Political scientists study multi-dimensional choices and preferences
Motivation

- Political scientists study multi-dimensional choices and preferences
- Survey experiments are a prominent tool for causal inference
  - In 2006–2010, APSR, AJPS and JOP published 72 articles using survey experiments

Yet, survey experiments are not perfect for testing causal theories

Example: Brader, Valentino and Suhay (2008)
- Treatment: Immigrants' country of origin in mock newspaper stories
- Mexican immigrant (vs. Russian) decreased support for immigration
- But what is it about the Mexican immigrant that caused the decrease?
  - Is it the country of origin itself, or perceived income, education, religion, or something else?

Treatments are typically composites of multiple causal components
- Effects of these attributes are aliased with one another
Political scientists study multi-dimensional choices and preferences

Survey experiments are a prominent tool for causal inference
- In 2006–2010, APSR, AJPS and JOP published 72 articles using survey experiments

Yet, survey experiments are not perfect for testing causal theories
Motivation

- Political scientists study multi-dimensional choices and preferences
- Survey experiments are a prominent tool for causal inference
  - In 2006–2010, APSR, AJPS and JOP published 72 articles using survey experiments
- Yet, survey experiments are not perfect for testing causal theories
- Example: Brader, Valentino and Suhay (2008)
  - Treatment: Immigrants' country of origin in mock newspaper stories
  - Mexican immigrant (vs. Russian) decreased support for immigration
Motivation

- Political scientists study multi-dimensional choices and preferences
- Survey experiments are a prominent tool for causal inference
  - In 2006–2010, APSR, AJPS and JOP published 72 articles using survey experiments
- Yet, survey experiments are not perfect for testing causal theories
- Example: Brader, Valentino and Suhay (2008)
  - Treatment: Immigrants’ country of origin in mock newspaper stories
  - Mexican immigrant (vs. Russian) decreased support for immigration
  - But what is it about the Mexican immigrant that caused the decrease?
Motivation

- Political scientists study multi-dimensional choices and preferences
- Survey experiments are a prominent tool for causal inference
  - In 2006–2010, APSR, AJPS and JOP published 72 articles using survey experiments
- Yet, survey experiments are not perfect for testing causal theories
- Example: Brader, Valentino and Suhay (2008)
  - Treatment: Immigrants’ country of origin in mock newspaper stories
  - Mexican immigrant (vs. Russian) decreased support for immigration
  - But what is it about the Mexican immigrant that caused the decrease?
  - Is it the country of origin itself, or perceived income, education, religion, or something else?
Motivation

- Political scientists study multi-dimensional choices and preferences
- Survey experiments are a prominent tool for causal inference
  - In 2006–2010, APSR, AJPS and JOP published 72 articles using survey experiments
- Yet, survey experiments are not perfect for testing causal theories
- Example: Brader, Valentino and Suhay (2008)
  - Treatment: Immigrants’ country of origin in mock newspaper stories
  - Mexican immigrant (vs. Russian) decreased support for immigration
  - But what is it about the Mexican immigrant that caused the decrease?
  - Is it the country of origin itself, or perceived income, education, religion, or something else?
- Treatments are typically composites of multiple causal components
- Effects of these attributes are aliased with one another
Typical experiments can only identify treatment effects as a whole.
Yet, scientific theories are often about effects of specific components.
Conjoint Analysis as a Tool for Causal Inference

- Typical experiments can only identify treatment effects as a whole.
- Yet, scientific theories are often about effects of specific components.
- **Conjoint analysis**: A popular experimental design in marketing research for analyzing multi-dimensional choices and preferences.
Typical experiments can only identify treatment effects as a whole

Yet, scientific theories are often about effects of specific components

Conjoint analysis: A popular experimental design in marketing research for analyzing multi-dimensional choices and preferences

- Introduced in 1971 (Green and Rao 1971)
- Widely used in business (e.g. Courtyard Marriott)
- Similar to the “vignettes” and “factorial experiments” in sociology (Jasso and Rossi 1977, Wallander 2009)
Typical experiments can only identify treatment effects as a whole.

Yet, scientific theories are often about effects of specific components.

Conjoint analysis: A popular experimental design in marketing research for analyzing multi-dimensional choices and preferences.

