

Causal Interaction in High Dimension

Kosuke Imai

Princeton University

McCourt School of Public Policy
Georgetown University

March 26, 2015

Joint work with Naoki Egami

Interaction Effects and Causal Heterogeneity

1 Moderation

- How do treatment effects vary across individuals?
- Who benefits from (or is harmed by) the treatment?
- Interaction between treatment and pre-treatment covariates

2 Causal interaction

- What aspects of a treatment are responsible for causal effects?
- What combination of treatments is most efficacious?
- Interaction between treatment variables

3 Individualized treatment regimes

- What combination of treatments is optimal for a given individual?

Two Interpretations of Causal Interaction

① Conditional effect interpretation:

- Does the effect of one treatment change as we vary the value of another treatment?
- Does the effect of being black change depending on whether an applicant is male or female?
- Useful for testing moderation among treatments

② Interactive effect interpretation:

- Does a combination of treatments induce an *additional effect* beyond the sum of separate effects attributable to each treatment?
- Does being a black female induce an additional effect beyond the effect of being black and that of being female?
- Useful for finding efficacious treatment combinations in high dimension

An Illustration in the 2×2 Case

- Two binary treatments: A and B
- Potential outcomes: $Y(a, b)$ where $a, b \in \{0, 1\}$
- **Conditional effect interpretation:**

$$\underbrace{[Y(1, 1) - Y(0, 1)]}_{\text{effect of } A \text{ when } B = 1} - \underbrace{[Y(1, 0) - Y(0, 0)]}_{\text{effect of } A \text{ when } B = 0}$$

- **Interactive effect interpretation:**

$$\underbrace{[Y(1, 1) - Y(0, 0)]}_{\text{effect of } A \text{ and } B} - \underbrace{[Y(1, 0) - Y(0, 0)]}_{\text{effect of } A \text{ when } B = 0} - \underbrace{[Y(0, 1) - Y(0, 0)]}_{\text{effect of } B \text{ when } A = 0}$$

- The same quantity but two different interpretations
- The interactive interpretation requires the specification of the **baseline condition**: $(A, B) = (0, 0)$ in this example

Causal Interaction in High Dimension

- In the 2×2 case, computing all four average potential outcomes gives a complete picture
- The dimensionality rapidly increases as the number of levels and treatments increase
- A motivating application: **Conjoint analysis** (Hainmueller *et al.* 2014)
 - survey experiments to measure immigration preferences
 - a representative sample of 1,396 American adults
 - gender², education⁷, origin¹⁰, experience⁴, plan⁴, language⁴, profession¹¹, application reason³, prior trips⁵
 - Over **1 million** treatment combinations
 - What combinations of profiles characterize (un)preferred immigrants?
- We focus on the interactive interpretation in high dimension

Difficulty of the Conventional Approach

- **Lack of invariance** to the baseline condition
- Inference depends on the choice of baseline condition
- 3×2 example:
 - Treatment $A \in \{a_0, a_1, a_2\}$ and Treatment $B \in \{b_0, b_1, b_2\}$
 - Regression model with the baseline condition (a_0, b_0) :

$$\mathbb{E}(Y | A, B) = 1 + a_1^* + a_2^* + b_2^* + a_1^*b_2^* + 2a_2^*b_2^* + 3a_2^*b_1^*$$

- Interaction effect for $(a_2, b_2) >$ Interaction effect for (a_1, b_2)
- Another equivalent model with the baseline condition (a_0, b_1) :

$$\mathbb{E}(Y | A, B) = 1 + a_1^* + 4a_2^* + b_2^* + a_1^*b_2^* - a_2^*b_2^* - 3a_2^*b_0^*$$

- Interaction effect for $(a_2, b_2) <$ Interaction effect for (a_1, b_2)
- Interaction effect for (a_2, b_1) is zero under the second model
- All interaction effects with at least one baseline value are zero

The Contributions of the Paper

- 1 Standard treatment interaction effects suffer from the lack of order and interval invariance to the choice of baseline condition
- 2 Propose the **marginal treatment interaction effect** that is invariant
- 3 Derive the identification condition and estimation strategy for this new quantity
- 4 Generalize these results to the K -way causal interaction
- 5 Illustrate the methods with the immigration survey experiment

