(\theta)$$

Def. 2: Bayes correlated equilibrium(BCE)

ϕ is BCE if

$$\sum_{\theta \in \Theta} \sum_{a_{-i} \in A_{-i}} \left(u_i\left(a_i, a_{-i}, \theta\right) - u_i\left(a_i', a_{-i}, \theta\right)\right) \phi\left(a_i, a_{-i}, \theta\right) \geq 0$$

for all i, a_i, a_i'

Joint Predictions and Counterfactuals

Joint Predictions with Fixed Information

- $\mathcal{G}^k = \left(A_i^k, u_i^k\right)_{i=1}^K$
- players simultaneously play $\mathcal{G}^1, \ldots, \mathcal{G}^K$
- Information structure is the same in each game

Def. 3: Joint Prediction

an outcome profile

$$\left(\phi^1, \ldots, \phi^K\right) \in \Delta\left(A^1 \times \Theta\right) \times \cdots \times \Delta\left(A^K \times \Theta\right)$$

is a joint prediction if there exists a prior μ, an information structure $\mathcal{I}$, and for each $k = 1, \ldots, K$ an equilibrium σ^k of $(\mu, \mathcal{G}^k, \mathcal{I})$ such that ϕ^k is induced by $(\mu, \mathcal{I}, \sigma^k)$.

Linked Game

Def. 4: Linked Game

The linked game $\overline{\mathcal{G}} = \left(\overline{A}_i, \overline{u}_i\right)_{i=1}^{N}$ is defined by, for each $i, \overline{A}_i = A_i^1 \times \cdots \times A_i^K$ and

$$\overline{u}_i(\overline{a}, \theta) = \sum_{k=1,\ldots,K} u_i^k \left(a^k, \theta\right)$$

where $\overline{a}_i = \left(a_i^1, \ldots, a_i^K\right)$

Linked Game

- $\mathcal{G}^k$: a component game of $\overline{\mathcal{G}}$
- An outcome $\overline{\phi}$ of $\overline{\mathcal{G}}$ can be identified with a joint distribution in $\Delta\left(A^1 \times \ldots \times A^K \times \Theta\right)$

Joint Predictions

Theorem 1

A tuple $(\phi^1, \dots, \phi^K)$ is a joint prediction for $\mathcal{G}^1, \dots, \mathcal{G}^K$ if and only if there exists a BCE $\overline{\phi}$ of $\overline{\mathcal{G}}$ for which the marginal of $\overline{\phi}$ on $A^k \times \Theta$ is ϕ^k for each $k = 1, \dots, K$.

Lemma 1

ϕ is a BCE if and only if there exists a prior μ, an information structure $\mathcal{I}$, and an equilibrium σ of $(\mu, \mathcal{G}, \mathcal{I})$ such that ϕ is induced by $(\mu, \mathcal{I}, \sigma)$ (Prop.1 in BCE)

Joint Predictions

- Fix a prior μ, an information structure $\mathcal{I}$ and strategy profile $\overline{\sigma}$ in $(\mu, \mathcal{I}, \overline{\mathcal{G}})$.
- For each k,σ_i^k is the strategy in $(\mu, \mathcal{G}^k, \mathcal{I})$ where $\sigma_i^k(\cdot \mid s_i)$ is the marginal of $\overline{\sigma}(\cdot \mid s_i)$ on A_i^k.
- Thus, the marginal of $\overline{\phi}$ on $A^k \times \Theta$ is ϕ^k

Lemma 2

$\overline{\sigma}$ is an equilibrium of $(\mu, \overline{\mathcal{G}}, \mathcal{I})$ if and only if σ^k is an equilibrium of $(\mu, \mathcal{G}^k, \mathcal{I})$ for each k.

