Estimating Heterogeneous Causal Effects of High-Dimensional Treatments Using Interpretable Machine Learning: Application to Conjoint Analysis

Kosuke Imai

Harvard University

Joint Statistical Meetings
August 11, 2021

Joint work with
Max Goplerud (U. of Pittsburgh) Nicole E. Pashley (Rutgers)

Causal Heterogeneity and Interaction Effects

1. Causal moderation (Heterogeneous treatment effects):
   - 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

3. Causal moderation + Causal interaction:
   - What treatment combinations are efficacious for what types of individuals?
   - Identify causal heterogeneity while maintaining interpretability

Conjoint Analysis

- Survey experiments with a high-dimensional 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

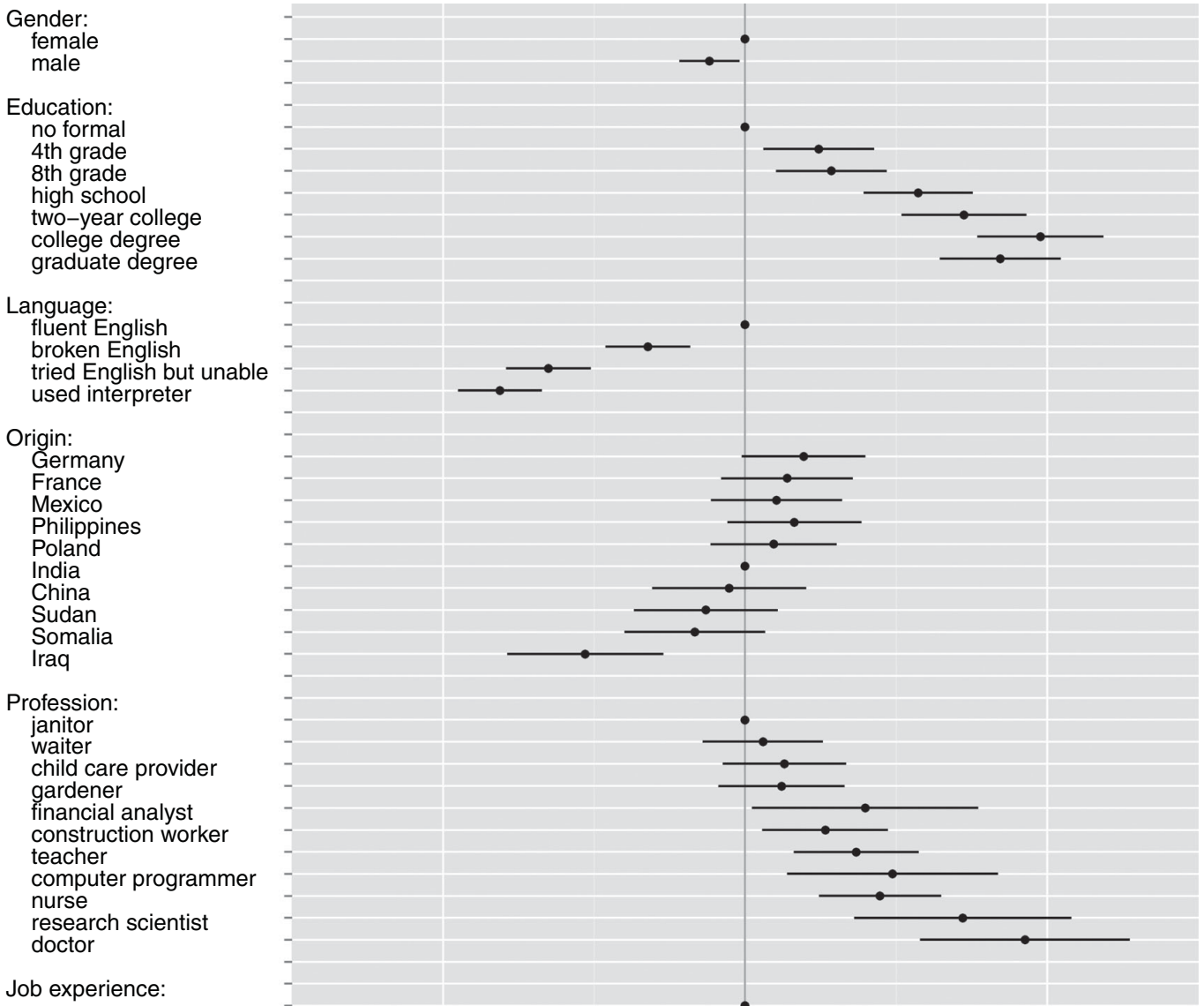
Please read the descriptions of the potential immigrants carefully. Then, please indicate which of the two immigrants you would personally prefer to see admitted to the United States.

