# Causal Interaction in Factorial Experiments: Application to Conjoint Analysis

Naoki Egami    Kosuke Imai

Princeton University

Talk at the Institute for Mathematics and Its Applications
University of Minnesota
September 15, 2017

# Causal Heterogeneity and Interaction Effects

1. Moderation:
   - How does the effect of a treatment vary across individuals?
   - Interaction between the treatment variable and pre-treatment covariates

2. Causal interaction:
   - What combination of treatments is efficacious?
   - Interaction among multiple treatment variables

# Conjoint Analysis

- Survey experiments with a factorial design
- Respondents evaluate several pairs of randomly selected profiles defined by multiple factors
- Social scientists use it to analyze multidimensional preferences
- Example: Immigration preference (Hopkins and Hainmueller 2014)
  - representative sample of 1,407 American adults
  - each respondent evaluates 5 pairs of immigrant profiles
  - `gender`$^{2}$, `education`$^{7}$, `origin`$^{10}$, `experience`$^{4}$, `plan`$^{4}$, `language`$^{4}$, `profession`$^{11}$, `application reason`$^{3}$, `prior trips`$^{5}$
  - What combinations of immigrant characteristics do Americans prefer?
  - High dimension: over 1 million treatment combinations
- Methodological challenges:
  - Many interaction effects $\rightsquigarrow$ false positives, difficulty of interpretation
  - Very few applied researchers study interaction

# Factorial Experiments with Two Treatments

- Two factorial treatments (e.g., gender and race):

$$\begin{aligned} A &\in \mathcal{A} = \{a_0, a_1, \ldots, a_{L_A - 1}\} \\ B &\in \mathcal{B} = \{b_0, b_1, \ldots, b_{L_B - 1}\} \end{aligned}$$

- 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

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

# Main Causal Estimands in Factorial Experiments

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

$$\tau_{AB}(a_\ell, b_m; a_0, b_0) \quad = \quad \mathbb{E}\{Y(a_\ell, b_m) - Y(a_0, b_0)\}$$

   - Effect of being Asian male

2. Average Marginal Effect (AME; Hainmueller *et al.* 2014; Dasgupta *et al.* 2015):
   - Average effect of treatment $A = a_\ell$ relative to the baseline condition $A = a_0$ averaging over the other treatment $B$

$$\psi_A(a_\ell, a_0) \quad = \quad \int \mathbb{E}\{Y(a_\ell, B) - Y(a_0, B)\} dF(B)$$

   - Effect of being male averaging over race

# The New Causal Interaction Effect

- Average Marginal Interaction Effect (AMIE):

$$\pi_{AB}(a_\ell, b_m; a_0, b_0) \;=\; \underbrace{\tau_{AB}(a_\ell, b_m; a_0, b_0)}_{\text{ACE of } (a_\ell, b_m)} - \underbrace{\psi_A(a_\ell, a_0)}_{\text{AME of } a_\ell} - \underbrace{\psi_B(b_m, b_0)}_{\text{AME of } b_m}$$

- Interpretation: additional effect induced by $A = a_\ell$ and $B = b_m$ together beyond the separate effect of $A = a_\ell$ and that of $B = b_m$
- Additional effect of being Asian male beyond the sum of separate effects for being male and being Asian
- Decomposition of ACE: $\tau_{AB} = \psi_A + \psi_B + \pi_{AB}$
- Invariance: Unlike the standard interaction effect, the *relative magnitude* of AMIE doesn't depend on the choice of baseline condition

- AMIEs depend on the distribution of treatment assignment:
  1. specified by one's experimental design
  2. motivated by a target population

# Higher-order Causal Interaction

- $J$ factorial treatments with $L_j$ levels each: $\mathbf{T} = (T_1, \ldots, T_J)$
- Assumptions:
  1. Full factorial design

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

  2. Independent treatment assignment

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

- 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

# Higher-order Average Marginal Interaction Effect

- General definition: the difference between ACE and the sum of all lower-order AMIEs (first-order AMIE = AME)
- Example: 3-way AMIE, $\pi_{1:3}(t_1, t_2, t_3; t_{01}, t_{02}, t_{03})$, equals

$$\underbrace{\tau_{1:3}(t_1, t_2, t_3; t_{01}, t_{02}, t_{03})}_{\text{ACE}}$$
$$-\underbrace{\left\{\pi_{1:2}(t_1, t_2; t_{01}, t_{02}) + \pi_{2:3}(t_2, t_3; t_{02}, t_{03}) + \pi_{1:3}(t_1, t_3; t_{01}, t_{03})\right\}}_{\text{sum of all 2-way AMIEs}}$$
$$-\underbrace{\left\{\psi(t_1; t_{01}) + \psi(t_2; t_{02}) + \psi(t_3; t_{03})\right\}}_{\text{sum of AMEs}}$$