Proof of Theorem 1

$\left(\phi^1, \ldots, \phi^K\right)$ is a joint prediction for $(\mathcal{G}^1, \ldots, \mathcal{G}^K)$

$\Leftrightarrow {}^{\exists}\mu, {}^{\exists}I, {}^{\exists}\sigma^k$ of $(\mu, \mathcal{G}^k, I)$ for each k s.t. ϕ^k induced by $(\mu, \mathcal{I}, \sigma^k)$

$\Leftrightarrow {}^{\exists}\mu, {}^{\exists}I, {}^{\exists}\overline{\sigma}$ of $(\mu, \overline{\mathcal{G}}, I)$ for each k

s.t. $\overline{\phi}$ is induced by $(\mu, \mathcal{I}, \overline{\sigma})$ and the marginal of $\overline{\phi}$ on $A^k \times \Theta$ is ϕ^k

$\Leftrightarrow \overline{\phi}$ is BCE of $\overline{\mathcal{G}}$ s.t. the marginal of $\overline{\phi}$ on $A^k \times \Theta$ is ϕ^k

- By Def.3, Lem.2,1

Proof of Lemma 2

$$\begin{aligned}
\overline{U}_i(\overline{\sigma}) &= \sum_{\theta\in\Theta}\int_{s\in S}\sum_{\overline{a}\in\overline{A}}\overline{u}_i(\overline{a},\theta)\overline{\sigma}(\overline{a}\mid s)\pi(ds\mid\theta)\mu(\theta)\\
&= \sum_{\theta\in\Theta}\int_{s\in S}\sum_{k=1,\dots,K}\left[\sum_{a\in A^k}u_i^k(a,\theta)\sigma^k(a\mid s)\right]\pi(ds\mid\theta)\mu(\theta)\\
&= \sum_{k=1,\dots,K}\sum_{\theta\in\Theta}\int_{s\in S}\left[\sum_{a\in A^k}u_i^k(a,\theta)\sigma^k(a\mid s)\right]\pi(ds\mid\theta)\mu(\theta)\\
&= \sum_{k=1,\dots,K}U_i^k\left(\sigma^k\right).
\end{aligned}$$

Proof of Lemma 2

- If $\overline{\sigma}$ is not an equilibrium, then there exist i and a strategy $\overline{\tau}_i$ such that

$$\sum_{k=1,\ldots,K} U_i^k\left(\sigma^k\right) = \overline{U}_i(\overline{\sigma}) < \overline{U}_i\left(\overline{\tau}_i, \overline{\sigma}_{-i}\right) = \sum_{k=1,\ldots,K} U_i^k\left(\tau_i^k, \sigma_{-i}\right)$$

  where τ_i^k is the marginal of $\overline{\tau}_i$ on A_i^k.
- Thus, for at least one k, τ_i^k is a profitable deviation in $\left(\mu, \mathcal{G}^k, \mathcal{I}\right)$.

Proof of Lemma 2

- Assume there is a profitable deviation in one of the component games, say to τ_i^k for player i in $\mathcal{G}^k$
- Then, the strategy defined by, for all $\overline{a}_i \in \overline{A}_i$,

$$\overline{\tau}_i \left(\overline{a}_i \mid s_i\right) = \tau_i \left(\overline{a}_i^k \mid s_i\right) \overline{\sigma}_i \left(\overline{a}_i^{-k} \mid s_i\right)$$

is a profitable deviation in the linked game.

Counterfactuals when Information is Latent and Fixed

- $\mathcal{G}$: observed game
- Analyst knows $\Theta, A, (u_i)_{i=1}^N$
- Analyst does not know $\mathcal{I}$
- Analyst knows ϕ of $\mathcal{G}$
  - lies in a set $M \subseteq \Delta(A \times \Theta)$
  - was generated under some prior μ and information structure $\mathcal{I}$
  - was induced by an equilibrium of $(\mu, \mathcal{G}, \mathcal{I})$
- The analyst wants to make counterfactual predictions for what might happen if the unobserved game $\widehat{\mathcal{G}}$ were played
- Analyst assumes that μ and $\mathcal{I}$ are the same in $\widehat{\mathcal{G}}$ as in $\mathcal{G}$

Counterfactual Predictions

Def. 5: Counterfactual Prediction

An outcome $\widehat{\phi} \in \Delta(\widehat{A} \times \Theta)$ is a counterfactual prediction if there exist $\mu, \mathcal{I}$, and equilibria σ and $\widehat{\sigma}$ of $(\mu, \mathcal{G}, \mathcal{I})$ and $(\mu, \widehat{\mathcal{G}}, \mathcal{I})$, respectively, such that the outcome ϕ induced by σ is in M and such that $\widehat{\phi}$ is induced by $\widehat{\sigma}$.