	Immigrant 1	Immigrant 2
Prior Trips to the U.S.	Entered the U.S. once before on a tourist visa	Entered the U.S. once before on a tourist visa
Reason for Application	Reunite with family members already in U.S.	Reunite with family members already in U.S.
Country of Origin	Mexico	Iraq
Language Skills	During admission interview, this applicant spoke fluent English	During admission interview, this applicant spoke fluent English
Profession	Child care provider	Teacher
Job Experience	One to two years of job training and experience	Three to five years of job training and experience
Employment Plans	Does not have a contract with a U.S. employer but has done job interviews	Will look for work after arriving in the U.S.
Education Level	Equivalent to completing two years of college in the U.S.	Equivalent to completing a college degree in the U.S.
Gender	Female	Male

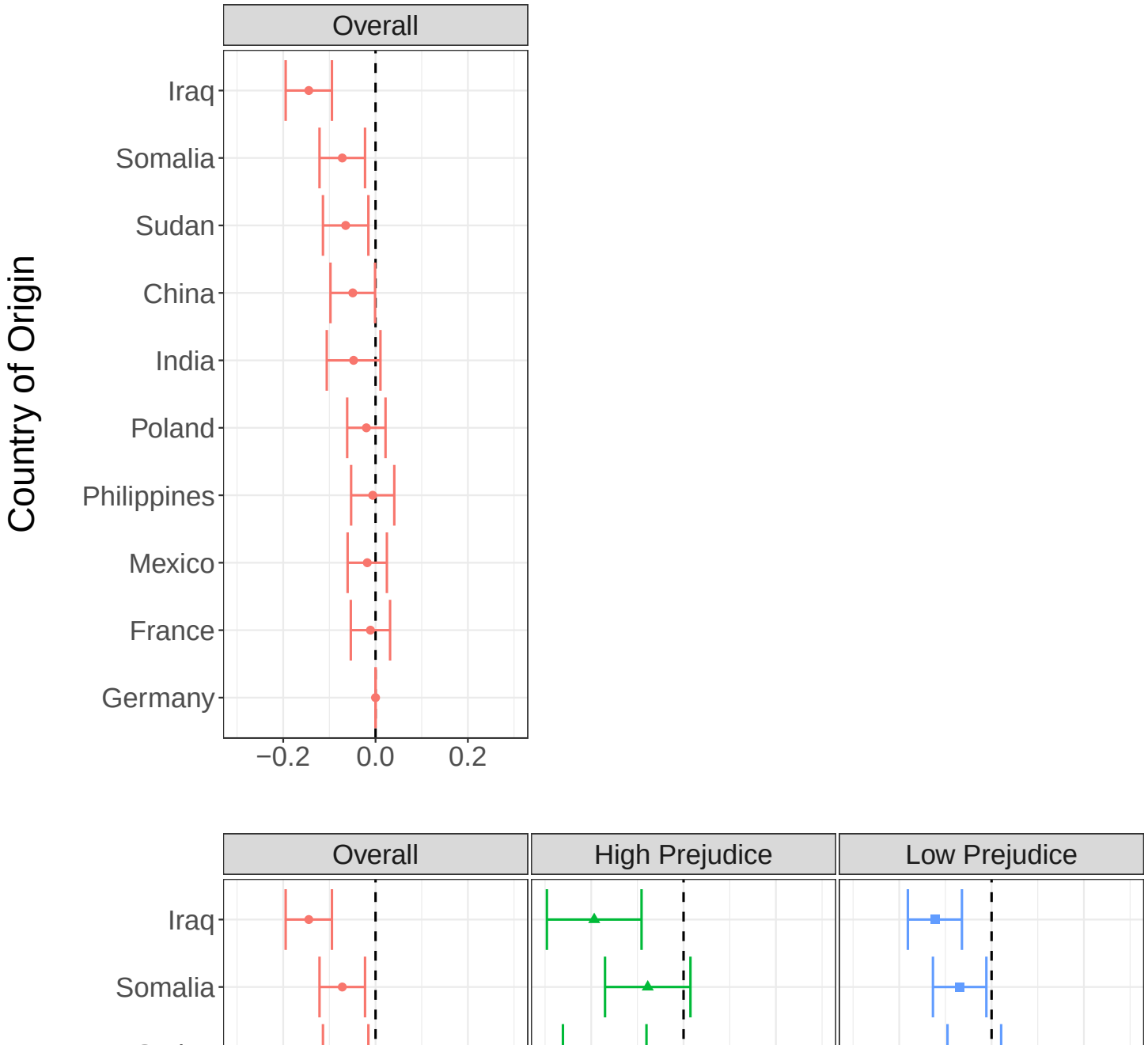
	Immigrant 1	Immigrant 2
If you had to choose between them, which of these two immigrants should be given priority to come to the United States to live?	○	○