Two-way Causal Interaction

- Two factorial treatments:

$$A \in \mathcal{A} = \{a_0, a_1, \dots, a_{D_A-1}\}$$

$$B \in \mathcal{B} = \{b_0, b_1, \dots, b_{D_B-1}\}$$

- Assumption: **Full factorial design**

- ① Randomization of treatment assignment

$$\{Y(a_\ell, b_m)\}_{a_\ell \in \mathcal{A}, b_m \in \mathcal{B}} \perp\!\!\!\perp \{A, B\}$$

- ② Non-zero probability for all treatment combination

$$\Pr(A = a_\ell, B = b_m) > 0 \quad \text{for all } a_\ell \in \mathcal{A} \quad \text{and} \quad b_m \in \mathcal{B}$$

- Fractional factorial design not allowed

- ① Use a small non-zero assignment probability
- ② Focus on a subsample
- ③ Combine treatments

Non-Interaction Effects of Interest

① Average Treatment Combination Effect (ATCE):

- Average effect of treatment combination $(A, B) = (a_\ell, b_m)$ relative to the baseline condition $(A, B) = (a_0, b_0)$

$$\tau(a_\ell, b_m; a_0, b_0) \equiv \mathbb{E}\{Y(a_\ell, b_m) - Y(a_0, b_0)\}$$

- Which treatment combination is most efficacious?

② Average Marginal Treatment Effect (AMTE; Hainmueller et al. 2014):

- Average effect of treatment $A = a_\ell$ relative to the baseline condition $A = a_0$ averaging over the other treatment B

$$\psi(a_\ell, a_0) \equiv \int_{\mathcal{B}} \mathbb{E}\{Y(a_\ell, B) - Y(a_0, B)\} dF(B)$$

- Which treatment is effective on average?

The Conventional Approach to Causal Interaction

- **Average Treatment Interaction Effect (ATIE):**

$$\xi(a_\ell, b_m; a_0, b_0) \equiv \mathbb{E}\{Y(a_\ell, b_m) - Y(a_0, b_m) - Y(a_\ell, b_0) + Y(a_0, b_0)\}$$

- Conditional effect interpretation:

$$\underbrace{\mathbb{E}\{Y(a_\ell, b_m) - Y(a_0, b_m)\}}_{\text{Effect of } A = a_\ell \text{ when } B = b_m} - \underbrace{\mathbb{E}\{Y(a_\ell, b_0) - Y(a_0, b_0)\}}_{\text{Effect of } A = a_\ell \text{ when } B = b_0}$$

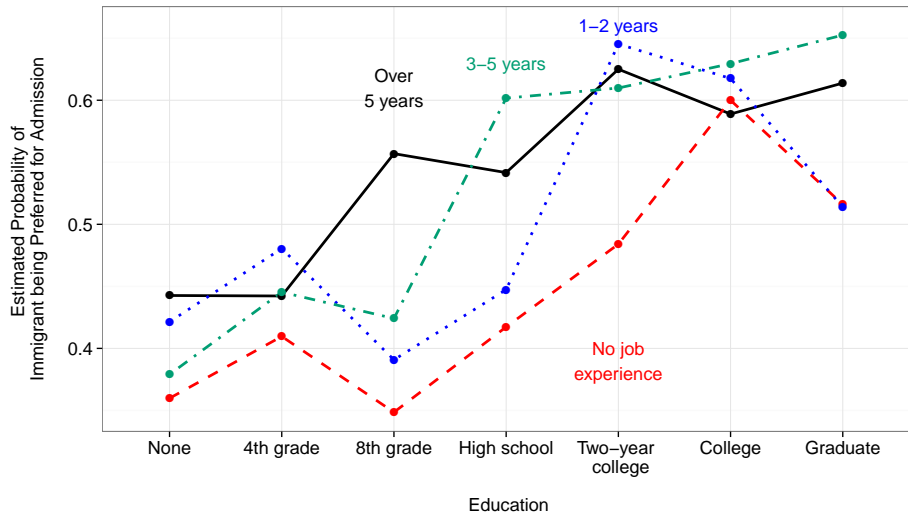
- Interactive effect interpretation:

$$\underbrace{\tau(a_\ell, b_m; a_0, b_0)}_{\text{ATCE}} - \underbrace{\mathbb{E}\{Y(a_\ell, b_0) - Y(a_0, b_0)\}}_{\text{Effect of } A = a_\ell \text{ when } B = b_0} - \underbrace{\mathbb{E}\{Y(a_0, b_m) - Y(a_0, b_0)\}}_{\text{Effect of } B = b_m \text{ when } A = a_0}$$

- Estimation: Linear regression with interaction terms

Ineffectiveness of Interaction Plot in High Dimension

Problem: it does not plot interaction effects themselves



Estimated Average Treatment Interaction Effect (ATIE)

Job experience	Education						
	None	4th grade	8th grade	High school	Two-year college	College	Graduate
None	0 (baseline)	0	0	0	0	0	0
1–2 years	0	0.009 (0.063)	−0.019 (0.063)	−0.032 (0.063)	0.100 (0.064)	−0.044 (0.064)	−0.064 (0.063)
3–5 years	0	0.016 (0.063)	0.056 (0.064)	0.165 (0.064)	0.107 (0.064)	0.010 (0.065)	0.117 (0.063)
> 5 years	0	−0.050 (0.064)	0.126 (0.064)	0.042 (0.063)	0.058 (0.064)	−0.094 (0.064)	0.015 (0.064)

The Effects of Changing the Baseline Condition

Job experience	Education						
	None	4th grade	8th grade	High school	Two-year college	College	Graduate
None	0.015 (0.064)	0.065 (0.062)	-0.111 (0.064)	-0.027 (0.061)	-0.043 (0.063)	0.109 (0.063)	0
1-2 years	0.078 (0.064)	0.138 (0.062)	-0.066 (0.062)	0.006 (0.061)	0.120 (0.062)	0.129 (0.062)	0
3-5 years	-0.102 (0.062)	-0.036 (0.062)	-0.172 (0.063)	0.021 (0.062)	-0.054 (0.061)	0.002 (0.062)	0
> 5 years	0	0	0	0	0	0	0 (baseline)

Lack of Invariance to the Baseline Condition

- Comparison between two ATIEs should not be affected by the choice of baseline conditions
- We prove that the ATIEs are neither interval or order invariant

- **Interval invariance:**

$$\begin{aligned} & \xi(a_\ell, b_m; a_0, b_0) - \xi(a_{\ell'}, b_{m'}; a_0, b_0) \\ &= \xi(a_\ell, b_m; a_{\tilde{\ell}}, b_{\tilde{m}}) - \xi(a_{\ell'}, b_{m'}; a_{\tilde{\ell}}, b_{\tilde{m}}), \end{aligned}$$

- **Order invariance:**

$$\begin{aligned} & \xi(a_\ell, b_m; a_0, b_0) \geq \xi(a_{\ell'}, b_{m'}; a_0, b_0) \\ \iff & \xi(a_\ell, b_m; a_{\tilde{\ell}}, b_{\tilde{m}}) \geq \xi(a_{\ell'}, b_{m'}; a_{\tilde{\ell}}, b_{\tilde{m}}). \end{aligned}$$

The New Causal Interaction Effect

- Average Marginal Treatment Interaction Effect (AMTIE):

$$\begin{aligned} & \pi(a_\ell, b_m; a_0, b_0) \\ \equiv & \underbrace{\tau(a_\ell, b_m; a_0, b_0)}_{\text{ATCE of } (A, B) = (a_\ell, b_m)} - \underbrace{\psi(a_\ell, a_0)}_{\text{AMTE of } A = a_\ell} - \underbrace{\psi(b_m, b_0)}_{\text{AMTE of } B = b_m} \end{aligned}$$

- Interactive effect interpretation: additional effect induced by $A = a_\ell$ and $B = b_m$ together beyond the separate effect of $A = a_\ell$ and that of $B = b_m$
- We prove that the AMTIEs are both interval and order invariant
- The AMTIEs do depend on the distribution of treatment assignment
 - ① specified by one's experimental design
 - ② motivated by the target population

