Multiple Hypothesis Testing in Conjoint Analysis

Yuki Shiraito

University of Michigan

Politics, Sandwiches, and Comments
Department of Government, Cornell University

November 5, 2021

Joint work with Guoer Liu (U-M)

Conjoint Analysis

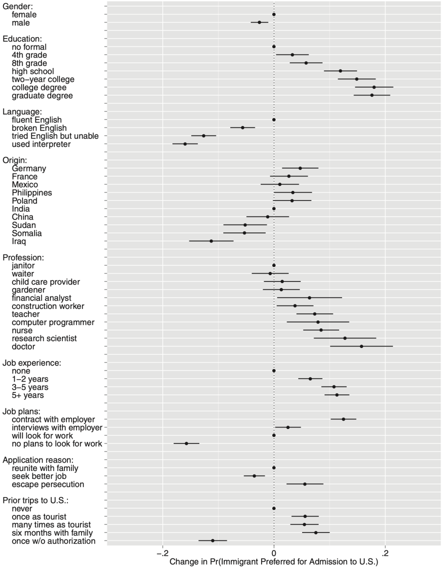
- Conjoint Design

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

- AMCE: test multiple causal hypotheses ***simultaneously***

Classic Conjoint Results

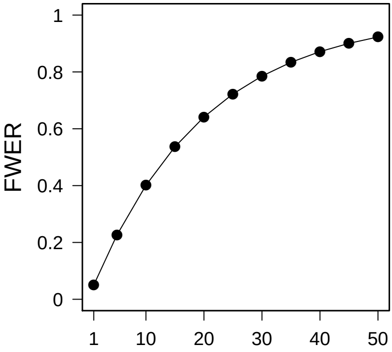


Multiple Hypothesis Testing

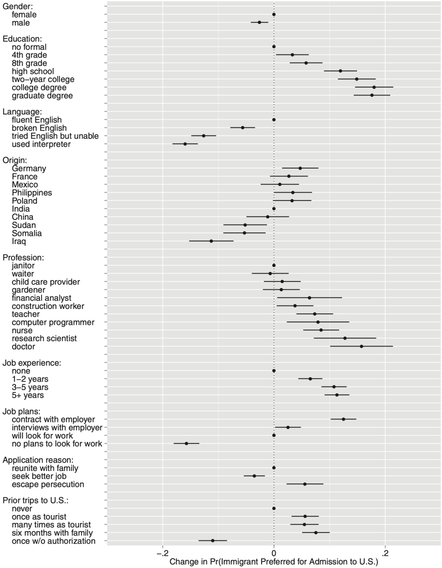
- Test one hypothesis, $\alpha \equiv \mathbb{P}(\text{Reject null} \mid \text{Null is true}) = 0.05$
- Test ten hypotheses simultaneously with $\alpha = 0.05$

$$\begin{aligned} \text{FWER} \equiv & \mathbb{P}(\text{At least one null is rejected} \mid \text{All nulls are true}) \\ = & 1 - (1-\alpha)^{10} \approx .4 \end{aligned}$$

- Theoretical results

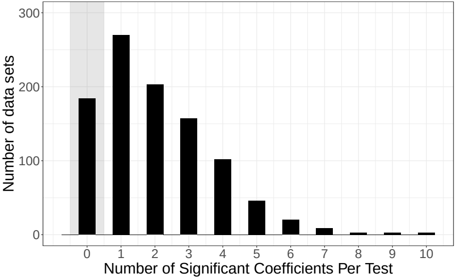


Number of Hypotheses in Conjoint Analysis: 41

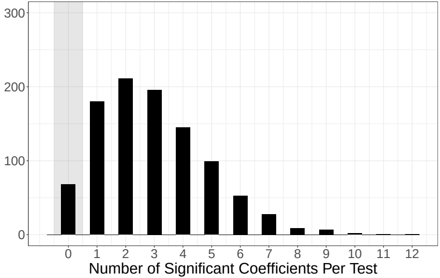


Quantifying the Problem by Simulations

- If AMCE is zero, in how many samples do you get false findings?
- Two scenarios for 41 attribute levels:
  1. No *individual* effect
  2. Nonzero individual effect, but zero *average* effect
- Number of samples for each number of false findings:



(a) Zero Individual MCE



(b) Nonzero Individual MCE but Zero AMCE

Correction Methods Overview

- Objective: contain *false positive* conclusions
- Trade-off: risk *false negative* conclusions