Average Marginal Component Effects (AMCE)

- Average marginal effect of one factor level relative to its baseline level averaging over the empirical distribution of the other factors



Effects of Country of Origin



Factorial Design

- Setup:
  - N units
  - J factors
  - $L_j \geq 2$ levels for factor j
  - Treatment: $T_{ij} \in \{0, 1, \ldots, L_j - 1\}$
  - Potential outcome: $Y_i(\boldsymbol{t})$ where $\boldsymbol{t} \in \mathcal{T}$
  - Observed outcome: $Y_i = Y_i(\boldsymbol{T}_i)$
  - Pre-treatment covariates: $\boldsymbol{X}_i$
- Randomization:

$$\{Y_i(\boldsymbol{t})\}_{\boldsymbol{t} \in \mathcal{T}} \perp\!\!\!\perp \boldsymbol{T}_i$$

- AMCE of factor j level l relative to baseline level l':

$$\delta_j(l, l') \;=\; \mathbb{E}\{Y_i(T_{ij} = l, \boldsymbol{T}_{i,-j}) - Y_i(T_{ij} = l', \boldsymbol{T}_{i,-j})\},$$

  - average over the distribution of $T_{i,-j}$
  - average over the distribution of units

Modeling Heterogeneous Effects

- Bayesian finite mixture of regularized regressions
  - regularized regression $\rightsquigarrow$ sparsity
  - finite mixture modeling $\rightsquigarrow$ heterogeneity
  - incorporate moderators $\rightsquigarrow$ predicting cluster membership

- K clusters

$$\Pr(Y_i = 1 \mid \boldsymbol{T}_i, \boldsymbol{X}_i) = \sum_{k=1}^{K} \pi_k(\boldsymbol{X}_i)\zeta_k(\boldsymbol{T}_i)$$

where

$$\zeta_k(\boldsymbol{T}_i) = \frac{\exp(\psi_k(\boldsymbol{T}_i))}{1 + \exp(\psi_k(\boldsymbol{T}_i))}, \quad \pi_k(\boldsymbol{X}_i) = \frac{\exp(\boldsymbol{X}_i^\top \phi_k)}{\sum_{k'=1}^{K} \exp(\boldsymbol{X}_i^\top \phi_{k'})}.$$