- Properties:
  1. $K$-way ACE = the sum of all $K$-way and lower-order AMIEs
  2. Invariance to the baseline condition

# Nonparametric Estimation of AMIE

1. Difference-in-means estimator
   - estimate ACE and AMEs using the difference-in-means estimators
   - estimate AMIE as $\hat{\pi}_{AB} = \hat{\tau}_{AB} - \hat{\psi}_A - \hat{\psi}_B$
   - higher-order AMIEs can be estimated sequentially
   - uses the empirical treatment assignment distribution
2. ANOVA based estimator
   - saturated ANOVA include all interactions up to the $J$th order
   - weighted zero-sum constraints: for all factors and levels,

$$\sum_{\ell=0}^{L_A-1} \Pr(A_i = a_\ell)\beta_\ell^A = 0, \quad \sum_{\ell=0}^{L_A-1} \Pr(A_i = a_\ell)\beta_{\ell m}^{AB} = 0,$$
$$\sum_{m=0}^{L_B-1} \Pr(B_i = b_m)\beta_m^B = 0, \quad \sum_{m=0}^{L_B-1} \Pr(B_i = b_m)\beta_{\ell m}^{AB} = 0, \quad \text{and so on}$$

   - AMIEs are differences of coefficients:

$$\mathbb{E}(\hat{\beta}_\ell^A - \hat{\beta}_0^A) = \psi_A(a_\ell; a_0), \quad \mathbb{E}(\hat{\beta}_{\ell m}^{AB} - \hat{\beta}_{00}^{AB}) = \pi_{AB}(a_\ell, b_m; a_0, b_0)$$

   - can use any marginal treatment assignment distribution of choice

# Conjoint Analysis of Ethnic Voting in Africa

- Ethnic voting and accountability: Carlson (2015, *World Politics*)
- Do voters prefer candidates of same ethnicity regardless of their prior performance? Do ethnicity and performance interact?

- Conjoint analysis in Uganda: 547 voters from 32 villages
- Each voter evaluates 3 pairs of hypothetical candidates
- 5 factors: `Coethnicity`$^2$, `Prior record`$^2$, `Prior office`$^4$, `Platform`$^3$, `Education`$^8$

- `Prior record` = `No` if `Prior office` = `businessman`
  $\rightsquigarrow$ combine these two factors into a single factor with 7 levels
- Collapse `Education` into 2 levels: `relevant degrees` (MA in business, law, economics, development) and `other degrees`

# A Statistical Model of Preference Differentials

- ANOVA regression with one-way and two-way effects:

$$Y_i(\mathbf{T}_i) \;=\; \mu + \sum_{j=1}^{J} \sum_{\ell=0}^{L_j-1} \beta_\ell^j \mathbf{1}\{T_{ij} = \ell\} + \sum_{j \neq j'} \sum_{\ell=0}^{L_j-1} \sum_{m=0}^{L_{j'}-1} \beta_{\ell m}^{jj'} \mathbf{1}\{T_{ij} = \ell, T_{ij'} = m\} + \epsilon_i$$

  with appropriate weighted zero-sum constraints

- In conjoint analysis, we observe the sign of preference differentials
- Linear probability model of preference differential:

$$\begin{aligned}
& \Pr(Y_i(\mathbf{T}_i^*) > Y_i(\mathbf{T}_i^\star) \mid \mathbf{T}_i^*, \mathbf{T}_i^\star) \\
= \;& \mu^* + \sum_{j=1}^{J} \sum_{\ell=0}^{L_j-1} \beta_\ell^j (\mathbf{1}\{T_{ij}^* = \ell\} - \mathbf{1}\{T_{ij}^\star = \ell\}) \\
& + \sum_{j \neq j'} \sum_{\ell=0}^{L_j-1} \sum_{m=0}^{L_{j'}-1} \beta_{\ell m}^{jj'} (\mathbf{1}\{T_{ij}^* = \ell, T_{ij'}^* = m\} - \mathbf{1}\{T_{ij}^\star = \ell, T_{ij'}^\star = m\})
\end{aligned}$$

  where $\mu^* = 0.5$ if the position of profile does not matter

- We apply regularized ANOVA method (Post and Bondell)