- Correction methods
  - Control *family-wise error rate* (FWER)
    - **Bonferroni Correction**
  - Control *false discovery rate* (FDR)
    - **Benjamini-Hochberg Procedure**
  - Control *false discovery rate* (FDR) & Reduce RMSE
    - **Adaptive Shrinkage**

- Proposal:

Bonf. **ASh** **BH**

Confirmatory ⟷ Exploratory

Bonferroni Correction

- Controls FWER to α
- Procedure: set $\alpha^* = \frac{\alpha}{\text{\# of tests}}$ for each test
- Strength: easy to construct confidence intervals
- Shortcomings:
  - high risk of false negative conclusions
  - ambiguous definition of "total number of tests"

Benjamini-Hochberg Procedure

- Controls FDR:

$$\mathbb{E}\left[\frac{\text{\# of false discoveries}}{\text{\# of total discoveries}}\right] \leq \alpha$$

- Solution:
  1. Rank p-values from smallest to largest
  2. Reject the null up to the largest p-value such that

$$p \leq \frac{\text{rank of } p}{\text{\# of tests}}\,\alpha$$

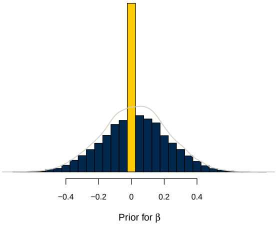
- Strength: less susceptive to false negative conclusion

- Shortcomings:
  - sensitive to pre-specified FDR
  - no uncertainty measures

Adaptive Shrinkage

- Regularizes β by placing a *spike-and-slab* prior

$$p(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}, \hat{\sigma}) \propto \underbrace{p(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta}, \hat{\sigma})}_{\text{Likelihood}} \underbrace{p(\boldsymbol{\beta}|\hat{\sigma})}_{\text{Prior}}$$



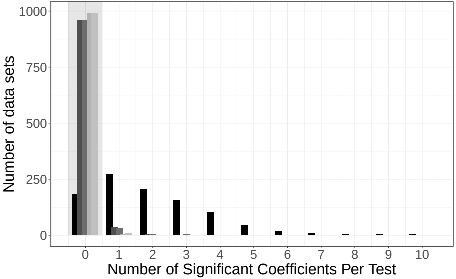
- Procedure: empirical Bayes post-estimation procedure
- Strength:

  transparent, flexible, credible interval
  more precise point estimates

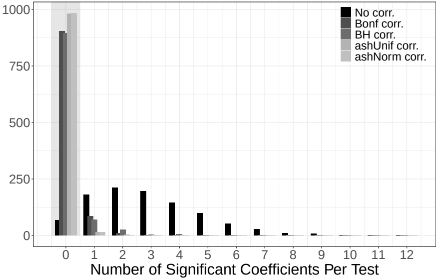
Simulations

- Design matrix based on the immigration conjoint by Heimuller et al.
- Avoiding false positives: zero AMCE
  1. No *individual* effect
  2. Nonzero individual effect, but zero *average* effect
- Avoiding both false positives and false negatives: nonzero AMCE
  1. Only *gender* has effect (appendix)
  2. All levels of *gender, education, English* have effects

Zero AMCE



(a) Zero Individual MCE



(b) Nonzero Individual MCE but Zero AMCE

Nonzero AMCE

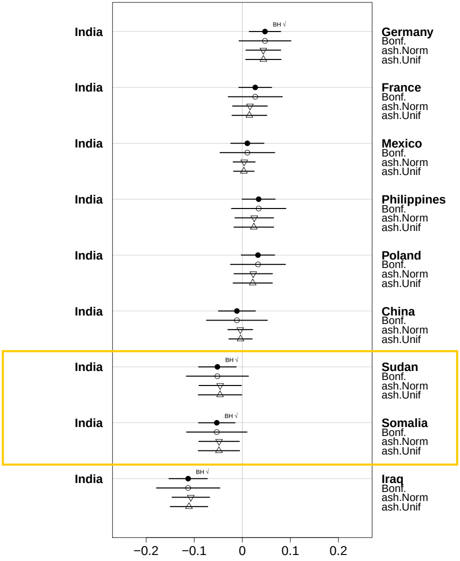
			No. of False Positives									
			0	1	2	3	4	5	6	7	8	9
No. of True Positives	No corr.	9	2	8	3	1	4	1				
		10	258	270	196	133	54	42	13	10	4	1
	Bonf corr.	8	38									
		9	305	6	2							
		10	623	25	1							
	BH corr.	8	4									
		9	47	25	4		1					
		10	607	208	66	23	7	6	2			
	ashUnif corr.	8	17	2								
		9	160	26	4	1		1				
		10	620	127	30	6	5	1				
	ashNorm corr.	8	21	2								
		9	172	29	3	1	1					
		10	647	99	14	7	4					