and we set $\phi_1 = 0$ for identification

The Outcome Model

- Linear predictor with main effects and two-way interactions:

$$\psi_k(\boldsymbol{T}_i) = \mu + \sum_{j=1}^{J}\sum_{l=0}^{L_j-1} 1\{T_{ij}=l\}\beta_{kl}^{j} + \sum_{j=1}^{J-1}\sum_{j'>j}\sum_{l=0}^{L_j-1}\sum_{l'=0}^{L_{j'}-1} 1\{T_{ij}=l, T_{ij'}=l'\}\beta_{kll'}^{jj'},$$

- ANOVA constraints:

$$\sum_{l=0}^{L_j-1}\beta_{kl}^{j}=0, \quad \text{and} \quad \sum_{l=0}^{L_j-1}\beta_{kll'}^{jj'}=0,$$

for each $j, j' = 1, 2, \ldots, J$ with $j' > j$, and $l' = 0, 1, \ldots, L_{j'}$

Regularization

- Goal: Fuse levels l_1 and l_2 of factor j when main effects and interaction effects are similar
  - main effects: $\beta^j_{l_1} \approx \beta^j_{l_2}$
  - interaction effects: $\beta^{jj'}_{l_1 l'} \approx \beta^{jj'}_{l_2 l'}$ for all j' and l'
- 2 factor example: ℓ_2 regularization for computational simplicity

$$\underbrace{\sqrt{(\beta^1_0 - \beta^1_1)^2 + (\beta^{12}_{00} - \beta^{12}_{10})^2 + (\beta^{12}_{01} - \beta^{12}_{11})^2}}_{\text{fuse levels 0 and 1 of factor 1}}$$

$$\underbrace{\sqrt{(\beta^2_0 - \beta^2_1)^2 + (\beta^{12}_{00} - \beta^{12}_{01})^2 + (\beta^{12}_{10} - \beta^{12}_{11})^2 + (\beta^{12}_{20} - \beta^{12}_{21})^2}}_{\text{fuse levels 0 and 1 of factor 2}}.$$

- Regularization as a Bayesian prior:

$$p\left(\boldsymbol{\beta}_k \mid \{\boldsymbol{\phi}_k\}_{k=2}^K\right) \propto \left(\lambda \bar{\pi}_k^\gamma\right)^m \exp\left(-\lambda \bar{\pi}_k^\gamma \sum_{g=1}^G \sqrt{\boldsymbol{\beta}_k^\top \boldsymbol{F}_g \boldsymbol{\beta}_k}\right),$$

where $\bar{\pi}_k = \sum_{i=1}^N \pi_k(\boldsymbol{X}_i)/N$ and $m = \text{rank}\left([\boldsymbol{F}_1, \cdots, \boldsymbol{F}_G]\right)$

Estimation and Inference

- BIC to choose the value of regularization parameter
- EM algorithm using data augmentation
  - Polya-Gamma augmentation for logistic regression
  - Another data augmentation for sparsity-inducing penalty
- Inference based on the log posterior given the fused levels