- Introduced in 1971 (Green and Rao 1971)
- Widely used in business (e.g. Courtyard Marriott)
- Similar to the “vignettes” and “factorial experiments” in sociology (Jasso and Rossi 1977, Wallander 2009)

Respondents read profiles which vary across multiple attributes.
Respondents then choose, rank, or rate profiles.
Repeat the exercise multiple times.
## Example: What Respondents See

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.

<table>
<thead>
<tr>
<th></th>
<th>Immigrant 1</th>
<th>Immigrant 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prior Trips to the U.S.</strong></td>
<td>Entered the U.S. once before on a tourist visa</td>
<td>Entered the U.S. once before on a tourist visa</td>
</tr>
<tr>
<td><strong>Reason for Application</strong></td>
<td>Reunite with family members already in U.S.</td>
<td>Reunite with family members already in U.S.</td>
</tr>
<tr>
<td><strong>Country of Origin</strong></td>
<td>Mexico</td>
<td>Iraq</td>
</tr>
<tr>
<td><strong>Language Skills</strong></td>
<td>During admission interview, this applicant spoke fluent English</td>
<td>During admission interview, this applicant spoke fluent English</td>
</tr>
<tr>
<td><strong>Profession</strong></td>
<td>Child care provider</td>
<td>Teacher</td>
</tr>
<tr>
<td><strong>Job Experience</strong></td>
<td>One to two years of job training and experience</td>
<td>Three to five years of job training and experience</td>
</tr>
<tr>
<td><strong>Employment Plans</strong></td>
<td>Does not have a contract with a U.S. employer but has done job interviews</td>
<td>Will look for work after arriving in the U.S.</td>
</tr>
<tr>
<td><strong>Education Level</strong></td>
<td>Equivalent to completing two years of college in the U.S.</td>
<td>Equivalent to completing a college degree in the U.S.</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td>Female</td>
<td>Male</td>
</tr>
</tbody>
</table>
This talk is based on: Hainmueller, Jens, Daniel J. Hopkins and Teppei Yamamoto, “Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments”
This talk is based on: Hainmueller, Jens, Daniel J. Hopkins and Teppei Yamamoto, “Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments”

The paper makes the following contributions:

- Revisit conjoint analysis as a tool for causal inference
- Formally analyze causal properties of conjoint analysis
- Show that average marginal component effects (AMCE) and their interactions are nonparametrically identified
- Propose simple estimators of those quantities of interest
- Empirical applications to vote choice and immigrant admission
This talk is based on: Hainmueller, Jens, Daniel J. Hopkins and Teppei Yamamoto, “Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments”

The paper makes the following contributions:

- Revisit conjoint analysis as a tool for causal inference
- Formally analyze causal properties of conjoint analysis
- Show that average marginal component effects (AMCE) and their interactions are nonparametrically identified
- Propose simple estimators of those quantities of interest
- Empirical applications to vote choice and immigrant admission

Today, we skip formal derivations (mostly) and focus on implementation
Example 1: Candidate Experiment

- Candidates in elections differ on a wide variety of attributes
- 6 pairs of hypothetical presidential candidates
- Candidates differ in 8 attributes:
  - Religion
  - Education
  - Profession
  - Income
  - Racial/ethnic background
  - Age
  - Military service experience
  - Gender

- 2 types of outcome variables: Choice and rating
- 611 respondents on Mechanical Turk
Question 1 of 6

Please carefully review the two candidates for President detailed below. Then please answer the questions about these two candidates below.

<table>
<thead>
<tr>
<th></th>
<th>Candidate 1</th>
<th>Candidate 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Religion</td>
<td>Evangelical Protestant</td>
<td>Mainline Protestant</td>
</tr>
<tr>
<td>Profession</td>
<td>High School Teacher</td>
<td>Farmer</td>
</tr>
<tr>
<td>Age</td>
<td>75</td>
<td>68</td>
</tr>
<tr>
<td>Annual Income</td>
<td>$54,000</td>
<td>$210,000</td>
</tr>
<tr>
<td>Race / Ethnicity</td>
<td>Caucasian</td>
<td>Black</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>Male</td>
</tr>
<tr>
<td>Military Service</td>
<td>Served in U.S. military</td>
<td>No military service</td>
</tr>
<tr>
<td>College Education</td>
<td>BA from small college</td>
<td>BA from Baptist college</td>
</tr>
</tbody>
</table>

Which of these two candidates would you prefer to see as President of the United States?