The Relationships between the ATIE and the AMTIE

- 1 The AMTIE is a linear function of the ATIEs:

$$\begin{aligned}\pi(a_\ell, b_m; a_0, b_0) &= \xi(a_\ell, b_m; a_0, b_0) - \sum_{a \in \mathcal{A}} \Pr(A_i = a) \xi(a, b_m; a_0, b_0) \\ &\quad - \sum_{b \in \mathcal{B}} \Pr(B_i = b) \xi(a_\ell, b; a_0, b_0)\end{aligned}$$

- 2 The ATIE is also a linear function of the AMTIEs:

$$\xi(a_\ell, b_m; a_0, b_0) = \pi(a_\ell, b_m; a_0, b_0) - \pi(a_\ell, b_0; a_0, b_0) - \pi(a_0, b_m; a_0, b_0)$$

- Absence of causal interaction:
All of the AMTIEs are zero if and only if all of the ATIEs are zero
- The AMTIEs can be estimated by first estimating the ATIEs

Higher-order Causal Interaction

- J factorial treatments: $\mathbf{T} = (T_1, \dots, T_J)$

- Assumptions:

- ① Full factorial design

$$Y(\mathbf{t}) \perp\!\!\!\perp \mathbf{T} \quad \text{and} \quad \Pr(\mathbf{T} = \mathbf{t}) > 0 \quad \text{for all } \mathbf{t}$$

- ② Independent treatment assignment

$$T_j \perp\!\!\!\perp \mathbf{T}_{-j} \quad \text{for all } j$$

- Assumption 2 is not necessary for identification but considerably simplifies estimation
- We are interested in the K -way interaction where $K \leq J$
- We extend all the results for the 2-way interaction to this general case

Difficulty of Interpreting the Higher-order ATIE

- Generalize the 2-way ATIE by marginalizing the other treatments $\underline{\mathbf{T}}^{1:2}$

$$\xi_{1:2}(t_1, t_2; t_{01}, t_{02}) \equiv \int \mathbb{E} \{ Y(t_1, t_2, \underline{\mathbf{T}}^{1:2}) - Y(t_{01}, t_2, \underline{\mathbf{T}}^{1:2}) \\ - Y(t_1, t_{02}, \underline{\mathbf{T}}^{1:2}) + Y(t_{01}, t_{02}, \underline{\mathbf{T}}^{1:2}) \} dF(\underline{\mathbf{T}}^{1:2})$$

- In the literature, the 3-way ATIE is defined as

$$\xi_{1:3}(t_1, t_2, t_3; t_{01}, t_{02}, t_{03}) \\ \equiv \underbrace{\xi_{1:2}(t_1, t_2; t_{01}, t_{02} \mid T_3 = t_3)}_{\text{2-way ATIE when } T_3 = t_3} - \underbrace{\xi_{1:2}(t_1, t_2; t_{01}, t_{02} \mid T_3 = t_{03})}_{\text{2-way ATIE when } T_3 = t_{03}}$$

- Higher-order ATIEs are similarly defined sequentially
- This representation is based on the **conditional effect interpretation**
- Problem: the conditional effect of conditional effects!

Interactive Interpretation of the Higher-order ATIE

- We show that the higher-order ATIE also has an **interactive effect interpretation**
- Example: 3-way ATIE, $\xi_{1:3}(t_1, t_2, t_3; t_{01}, t_{02}, t_{03})$, equals

$$\underbrace{\tau_{1:3}(t_1, t_2, t_3; t_{01}, t_{02}, t_{03})}_{\text{ATCE}}$$

$$\begin{aligned} & - \left\{ \xi_{1:2}(t_1, t_2; t_{01}, t_{02} \mid T_3 = t_{03}) + \xi_{2:3}(t_2, t_3; t_{02}, t_{03} \mid T_1 = t_{01}) \right. \\ & \quad \left. + \xi_{1,3}(t_1, t_3; t_{01}, t_{03} \mid T_2 = t_{02}) \right\} \quad \text{sum of 2-way conditional ATIEs} \\ & - \left\{ \tau_1(t_1, t_{02}, t_{03}; t_{01}, t_{02}, t_{03}) + \tau_2(t_{01}, t_2, t_{03}; t_{01}, t_{02}, t_{03}) \right. \\ & \quad \left. + \tau_3(t_{01}, t_{02}, t_3; t_{01}, t_{02}, t_{03}) \right\} \quad \text{sum of (1-way) ATCEs} \end{aligned}$$