Empirical Analysis: Forced Choice Design

- The symmetry assumption leads to

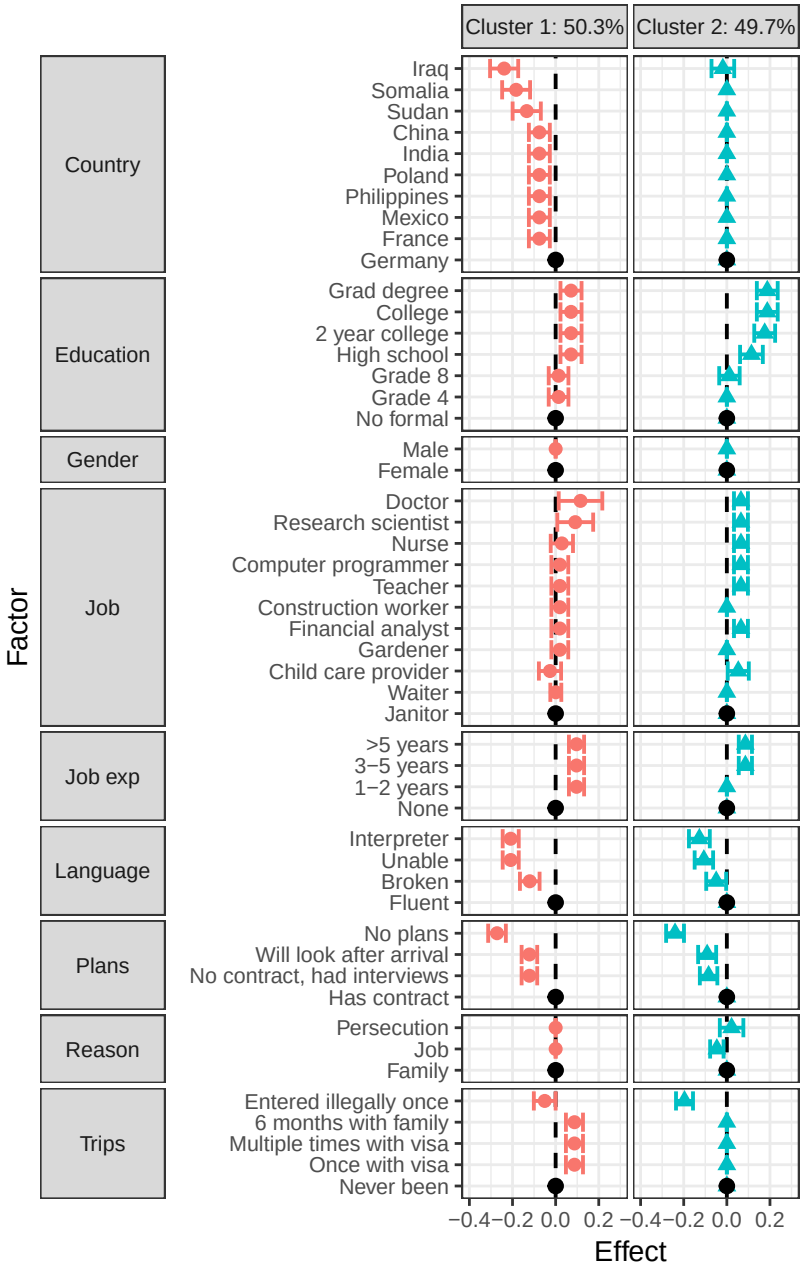
$$\psi_k(\boldsymbol{T}_i^L, \boldsymbol{T}_i^R) = \mu + \sum_{j=1}^{J} \sum_{l \in L_j} \beta_{kl}^j \left(1\left\{T_{ij}^L = l\right\} - 1\left\{T_{ij}^R = l\right\}\right)$$
$$+ \sum_{j=1}^{J-1} \sum_{j'>j} \sum_{l \in L_j} \sum_{l' \in L_{j'}} \beta_{kll'}^{jj'} \left(1\left\{T_{ij}^L = l, T_{ij'}^L = l'\right\} - 1\left\{T_{ij}^R = l, T_{ij'}^R = l'\right\}\right)$$

  where $\boldsymbol{T}_i^L$ and $\boldsymbol{T}_i^R$ represent the factors for the left and right profiles