<table>
<thead>
<tr>
<th>Candidate 1</th>
<th>Candidate 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

On a scale from 1 to 7, where 1 indicates that you would never support this candidate, and 7 indicates that you would always support this candidate, where would you place Candidate 1?

<table>
<thead>
<tr>
<th>Never Support</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Definitely Support</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Never Support</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Definitely Support</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

Why do you prefer this candidate? Please answer in one sentence.
Example 2: Immigration Experiment

- Existing evidence suggests a variety of immigrant attributes affecting policy preferences
- 5 pairs of hypothetical immigrants applying for admission to U.S.
- 9 attributes:
  - Gender
  - Education
  - Employment plans
  - Job experience
  - Profession
  - Language skills
  - Country of origin
  - Reason for applying
  - Prior trips to the U.S.
- Two restrictions on the distribution of the attributes:
  - Immigrants fleeing persecution must come from Iraq, Sudan, etc.
  - High-skill occupations must have at least two years of college education
- Knowledge Networks panel, 1407 completed respondents
Standard survey experiment: One binary treatment
Potential Outcomes Framework for Causal Inference

- Standard survey experiment: One binary treatment
- Respondent: $i \in \{1, \ldots, N\}$, drawn from population $\mathcal{P}$
Potential Outcomes Framework for Causal Inference

- Standard survey experiment: One binary treatment
- Respondent: \( i \in \{1, \ldots, N\} \), drawn from population \( P \)
- Treatment: \( T_i \in \{0, 1\} \)

Potential outcomes: \( Y_i(1) \) and \( Y_i(0) \)

- \( Y_i(1) \) indicates the response of respondent \( i \) we would observe if \( i \) was assigned to the treatment condition (\( T_i = 1 \))
- \( Y_i(0) \) indicates the response of respondent \( i \) we would observe if \( i \) was assigned to the control condition (\( T_i = 0 \))

In a given experiment, we only observe either \( Y_i(1) \) or \( Y_i(0) \)

Unit causal (treatment) effect: \( \tau_i \equiv Y_i(1) - Y_i(0) \)

Fundamental problem of causal inference (Holland 1986): We can never observe both potential outcomes at once!
Potential Outcomes Framework for Causal Inference

- Standard survey experiment: One binary treatment
- Respondent: $i \in \{1, \ldots, N\}$, drawn from population $\mathcal{P}$
- Treatment: $T_i \in \{0, 1\}$
- Potential outcomes: $Y_i(1)$ and $Y_i(0)$

$Y_i(1)$ indicates the response of respondent $i$ we would observe if $i$ was assigned to the treatment condition ($T_i = 1$).

$Y_i(0)$ indicates the response of respondent $i$ we would observe if $i$ was assigned to the control condition ($T_i = 0$).

In a given experiment, we only observe either $Y_i(1)$ or $Y_i(0)$.

Unit causal (treatment) effect:

$\tau_i \equiv Y_i(1) - Y_i(0)$

Fundamental problem of causal inference (Holland 1986): We can never observe both potential outcomes at once!
Potential Outcomes Framework for Causal Inference

- Standard survey experiment: One binary treatment
- Respondent: \( i \in \{1, \ldots, N\} \), drawn from population \( \mathcal{P} \)
- Treatment: \( T_i \in \{0, 1\} \)

**Potential outcomes:** \( Y_i(1) \) and \( Y_i(0) \)
- \( Y_i(1) \) indicates the response of respondent \( i \) we *would* observe if \( i \) was assigned to the treatment condition (\( T_i = 1 \))
- \( Y_i(0) \) indicates the response of respondent \( i \) we *would* observe if \( i \) was assigned to the control condition (\( T_i = 0 \))
Standard survey experiment: One binary treatment

Respondent: \( i \in \{1, \ldots, N\} \), drawn from population \( \mathcal{P} \)

Treatment: \( T_i \in \{0, 1\} \)