- Problems:
 - ① Lower-order *conditional* ATIEs rather than lower-order ATIEs are used
 - ② K -way ATCE \neq sum of all K -way and lower-order ATIEs
 - ③ (We prove) Lack of invariance to the baseline conditions

The K -way Average Marginal Treatment Interaction Effect

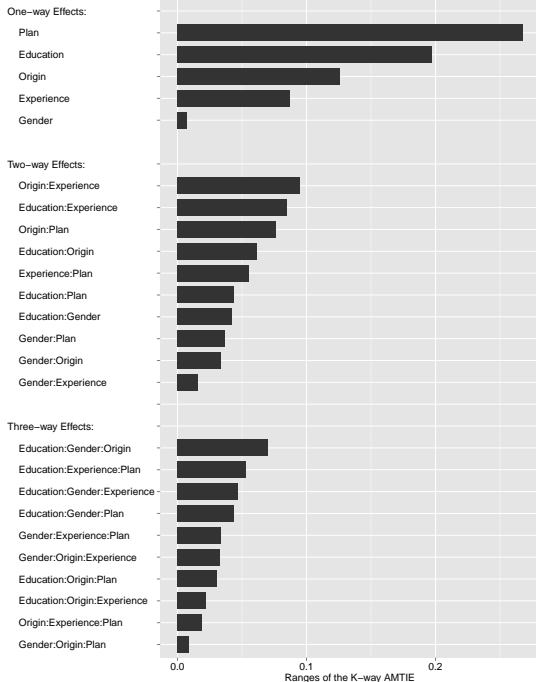
- Definition: the difference between the ATCE and the sum of lower-order AMTIEs
- **Interactive effect interpretation**
- Example: 3-way AMTIE, $\pi_{1:3}(t_1, t_2, t_3; t_{01}, t_{02}, t_{03})$, equals

$$\begin{aligned} & \underbrace{\tau_{1:3}(t_1, t_2, t_3; t_{01}, t_{02}, t_{03})}_{\text{ATCE}} \\ & - \underbrace{\left\{ \pi_{1:2}(t_1, t_2; t_{01}, t_{02}) + \pi_{2:3}(t_2, t_3; t_{02}, t_{03}) + \pi_{1,3}(t_1, t_3; t_{01}, t_{03}) \right\}}_{\text{sum of 2-way AMTIEs}} \\ & - \underbrace{\left\{ \psi(t_1; t_{01}) + \psi(t_2; t_{02}) + \psi(t_3; t_{03}) \right\}}_{\text{sum of (1-way) AMTEs}} \end{aligned}$$

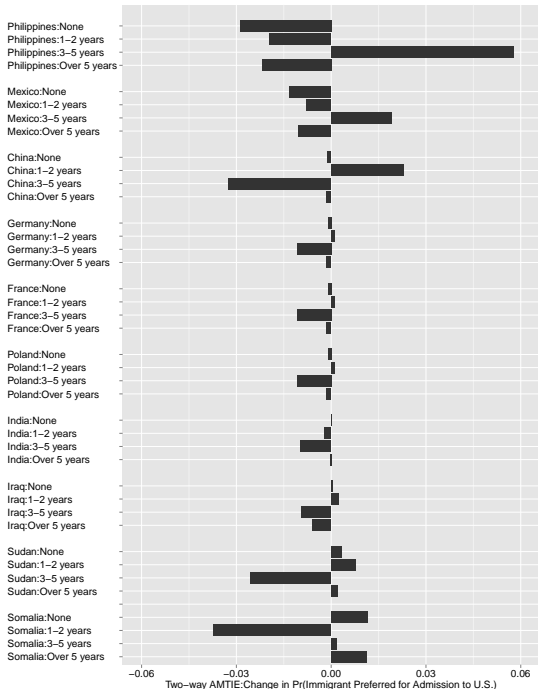
- Properties:
 - ① K -way ATCE = the sum of all K -way and lower-order AMTIEs
 - ② Interval and order invariance to the baseline condition
 - ③ Derive the relationships between the AMTIEs and ATIEs for any order