- $\widehat{\Phi}$: Set of counterfactual predictions $\widehat{\phi}$

Counterfactual Predictions

Theorem 2

An outcome $\widehat{\phi} \in \Delta(\widehat{A} \times \Theta)$ is in $\widehat{\Phi}$ if and only if there is a BCE $\overline{\phi}$ of $\overline{\mathcal{G}}$ such that (i) the marginal of $\overline{\phi}$ on $A \times \Theta$ is in M and (ii) $\widehat{\phi}$ is the marginal of $\overline{\phi}$ on $\widehat{A} \times \Theta$.

- M is obtained from data
  - $|M| = 1$, if the analyst observed ϕ
  - M contains all the outcomes whose marginal distribution of actions coincides with the data if the distribution of actions is observed
- (i) and (ii) are described as an intersection of a finite number of linear inequalities

Proof of Theorem 2

$$\widehat{\phi} \in \widehat{\Phi} \Leftrightarrow {}^{\exists}\mu, {}^{\exists}I, {}^{\exists}\sigma \text{ of } (\mu, G, I), {}^{\exists}\hat{\sigma} \text{ of } (\mu, G, I)$$

s.t. ϕ induced by σ is in M and $\hat{\phi}$ induced by $\hat{\sigma}$

$\Leftrightarrow (\phi, \widehat{\phi})$ is a joint prediction for $\mathcal{G}, \widehat{\mathcal{G}}$ s.t. $\phi \in M$

$\Leftrightarrow {}^{\exists}$BCE $\widehat{\phi}$ of $\widehat{\mathcal{G}}$

s.t. $\overline{\phi}$'s marginals on $A \times \Theta$ and $\widehat{A} \times \Theta$ are $\phi \in M$ and $\widehat{\phi}$

by Def 5,3 and Thm 1

One-Player Games

Model

- Consider the decision-making by a single agent.
- Observed game is given as follows:
  - Action: $A = \{0, 1\}$
  - State: $\Theta = \{-1, 1\}$ w.p. $1/2$
  - Payoff function: $u(a, \theta) = a\theta$
- Counterfactual game is given as follows:
  - Action: $\hat{A} = \{0, 1\}$
  - State: $\Theta = \{-1, 1\}$ w.p. $1/2$
  - Payoff function: $\hat{u}(\hat{a}, \theta) = \hat{a}(\theta + z)$ where $z \in \mathbb{R}$

Model

- Payoff matrix:

$a \backslash \theta$	-1	1
0	0	0
1	$-1+z$	$1+z$

- Observed distribution on (a, θ) is $M = \{\phi\}$ where $\phi : A \times \Theta \to [0,1]$ satisfies $\alpha \in [1/4, 1/2]$ and

$a \backslash \theta$	-1	1
0	α	$1/2 - \alpha$
1	$1/2 - \alpha$	α

Model

- Let $\overline{\phi} \in \Delta\left(A \times \hat{A} \times \Theta\right)$ be an outcome in the linked game.
- We want to investigate the maximal and minimal counterfactual welfare; for the maximal welfare, solve

$$\max_{\overline{\phi} \geq 0} \sum_{(a,\hat{a},\theta)} \overline{\phi}(a,\hat{a},\theta)\hat{a}(\theta + z),$$

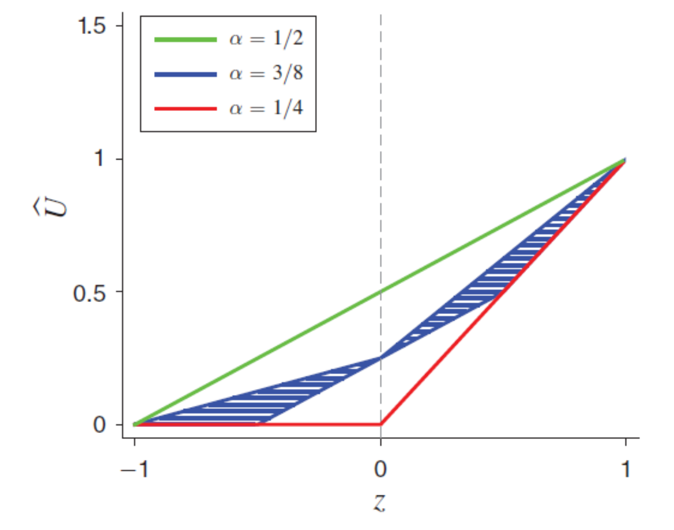
subject to

Model

$$\sum_{\hat{a}} \overline{\phi}(a, \hat{a}, \theta) = \begin{cases} \alpha, & \text{if } (a, \theta) \in \{(0, -1), (1, 1)\}, \\ 1/2 - \alpha, & \text{otherwise}, \end{cases}$$

and the obedience constraints for the linked game.