Potential outcomes: \( Y_i(1) \) and \( Y_i(0) \)
- \( Y_i(1) \) indicates the response of respondent \( i \) we would observe if \( i \) was assigned to the treatment condition (\( T_i = 1 \))
- \( Y_i(0) \) indicates the response of respondent \( i \) we would observe if \( i \) was assigned to the control condition (\( T_i = 0 \))
- In a given experiment, we only observe either \( Y_i(1) \) or \( Y_i(0) \)
Standard survey experiment: One binary treatment

Respondent: \( i \in \{1, \ldots, N\} \), drawn from population \( \mathcal{P} \)

Treatment: \( T_i \in \{0, 1\} \)

Potential outcomes: \( Y_i(1) \) and \( Y_i(0) \)

- \( Y_i(1) \) indicates the response of respondent \( i \) we would observe if \( i \) was assigned to the treatment condition \( (T_i = 1) \)
- \( Y_i(0) \) indicates the response of respondent \( i \) we would observe if \( i \) was assigned to the control condition \( (T_i = 0) \)
- In a given experiment, we only observe either \( Y_i(1) \) or \( Y_i(0) \)

Unit causal (treatment) effect: \( \tau_i \equiv Y_i(1) - Y_i(0) \)
Potential Outcomes Framework for Causal Inference

- Standard survey experiment: One binary treatment
- Respondent: \( i \in \{1, \ldots, N\} \), drawn from population \( P \)
- Treatment: \( T_i \in \{0, 1\} \)

- **Potential outcomes:** \( Y_i(1) \) and \( Y_i(0) \)
  - \( Y_i(1) \) indicates the response of respondent \( i \) we *would* observe if \( i \) was assigned to the treatment condition \( (T_i = 1) \)
  - \( Y_i(0) \) indicates the response of respondent \( i \) we *would* observe if \( i \) was assigned to the control condition \( (T_i = 0) \)
  - In a given experiment, we only observe either \( Y_i(1) \) or \( Y_i(0) \)

- **Unit causal (treatment) effect:** \( \tau_i \equiv Y_i(1) - Y_i(0) \)

- **Fundamental problem of causal inference** (Holland 1986): We can never observe both potential outcomes at once!
Suppose we have a population of four units.
Suppose we have a population of four units.

What we observe from data:

<table>
<thead>
<tr>
<th>Unit (i)</th>
<th>$T_i$</th>
<th>$Y_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Suppose we have a population of four units.

What’s really going on:

<table>
<thead>
<tr>
<th>Unit (i)</th>
<th>(T_i)</th>
<th>(Y_i)</th>
<th>(Y_i(1))</th>
<th>(Y_i(0))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>?</td>
<td>1</td>
</tr>
</tbody>
</table>
Suppose we have a population of four units.

Causal effects are not identified because of missing data:

<table>
<thead>
<tr>
<th>Unit (i)</th>
<th>$T_i$</th>
<th>$Y_i$</th>
<th>$Y_i(1)$</th>
<th>$Y_i(0)$</th>
<th>$\tau_i = Y_i(1) - Y_i(0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>?</td>
<td>1</td>
<td>?</td>
</tr>
</tbody>
</table>
Suppose we have a population of four units.

It could be this (null effect):

<table>
<thead>
<tr>
<th>Unit (i)</th>
<th>$T_i$</th>
<th>$Y_i$</th>
<th>$Y_i(1)$</th>
<th>$Y_i(0)$</th>
<th>$\tau_i = Y_i(1) - Y_i(0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Suppose we have a population of four units.

Or this (selection bias):

<table>
<thead>
<tr>
<th>Unit ($i$)</th>
<th>$T_i$</th>
<th>$Y_i$</th>
<th>$Y_i(1)$</th>
<th>$Y_i(0)$</th>
<th>$\tau_i = Y_i(1) - Y_i(0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>-1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>-3</td>
<td>0</td>
<td>-3</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>-4</td>
<td>1</td>
<td>-5</td>
</tr>
</tbody>
</table>
Causal Inference in a Standard Survey Experiment

- We cannot possibly identify $\tau_i$ because of the fundamental missing data problem
- But in a randomized experiment, we can identify the average treatment effect:

$$\mathbb{E}[\tau_i] = \mathbb{E}[Y_i(1) - Y_i(0)]$$
We cannot possibly identify $\tau_i$ because of the fundamental missing data problem.