- Correct number of positives: 10

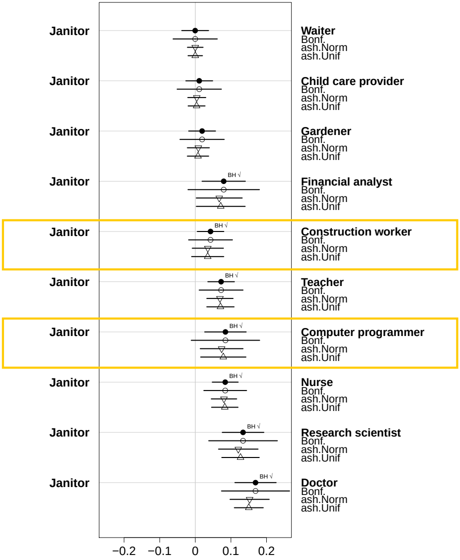
Reanalysis

- Immigrants preferred by the U.S. public (Hainmueller et al. 2014)
  - Focus on *Country of Origin* and *Profession*
  - To show:
    1. How corrected results differ
    2. ASh attains the middle
- Trading partners preferred in Vietnam (Spiker et al. 2016)
  - Focus on *Military Ally* and *Environmental Standards*
  - To show:
    1. Bonf. and ASh recovers the null correctly
    2. BH does not correct at all with few number of discoveries

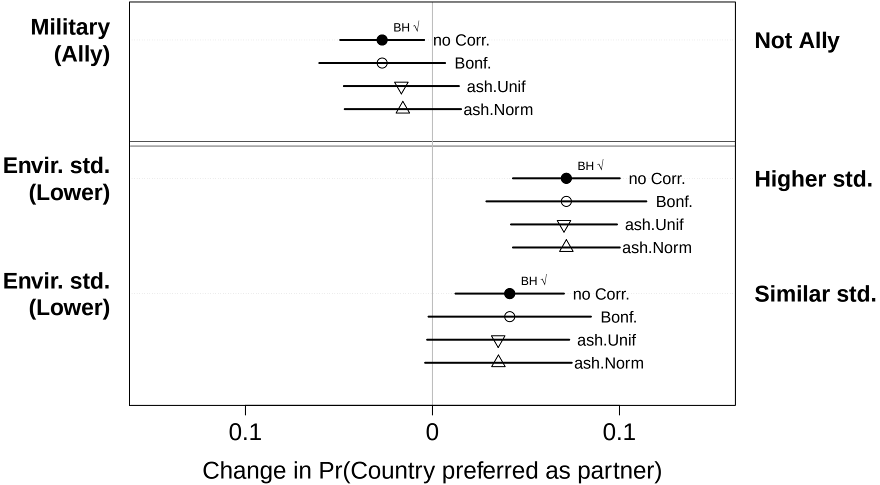
Country of Origin



Profession


Janitor
Waiter
Bonf.
ash.Norm
ash.Unif
Janitor
Child care provider
Bonf.
ash.Norm
ash.Unif
Janitor
Gardener
Bonf.
ash.Norm
ash.Unif
Janitor
BH √
Financial analyst
Bonf.
ash.Norm
ash.Unif
Janitor
BH √
Construction worker
Bonf.
ash.Norm
ash.Unif
Janitor
BH √
Teacher
Bonf.
ash.Norm
ash.Unif
Janitor
BH √
Computer programmer
Bonf.
ash.Norm
ash.Unif
Janitor
BH √
Nurse
Bonf.
ash.Norm
ash.Unif
Janitor
BH √
Research scientist
Bonf.
ash.Norm
ash.Unif
Janitor
BH √
Doctor
Bonf.
ash.Norm
ash.Unif
−0.2
−0.1
0
0.1
0.2

Selecting Trading Partners in Vietnam



Concluding Remarks

- Conjoint analysis inherently needs multiple hypothesis testing
- No correction $\rightsquigarrow$ danger of false findings
- Correction methods
  - Bonferroni Correction (Most conservative)
  - Benjamini-Hochberg Procedure (Least conservative)
  - Adaptive shrinkage (middle-ground)