Counterfactual Welfare



Counterfactual Welfare

- Let $z \in (-1, 1)$ for simplicity.
- Consider the case $\alpha = 1/2$.
  - Observed distribution:

$a \backslash \theta$	-1	1
0	$1/2$	0
1	0	$1/2$

- Information structure should be **"full information"**.
- $\hat{a} = 0$ is taken if $a = 0$, and $\hat{a} = 1$ is taken if $a = 1$. So

$$\text{welfare} = 1/2 \cdot 0 + 1/2 \cdot 1 \cdot (1 + z) = (1 + z)/2.$$

Counterfactual Welfare

- Consider the case $\alpha = 1/4$.
  - Observed distribution:

$a\backslash\theta$	-1	1
0	$1/4$	$1/4$
1	$1/4$	$1/4$

- Information structure should be **"no information"**.
- $\hat{a} = 0$ is taken if $-1 < z \leq 0$, and $\hat{a} = 1$ is taken if $0 \leq z < 1$. So the welfare is 0 if $-1 < z \leq 0$, and

$$\text{welfare} = 1/2 \cdot 1 \cdot (-1 + z) + 1/2 \cdot 1 \cdot (1 + z) = z$$

if $0 \leq z < 1$.

Counterfactual Welfare

- Consider the case $\alpha = 3/8$.
  - Observed distribution:

$a \backslash \theta$	-1	1
0	$3/8$	$1/8$
1	$1/8$	$3/8$

- There may be multiple candidates for information structure, so the maximal and minimal counterfactual welfare may differ.
- We first derive the maximal welfare.

Counterfactual Welfare

- By the payoff function u_i, $a = 0$ is chosen only if the agent's posterior belief satisfies $\Pr(\theta = 1|s, a = 0) \in [0, 1/2]$ for all $s \in S$.
- Also, since $\Pr(\theta = 1|a = 0) = 1/4$, a family of posterior beliefs $\{\Pr(\theta = 1|s, a = 0)\}_{s \in S}$ satisfies $\mathrm{E}[\Pr(\theta = 1|s, a = 0)] = 1/4$.
- The most (Blackwell) informative signal should induce the posterior beliefs $\Pr(\theta = 1|s, a = 0) = 0$ and $\Pr(\theta = 1|s, a = 0) = 1/2$.
- Actually, a more informative signal leads to higher expected utility in a single agent's decision problem.

Counterfactual Welfare

- For $a = 1$, by the similar argument, $\Pr(\theta = 1|s, a = 1) \in [1/2, 1]$ for all $s \in S$.
- Also, since $\Pr(\theta = 1|a = 1) = 3/4$, a family of posterior beliefs $\{\Pr(\theta = 1|s, a = 0)\}_{s \in S}$ satisfies $\mathrm{E}[\Pr(\theta = 1|s, a = 0)] = 3/4$.
- The most informative signal splits the posterior belief $\Pr(\theta = 1|s, a = 0)$ into $1/2$ and 1.
- Hereafter, we derive the information structure.
- Let (S, π) be the most informative signal and define $S = A \times \hat{A}$ and $\pi \in \Delta(A \times \hat{A} \times \Theta)$.

Counterfactual Welfare

- By the obedience condition, $\overline{\phi}(a, \hat{a}, \theta) = \pi(a, \hat{a}, \theta)$. So π should satisfy

$$\pi(a,1,1) + \pi(a,0,1) = \phi(a,1) \quad \forall a \in \{0,1\},$$
$$\pi(a,1,-1) + \pi(a,0,-1) = \phi(a,-1) \quad \forall a \in \{0,1\}.$$

- Also, by the argument above, π should satisfy the requirement for posterior belief:

$$\Pr(\theta = 1|a,1) = \frac{\pi(a,1,1)}{\pi(a,1,1) + \pi(a,1,-1)} = \begin{cases} 1/2 & \text{if } a = 0, \\ 1 & \text{if } a = 1, \end{cases}$$

$$\Pr(\theta = 1|a,0) = \frac{\pi(a,0,1)}{\pi(a,0,1) + \pi(a,0,-1)} = \begin{cases} 0 & \text{if } a = 0, \\ 1/2 & \text{if } a = 1. \end{cases}$$