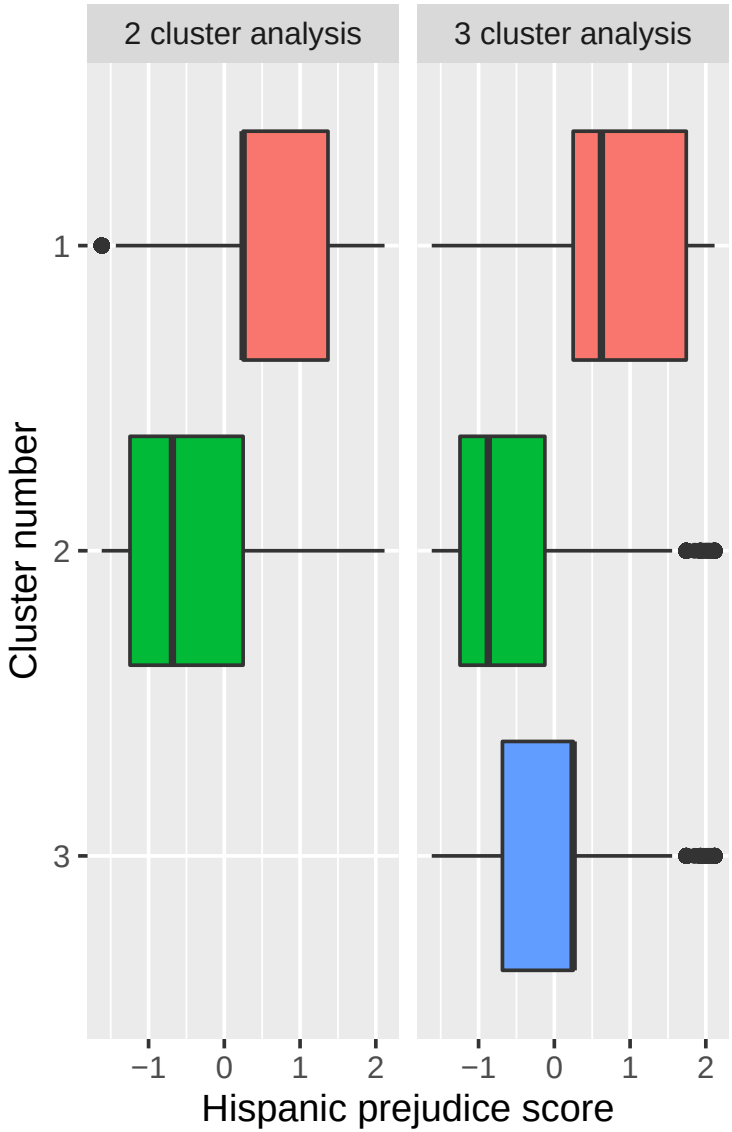
- AMCE:

$$\frac{1}{2}\mathbb{E}\left[\left\{\Pr\left(Y_i = 1 \mid Z_i = k, T_{ij}^L = l, \boldsymbol{T}_{i,-j}^L, \boldsymbol{T}_i^R\right)\right.\right.$$
$$\left. - \Pr\left(Y_i = 1 \mid Z_i = k, T_{ij}^L = l', \boldsymbol{T}_{i,-j}^L, \boldsymbol{T}_i^R\right)\right\}$$
$$+ \left\{\Pr\left(Y_i = 0 \mid Z_i = k, T_{ij}^R = l, \boldsymbol{T}_{i,-j}^R, \boldsymbol{T}_i^L\right)\right.$$
$$\left.\left. - \Pr\left(Y_i = 0 \mid Z_i = k, T_{ij}^R = l', \boldsymbol{T}_{i,-j}^R, \boldsymbol{T}_i^L\right)\right\}\right].$$