# Ranges of Estimated AMEs and AMIEs

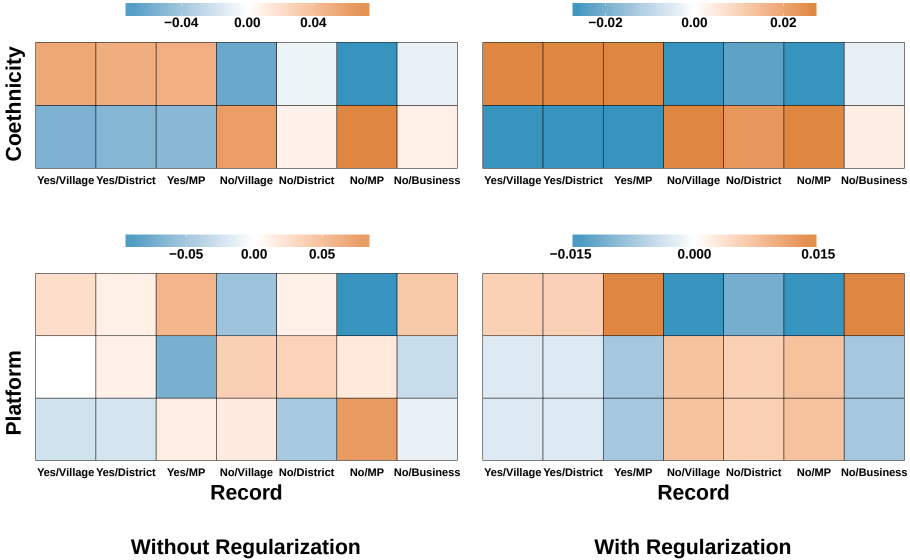
| | Range | Selection prob. |
|---|---|---|
| **AME** | | |
| `Record` | 0.122 | 1.00 |
| `Coethnicity` | 0.053 | 1.00 |
| `Platform` | 0.023 | 0.93 |
| `Degree` | 0.000 | 0.33 |
| **AMIE** | | |
| `Coethnicity` × `Record` | 0.053 | 1.00 |
| `Record` × `Platform` | 0.030 | 0.92 |
| `Platform` × `Coethnic` | 0.008 | 0.64 |
| `Coethnicity` × `Degree` | 0.000 | 0.62 |
| `Platform` × `Degree` | 0.000 | 0.35 |
| `Record` × `Degree` | 0.000 | 0.09 |

- Factor selection probability based on bootstrap

# Close Look at the Estimated AMEs

| Factor | AME | Selection prob. |
|---|---|---|
| Record | | |
| Yes/Village | 0.122 | ⟩ 0.71 |
| Yes/District | 0.122 | ⟩ 0.77 |
| Yes/MP | 0.101 | ⟩ 1.00 |
| No/Village | 0.047 | ⟩ 0.74 |
| No/District | 0.051 | ⟩ 0.74 |
| No/MP | 0.047 | ⟩ 1.00 |
| No/Businessman | base | |
| Platform | | |
| Jobs | $-0.023$ | ⟩ 0.56 |
| Clinic | $-0.023$ | ⟩ 0.94 |
| Education | base | |
| Coethnicity | 0.054 | 1.00 |
| Degree | 0.000 | 0.33 |

# Effect of Regularization on AMIEs



# Decomposition and Conditional Effects

- Decomposition of ACE (Coethnicity × Record interaction):

$$\underbrace{\tau(\texttt{Coethnic, No/Business}; \texttt{Non-coethnic, No/MP})}_{-2.4}$$

$$= \underbrace{\psi(\texttt{Coethnic}; \texttt{Non-coethnic})}_{5.4} + \underbrace{\psi(\texttt{No/Business}; \texttt{No/MP})}_{-4.7}$$

$$+ \underbrace{\pi(\texttt{Coethnic, No/Business}; \texttt{Non-coethnic, No/MP})}_{-3.1}$$

- Conditional effects (Platform × Record interaction):
  - AMIE: $\pi(\texttt{Education, No/MP}\}; \{\texttt{Job, No/MP}\}) = -2.3$
  - Conditional effect of Education relative to Job for No/MP is approximately zero
  - AME: $\psi(\texttt{Education}; \texttt{Job}) = 2.3$

# Concluding Remarks

- Interaction effects play an essential role in causal heterogeneity
  1. moderation
  2. causal interaction

- Randomized experiments with a factorial design
  1. useful for testing multiple treatments and their interactions
  2. social science applications: audit studies, conjoint analysis
  3. challenge: estimation and interpretation in high dimension

- Average Marginal Interaction Effect (AMIE)
  1. invariant to baseline condition
  2. straightforward interpretation even for high order interaction
  3. enables effect decomposition
  4. enables regularization through ANOVA