Counterfactual Welfare

- By tedious calculation, we have

$\theta \backslash s$	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
-1	$1/4$	$1/8$	$1/8$	0
1	0	$1/8$	$1/8$	$1/4$

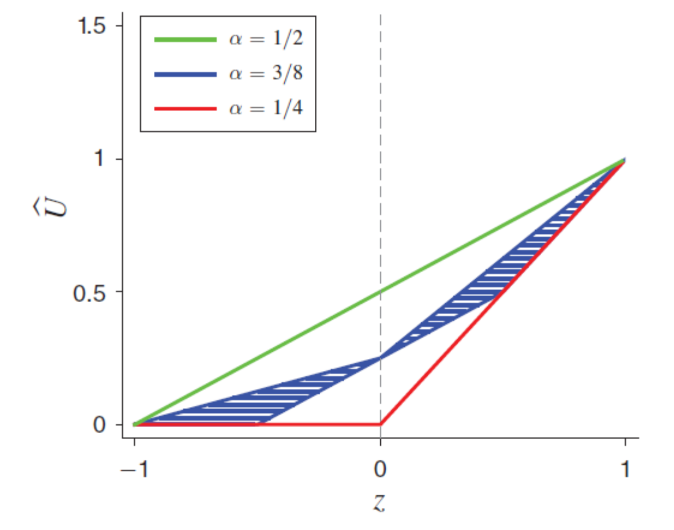
- Actually, $s = (0,1)$ and $s = (1,0)$ are mutually redundant, so we also have

$\theta \backslash s$	0	$1/2$	1
-1	$1/4$	$1/4$	0
1	0	$1/4$	$1/4$

Counterfactual Welfare

- In the second information structure,
  - $s = 0$ or $1 \Rightarrow$ full information (same as in the case $\alpha = 1/2$)
  - $s = 1/2 \Rightarrow$ no information (same as in the case $\alpha = 1/4$)
- So the welfare can be calculated by **taking the expectation of full information case and no information case**.

Counterfactual Welfare



Counterfactual Welfare

- Next, we derive the minimal welfare.
- Consider the following information structure:

$\theta \backslash s$	0	1
-1	$3/8$	$1/8$
1	$1/8$	$3/8$

  where $a = 1$ if $s = 1$ and $a = 0$ if $s = 0$.
- In the least informative case, the chosen a itself is a signal, so there is no more information except for the information obtained from a.

Counterfactual Welfare

- Then, derive the counterfactual Welfare.
- Given a, then $\hat{a} = a$ holds if

$$
\begin{aligned}
0 &\geq \sum_{\theta} \pi(-1, \theta)(\theta + z) = 3/8(-1+z) + 1/8(1+z) \iff z \leq 1/2, \\
0 &\leq \sum_{\theta} \pi(1, \theta)(\theta + z) = 1/8(-1+z) + 3/8(1+z) \iff z \geq -1/2.
\end{aligned}
$$

- So if $z \in [-1/2, 1/2]$, the welfare is

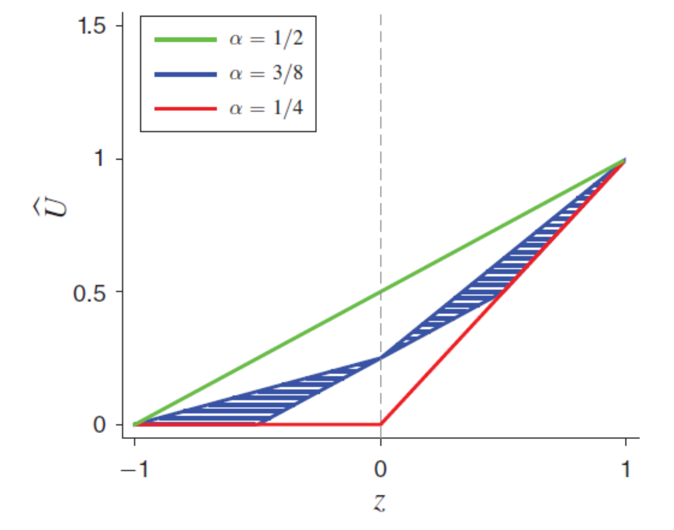
$$
1/8(-1+z) + 3/8(1+z) = 1/4 + z/2.
$$

Counterfactual Welfare

- If $z < -1/2$, $\hat{a} = 0$ is always optimal. Then the welfare is 0.
- If $z > 1/2$, $\hat{a} = 1$ is always optimal. Then the welfare is