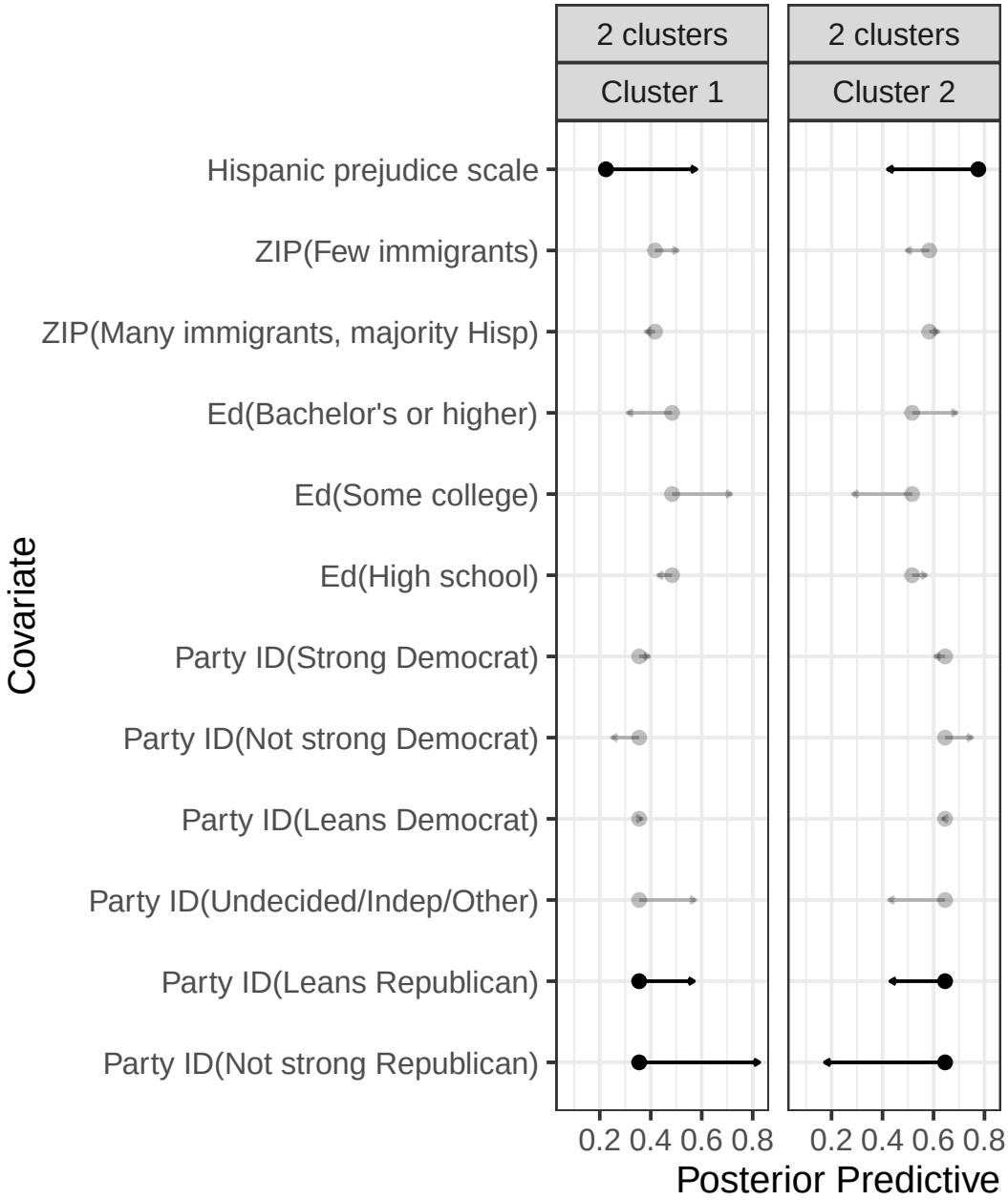
Estimated AMCEs: 2-cluster and 3-cluster Models