Bonf. **ASh** **BH**

Confirmatory Exploratory

- Do correction, or you will get at least one false result

ASh Model

- Model: $\boldsymbol{\beta} = (\beta_1, ..., \beta_J)$; est. $\hat{\boldsymbol{\beta}}$, std.err $\hat{\sigma}$

$$p(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}, \hat{\sigma}) \propto \underbrace{p(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta}, \hat{\sigma})}_{\text{Likelihood}} \underbrace{p(\boldsymbol{\beta}|\hat{\sigma})}_{\text{Prior}}$$

$$\beta_1, ..., \beta_J \overset{iid}{\sim} g$$

where

$$g(\cdot; \boldsymbol{\pi}) = \pi_0 \delta_0(\cdot) + \sum_{k=1}^{K} \pi_k \mathcal{N}(\cdot; 0, \delta_k^2),$$

$$\sum_{k=0}^{K} \pi_k = 1 \quad \text{and} \quad \pi_k \geq 0$$

- Emprical Bayes estimates:

$$\hat{\boldsymbol{\pi}} = \underset{\boldsymbol{\pi}}{\text{argmax}} \prod_{j=1}^{J} \sum_{k=0}^{K} \pi_k \mathcal{N}(\hat{\beta}_j; 0, \delta_k^2 + \hat{s}_j^2)$$

Simulation Result: Only One Nonzero AMCE

			No. of False Positives								
			0	1	2	3	4	5	6	7	8
No. of True Positives	No corr.	1	230	290	215	123	69	42	19	9	3
	Bonf. corr.	1	966	32	2						
	BH corr.	1	931	61	7	1					
	ashUnif corr.	1	996	4							
	ashNorm corr.	1	998	2							

$$\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, 0.01^2)$$

Simulation Result: Only One Nonzero AMCE

			No. of False Positives												
			0	1	2	3	4	5	6	7	8	9	10	11	12
No. of True Positives	No corr.	1	237	253	223	134	83	38	17	6	2	6			1
	Bonf. corr.	1	962	37	1										
	BH corr.	1	930	55	7	5	1	1	1						
	ashUnif corr.	1	984	14	2										
	ashNorm corr.	1	987	12	1										

$$\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, 0.1^2)$$

Simulation Result: Nonzero AMCE in Each Attribute

			No. of False Positives												
			0	1	2	3	4	5	6	7	8	9	10	11	12
	No corr.	7		2				1							
		8	10	22	27	16	22	8	2	3	1				
		9	118	194	179	169	86	58	39	19	13	7	2	1	1
	Bonf corr.	5	7	3											
		6	77	5	2										
		7	244	15	7										
		8	396	37	5										
No. of True Positives		9	180	20	2										
	BH corr.	6	5	2											
		7	37	15	5	1	1								
		8	147	89	36	11	4	1		3					
		9	321	187	75	35	12	8	1	3	1				
	ashUnif corr.	6	12	3	1	1									
		7	84	25	4	1	1								
		8	220	99	23	12	1	1							
		9	294	130	46	29	8	2	2		1				
	ashNorm corr.	5		1											
		6	11	5	2	1									
		7	98	21	5	2									
		8	224	100	24	10	1	1							
		9	295	124	42	21	7	2	2	1					

Figure: The true AMCE for each attribute has one significant levels I.

Simulation Result: Nonzero AMCE in Each Attribute

			No. of False Positives													
			0	1	2	3	4	5	6	7	8	9	10	11	12	13
No. of True Positives	No corr.	6			5	7	4	4	1	1						
		7			41	46	34	17	8	9	3					
		8			115	100	88	52	22	16	12	6	1		2	1
		9			100	116	82	49	31	17	5	1	4			
	Bonf corr.	4		1	37											
		5		2	247	14	1									
		6		4	365	15	1									
		7		4	224	7	3	1								
		8		2	63	2										
		9			7											
	BH corr.	4			3											
		5			32	4	2									
		6			106	28	7	4	2							
		7			212	70	17	8	1	1						
		8			229	82	38	9	7	2	1	1				
		9			77	34	13	5	3	2						
	ashUnif corr.	4			2	1		1								
		5		1	52	13	4									
		6		1	176	50	13	5								
		7			233	72	14	11	1	1						
		8			180	62	23	6	1	1	2					
		9			40	20	10	2	1	1						
	ashNorm corr.	4			4			1								
		5		1	47	13	4									
		6		1	174	49	11	3								
		7		234	71	17	8		1							
		8			187	63	23	7	1	2	1					
		9			43	20	11	1	2							

Figure: The true AMCE for each attribute has one significant levels II. The standard deviation for the reference category of *Job Experience* is four times larger.

Simulation Result: ASh RMSE