$$1/2(-1 + z) + 1/2(1 + z) = z.$$

Counterfactual Welfare



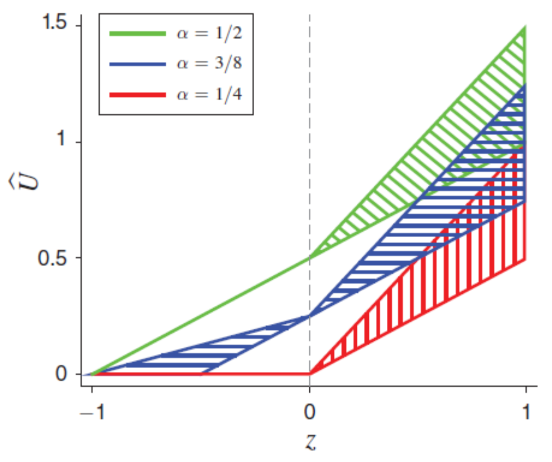
Counterfactual Welfare with Partially Observed Outcome

- In many cases, the data is censored.
- Here, the distribution of the state is observed when $a = 1$ but unobserved when $a = 0$:

$a \backslash \theta$	-1	1
0	$?$	$?$
1	$1/2 - \alpha$	α

- Then, the constraints on the outcome are relaxed.

Counterfactual Welfare with Partially Observed Outcome



Two-Player Games

Model

- Consider a simple entry game with probabilistic entry costs.
- The Game is given as follows:
  - Action: $A = \hat{A} = \{N, E\} \times \{N, E\}$
  - State: $\theta = (c_1, c_2), \Theta = \{(0,0), (0,2), (2,0), (2,2)\}$
  - Payoff matrix ($z = 0$ for the observed game):

$a_1 \backslash a_2$	N	E
N	$0, 0$	$0, 3 - c_2 + z$
E	$3 - c_1 + z, 0$	$1 - c_1 + z, 1 - c_2 + z$

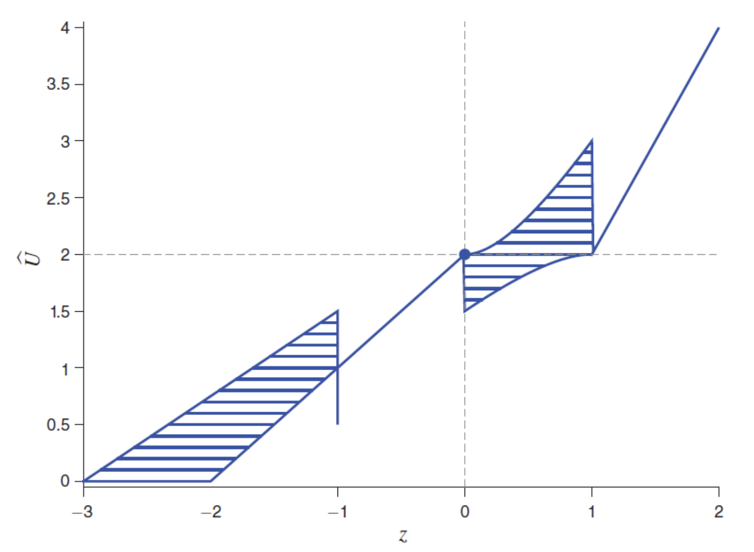
Model

- Observed outcome is

$$\phi(a_1, a_2, c_1, c_2) = 1/4$$

  if $(a_i, c_i) \in \{(E, 0), (N, 2)\}$ for all $i \in \{1, 2\}$, and 0 otherwise.
- We want to predict the counterfactual producer surplus for each z.

Counterfactual Surplus



Conclusion

- Suppose that we want to predict the outcome of counterfactual games using the data of an observed game in hand.
- **Under the assumption that there is a common information structure among the observed game and the counterfactual games, we can obtain the prediction of the counterfactual games by the statements of Theorem 1 and Theorem 2**.
- The outcomes are characterized using the BCE of the linked game.