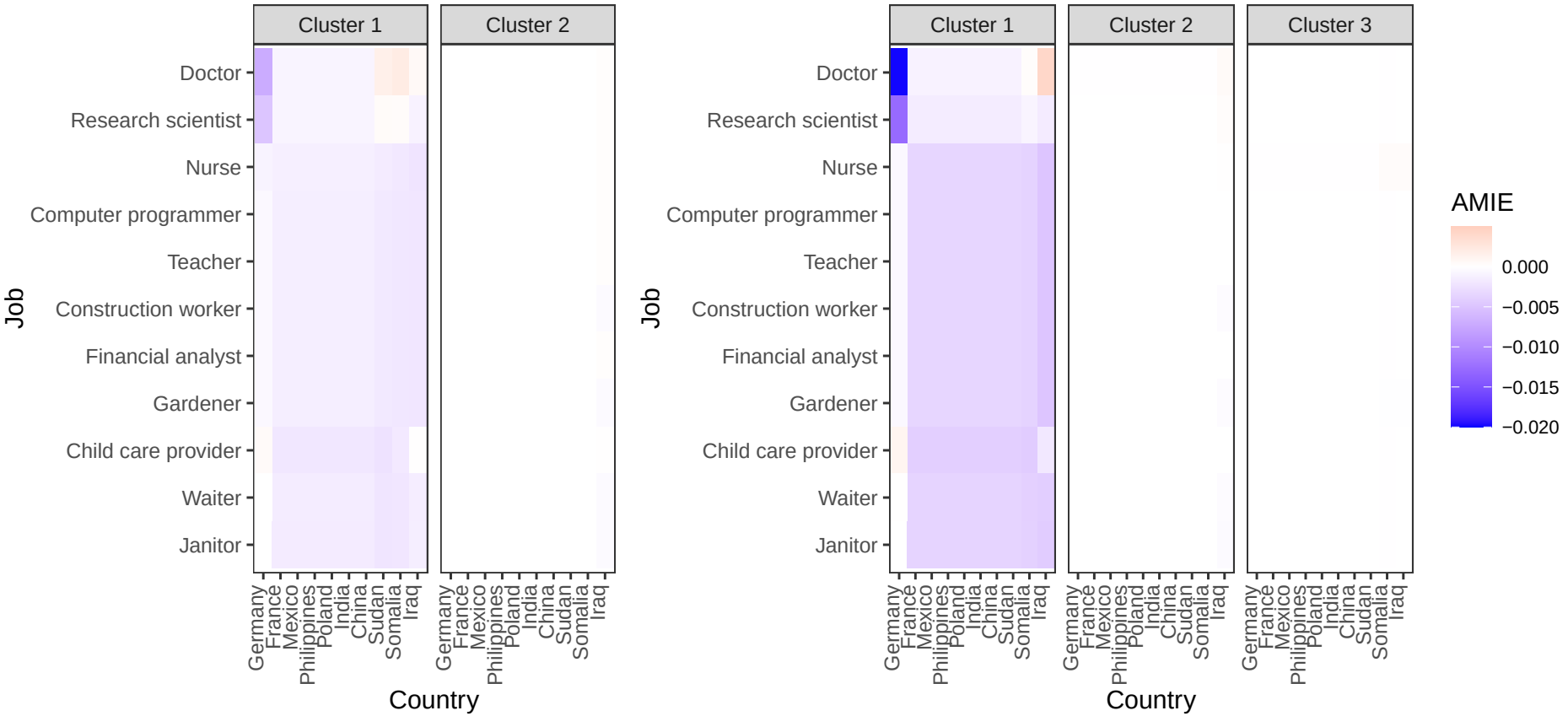
Moderators and Cluster Membership



Marginal Effects of Moderators



Interaction Effects



- Very small interaction effects
- But, consistent with the "skill premium theory" of Newman and Malhotra (2019)

Concluding Remarks

- Interaction effects play an essential role in causal heterogeneity
  1. causal moderation
  2. causal interaction
- Need for interpretable machine learning methods

- High-dimensional factorial design: moderation + interaction
  1. social science applications: conjoint analysis, audit studies,
  2. finite mixture of regularized regressions

- Conjoint analysis of immigration preferences
- Preliminary finding: group of respondents who give priority to country of origin, are less educated, more likely to be Republicans, and tend to live in areas with few immigrants.