But in a randomized experiment, we can identify the average treatment effect:

$$E[\tau_i] = E[Y_i(1) - Y_i(0)]$$

Because of randomization, we have

$$E[Y_i(t)] = E[Y_i|T_i = t].$$

That is, the ATE is identified as

$$E[Y_i|T_i = 1] - E[Y_i|T_i = 0].$$

<table>
<thead>
<tr>
<th>Unit ($i$)</th>
<th>$T_i$</th>
<th>$Y_i$</th>
<th>$Y_i(1)$</th>
<th>$Y_i(0)$</th>
<th>$\tau_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>?</td>
<td>1</td>
<td>?</td>
</tr>
</tbody>
</table>
Causal Inference in a Standard Survey Experiment

- We cannot possibly identify $\tau_i$ because of the fundamental missing data problem.
- But in a randomized experiment, we can identify the average treatment effect:
  \[ \mathbb{E}[\tau_i] = \mathbb{E}[Y_i(1) - Y_i(0)] \]

<table>
<thead>
<tr>
<th>Unit ($i$)</th>
<th>$T_i$</th>
<th>$Y_i$</th>
<th>$\mathbb{E}[Y_i(1)]$</th>
<th>$\mathbb{E}[Y_i(0)]$</th>
<th>$\mathbb{E}[\tau_i]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>.5</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>.5</td>
<td>1.5</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>.5</td>
<td>1.5</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

- Because of randomization, we have $\mathbb{E}[Y_i(t)] = \mathbb{E}[Y_i \mid T_i = t]$. 
Causal Inference in a Standard Survey Experiment

- We cannot possibly identify $\tau_i$ because of the fundamental missing data problem.
- But in a randomized experiment, we can identify the average treatment effect:

$$E[\tau_i] = E[Y_i(1) - Y_i(0)]$$

<table>
<thead>
<tr>
<th>Unit ($i$)</th>
<th>$T_i$</th>
<th>$Y_i$</th>
<th>$E[Y_i(1)]$</th>
<th>$E[Y_i(0)]$</th>
<th>$E[\tau_i]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>.5</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>.5</td>
<td>1.5</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>.5</td>
<td>1.5</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

- Because of randomization, we have $E[Y_i(t)] = E[Y_i \mid T_i = t]$.
- That is, the ATE is identified as $E[Y_i \mid T_i = 1] - E[Y_i \mid T_i = 0]$. 
Potential Outcomes Framework for Conjoint Analysis

- **Respondent**: $i \in \{1, \ldots, N\}$, drawn from population $P$
- **Choice (rating) task**: $k \in \{1, \ldots, K\}$
- **Profile**: $j \in \{1, \ldots, J\}$
- **Attribute**: $l \in \{1, \ldots, L\}$, each discrete with $D_l$ levels
Potential Outcomes Framework for Conjoint Analysis

- Respondent: $i \in \{1, ..., N\}$, drawn from population $\mathcal{P}$
- Choice (rating) task: $k \in \{1, ..., K\}$
- Profile: $j \in \{1, ..., J\}$
- Attribute: $l \in \{1, ..., L\}$, each discrete with $D_l$ levels

- Treatment: $T_{ijk} \equiv [T_{ijk1} \cdots T_{ijkL}]^\top \in \mathcal{T}$
  - Vector of attributes of profile $j$ that respondent $i$ sees in choice task $k$
  - $\mathcal{T}$: most simply, all possible combinations of attribute levels
  - Often, researcher wants to exclude certain profiles from $\mathcal{T}$
    (e.g. a research scientist with no formal education)
Potential Outcomes Framework for Conjoint Analysis

- **Respondent:** \( i \in \{1, \ldots, N\} \), drawn from population \( \mathcal{P} \)
- **Choice (rating) task:** \( k \in \{1, \ldots, K\} \)
- **Profile:** \( j \in \{1, \ldots, J\} \)
- **Attribute:** \( l \in \{1, \ldots, L\} \), each discrete with \( D_l \) levels