Empirical Analysis of the Immigration Survey Experiment

- 5 factors (gender², education⁷, origin¹⁰, experience⁴, plan⁴)
 - ① full factorial design assumption
 - ② computational tractability
- Matched-pair conjoint analysis: randomly choose one profile
- Binary outcome: whether a profile is selected
- Model with one-way, two-way, and three-way interaction terms
- The “ $p > n$ ” problem: $p = 1,575$ and $n = 1,396$
- Curse of dimensionality \implies sparsity assumption
- Support vector machine with a lasso constraint (Imai & Ratkovic, 2013)
- Over-identified model that includes baseline conditions
- 99 non-zero and 1,476 zero coefficients
- Cross-validation for selecting a tuning parameter
- **FindIt: Finding heterogeneous treatment effects**



- Range of AMTIEs
- Variation within a factor interaction
- Sparsity-of-effects principle
- gender appears to play a significant role in three-way interactions



- origin × experience interaction
- Baseline: India, None
- Only relative magnitude matters
- Little interaction for European origin
- Similar pattern for Mexico and Phillipines
- Another similar pattern for China, Sudan, and Somalia

Decomposing the Average Treatment Combination Effect

- Two-way effect example (origin \times experience):

$$\begin{aligned} & \underbrace{\tau(\text{Somalia, 1-2 years; India, None})}_{-3.74} \\ = & \underbrace{\psi(\text{Somalia; India})}_{-5.14} + \underbrace{\psi(1 - 2\text{years; None})}_{5.12} + \underbrace{\pi(\text{Somalia, 1 - 2years; India, None})}_{-3.72} \end{aligned}$$

- Three-way examples (education \times gender \times origin):

$$\begin{aligned} & \underbrace{\tau(\text{Graduate, Male, India; Graduate, Female, India})}_{7.46} \\ = & \underbrace{\psi(\text{Male; Female})}_{-0.77} + \underbrace{\pi(\text{Graduate, Male; Graduate, Female})}_{-0.34} \\ & + \underbrace{\pi(\text{Male, India; Female, India})}_{1.56} + \underbrace{\pi(\text{Graduate, Male, India; Graduate, Female, India})}_{7.01} \end{aligned}$$

$$\begin{aligned}
& \underbrace{\tau(\text{High school, Male, Germany; High school, Female, Germany})}_{-11.52} \\
= & \underbrace{\psi(\text{Male; Female})}_{-0.77} + \underbrace{\pi(\text{High school, Male; High school, Female})}_{-0.67} \\
& + \underbrace{\pi(\text{Male, Germany; Female, Germany})}_{-3.34} \\
& + \underbrace{\pi(\text{High school, Male, Germany; High school, Female, Germany})}_{-6.74}.
\end{aligned}$$

Concluding Remarks

- Interaction effects play an essential role in causal heterogeneity
 - ① moderation
 - ② causal interaction
- Two interpretations of causal interaction
 - ① conditional effect interpretation (problematic in high dimension)
 - ② interactive effect interpretation
- Average Marginal Treatment Interaction Effect
 - ① interactive effect in high-dimension
 - ② invariant to baseline condition
 - ③ enables effect decomposition
- Estimation challenges in high dimension

References

- ① Imai, Kosuke and Marc Ratkovic. (2013). “Estimating Treatment Effect Heterogeneity in Randomized Program Evaluation.” *Annals of Applied Statistics*, Vol. 7, No. 1 (March), pp. 443–470.
- ② Egami, Naoki and Kosuke Imai. (2015). “Causal Interaction in High Dimension.” Working Paper available at <http://imai.princeton.edu/research/int.html>

Send comments and suggestions to kimai@Princeton.Edu