- **Treatment:** \( T_{ijk} \equiv [T_{ijk1} \cdots T_{ijkL}]^\top \in \mathcal{T} \)
  - Vector of attributes of profile \( j \) that respondent \( i \) sees in choice task \( k \)
  - \( \mathcal{T} \): most simply, all possible combinations of attribute levels
  - Often, researcher wants to exclude certain profiles from \( \mathcal{T} \) (e.g. a research scientist with no formal education)

- **\( T_{ik} \equiv [T_{i1k} \cdots T_{iJk}]^\top \):** \( J \times L \) matrix
  - e.g. 2 candidates w/ 3 binary attributes \((J = 2, L = 3, D_1 = D_2 = D_3 = 2)\):
    \[
    T_{ik} = \begin{bmatrix}
    \text{Profile of candidate 1} \\
    \text{Profile of candidate 2}
    \end{bmatrix} = \begin{bmatrix}
    \text{female} & \text{rich} & \text{college} \\
    \text{male} & \text{poor} & \text{college}
    \end{bmatrix}
    \]
Potential Outcomes and Assumptions

**Assumption 1: No carryover effect**

= profiles in other choice tasks do not affect how respondents evaluate the profiles in the current task
Potential Outcomes and Assumptions

**Assumption 1: No carryover effect**

Profiles in other choice tasks do not affect how respondents evaluate the profiles in the current task.

- Potential outcomes: $Y_i(t) \equiv [Y_{i1}(t) \cdots Y_{iJ}(t)]$
  - Respondent $i$’s response if she saw profiles $t$
- Choice outcome: $Y_{ij}(t) \in \{0, 1\}$, s.t. $\sum_{j=1}^{J} Y_{ij}(t) = 1$ for all $i$
- Rating outcome: $Y_{ij}(t) \in [y_l, y_u]$
**Assumption 1**: No carryover effect

≡ profiles in other choice tasks do not affect how respondents evaluate the profiles in the current task

- Potential outcomes: \( Y_i(t) \equiv [Y_{i1}(t) \cdots Y_{iJ}(t)] \)
  - Respondent \( i \)'s response if she saw profiles \( t \)
- Choice outcome: \( Y_{ij}(t) \in \{0, 1\} \), s.t. \( \sum_{j=1}^{J} Y_{ij}(t) = 1 \) for all \( i \)
- Rating outcome: \( Y_{ij}(t) \in [y_l, y_u] \)
- Observed outcome: \( Y_{ik} = Y_i(T_{ik}) = [Y_{i1}(T_{ik}) \cdots Y_{iJ}(T_{ik})]^\top \)
Potential Outcomes and Assumptions

Assumption 1: No carryover effect
= profiles in other choice tasks do not affect how respondents evaluate the profiles in the current task

- Potential outcomes: \( Y_i(t) \equiv [Y_{i1}(t) \cdots Y_{iJ}(t)] \)
  - Respondent \( i \)'s response if she saw profiles \( t \)
- Choice outcome: \( Y_{ij}(t) \in \{0, 1\} \), s.t. \( \sum_{j=1}^{J} Y_{ij}(t) = 1 \) for all \( i \)
- Rating outcome: \( Y_{ij}(t) \in [y_l, y_u] \)
- Observed outcome: \( Y_{ik} = Y_i(T_{ik}) = [Y_{i1}(T_{ik}) \cdots Y_{iJ}(T_{ik})]^\top \)

Assumption 2: No profile-order effect
= order of profiles within a choice task does not matter:

\[ Y_{ij}(T_{ik}) = Y_{ij'}(T'_{ik}) \quad \text{if} \quad T_{ijk} = T'_{ij'k} \quad \text{and} \quad T_{ij'k} = T'_{ijk}, \]

for any \( i, j, j' \) and \( k \).

- Potential outcomes are now: \( Y_i(t) = [Y_i(t_1, t_{[-1]}) \cdots Y_i(t_J, t_{[-J]})]^\top \)
**Potential Outcomes and Assumptions**

**Assumption 1: No carryover effect**

Profiles in other choice tasks do not affect how respondents evaluate the profiles in the current task.

- Potential outcomes: $Y_i(t) \equiv [Y_{i1}(t) \cdots Y_{iJ}(t)]$
  - Respondent $i$’s response if she saw profiles $t$
- Choice outcome: $Y_{ij}(t) \in \{0, 1\}$, s.t. $\sum_{j=1}^{J} Y_{ij}(t) = 1$ for all $i$
- Rating outcome: $Y_{ij}(t) \in [y_l, y_u]$
- Observed outcome: $Y_{ik} = Y_i(T_{ik}) = [Y_{i1}(T_{ik}) \cdots Y_{iJ}(T_{ik})]^T$

**Assumption 2: No profile-order effect**

Order of profiles within a choice task does not matter:

$$Y_{ij}(T_{ik}) = Y_{ij'}(T'_{ik}) \text{ if } T_{ijk} = T'_{ij'k} \text{ and } T_{ij'k} = T'_{ijk},$$

for any $i, j, j'$ and $k$.

- Potential outcomes are now: $Y_i(t) = [Y_i(t_1, t_{[-1]}) \cdots Y_i(t_J, t_{[-J]})]^T$

Assumptions 1 and 2 can be (partially) tested empirically.
**Assumption 3: Randomization of Profiles**

\[ Y_{i(t)} \perp T_{ijkl} \text{ for all } t, i, j, k \text{ and } l \]

\[ 0 < p(t) < 1 \text{ for all } t \]

- Attributes must be randomly generated
- All possible attribute combinations (in \( T \)) must be generated with positive probability
- Can be guaranteed to hold by properly implementing randomization
We want:
A causally-interpretable, summary measure of an attribute’s overall effect
We want:
A causally-interpretable, summary measure of an attribute’s overall effect

Idea: Average the effects over the distribution of the other attributes
We want:
A causally-interpretable, summary measure of an attribute’s overall effect

Idea: Average the effects over the distribution of the other attributes

For example, the effect for candidate gender is

\[
= \mathbb{E} \left[ Y_i \left( \frac{\text{female}}{\text{male}}, \frac{\text{rich}}{\text{rich}}, \frac{\text{college}}{\text{no college}} \right) - Y_i \left( \frac{\text{male}}{\text{male}}, \frac{\text{rich}}{\text{rich}}, \frac{\text{college}}{\text{no college}} \right) \right] \\
\times \Pr \left( \frac{\text{rich}}{\text{male}}, \frac{\text{rich}}{\text{rich}}, \frac{\text{college}}{\text{no college}} \right)
\]

\[
+ \mathbb{E} \left[ Y_i \left( \frac{\text{female}}{\text{female}}, \frac{\text{poor}}{\text{rich}}, \frac{\text{college}}{\text{college}} \right) - Y_i \left( \frac{\text{male}}{\text{male}}, \frac{\text{poor}}{\text{rich}}, \frac{\text{college}}{\text{college}} \right) \right] \\
\times \Pr \left( \frac{\text{poor}}{\text{male}}, \frac{\text{rich}}{\text{rich}}, \frac{\text{college}}{\text{college}} \right)
\]

\[
+ \mathbb{E} \left[ Y_i \left( \frac{\text{female}}{\text{female}}, \frac{\text{rich}}{\text{rich}}, \frac{\text{no college}}{\text{no college}} \right) - Y_i \left( \frac{\text{male}}{\text{male}}, \frac{\text{rich}}{\text{rich}}, \frac{\text{no college}}{\text{no college}} \right) \right] \\
\times \Pr \left( \frac{\text{rich}}{\text{female}}, \frac{\text{poor}}{\text{poor}}, \frac{\text{college}}{\text{college}} \right)
\]

\[
+ \cdots \cdots
\]
Quantities of Interest: Effect of Attributes

Formally: Average Marginal Component Effect (AMCE):

\[\bar{\pi}_l(t_1, t_0, p(t)) \equiv \mathbb{E}[Y_i(t_1, T_{ijk\l l}, T_{i\l j}k) - Y_i(t_0, T_{ijk\l l}, T_{i\l j}k)]\]

\[= \sum_{[t_{\l l}, t_{\l j}] \in \tilde{T}} \mathbb{E}[Y_i(t_1, t_{\l l}, t_{\l j}) - Y_i(t_0, t_{\l l}, t_{\l j})]p(t_{\l l}, t_{\l j}),\]

where

\[
\{\begin{array}{l}
T_{ijk\l l} = [T_{ijk1} \cdots T_{ijk(l-1)} T_{ijk(l+1)} \cdots T_{ijkL}] \quad \text{(all components but } l\text{th}) \\
T_{i\l j}k = [T_{i1k} \cdots T_{i(j-1)k} T_{i(j+1)k} \cdots T_{ijk}]^T \quad \text{(all profiles but } j\text{th})
\end{array}\]
Quantities of Interest: Effect of Attributes

Formally: Average Marginal Component Effect (AMCE):

\[ \bar{\pi}_l(t_1, t_0, p(t)) \equiv \mathbb{E}[Y_i(t_1, T_{ijk[-l]}, T_{i[-j]k}) - Y_i(t_0, T_{ijk[-l]}, T_{i[-j]k})] \]

\[ = \sum_{[t_{[-l]}, t_{[-j]}] \in \tilde{T}} \mathbb{E}[Y_i(t_1, t_{[-l]}, t_{[-j]}) - Y_i(t_0, t_{[-l]}, t_{[-j]})]p(t_{[-l]}, t_{[-j]}), \]

where

\[ \begin{align*}
T_{ijk(-l)} &= [T_{ijk1} \cdots T_{ijk(l-1)} T_{ijk(l+1)} \cdots T_{ijkL}] \quad \text{(all components but } l\text{th)} \\
T_{i(-j)k} &= [T_{i1k} \cdots T_{i(j-1)k} T_{i(j+1)k} \cdots T_{iJk}]^\top \quad \text{(all profiles but } j\text{th)}
\end{align*} \]

\[ \tilde{T} \] excludes “empty counterfactuals” (e.g. a research scientist with no formal education)

- Nonparametrically identified under Assumptions 1, 2 and 3
- Interaction effects can be similarly defined and identified
If profiles are not restricted, attributes can be completely randomized:

\[ T_{ijkl} \perp \{ T_{ijk[-l]}, T_{i[-j]k} \} \text{ for all } i, j, k. \]
If profiles are not restricted, attributes can be completely randomized:

\[ T_{ijkl} \perp \{ T_{ijk[-l]}, T_{i[-j]k} \} \text{ for all } i, j, k. \]

In this case, the simple difference-in-means estimator is unbiased:

\[
\hat{\pi}_l(t_1, t_0, p(t)) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{ijk} 1\{ T_{ijkl} = t_1 \}}{n_1} - \frac{\sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{ijk} 1\{ T_{ijkl} = t_0 \}}{n_0},
\]

where \( n_1 \) and \( n_0 \) are the numbers of profiles in the two conditions.
Nonparametric Estimation without Restrictions

- If profiles are not restricted, attributes can be completely randomized:

\[ T_{ijkl} \perp \perp \{ T_{ijk[-l]}, T_{i[-j]k} \} \text{ for all } i, j, k. \]

- In this case, the simple difference-in-means estimator is unbiased:

\[
\hat{\pi}_l(t_1, t_0, p(t)) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{ijk} 1\{ T_{ijkl} = t_1 \}}{n_1} - \frac{\sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{ijk} 1\{ T_{ijkl} = t_0 \}}{n_0},
\]

where \( n_1 \) and \( n_0 \) are the numbers of profiles in the two conditions.

- A numerically equivalent estimate can be obtained by a regression (regress \( Y_{ijk} \) on a set of dummies for the levels of \( T_{ijkl} \))

- Works for both choice and rating outcomes
Suppose we are interested in the AMCE of candidate age on ratings.

To obtain nonparametric estimates, we run the following linear regression:

\[
\text{rating}_{ijk} = \theta_0 + \theta_1 \text{[age}_{ijk} = 75] + \theta_2 \text{[age}_{ijk} = 68] + \theta_3 \text{[age}_{ijk} = 60] + \theta_4 \text{[age}_{ijk} = 52] + \theta_5 \text{[age}_{ijk} = 45] + \epsilon_{ijk},
\]

where \([\text{age}_{ijk} = XX]\) indicates a dummy variable for age \(XX\).

Then, \(\hat{\theta}_1, \hat{\theta}_2, \ldots\) are the estimated AMCE for ages 75, 68, etc. compared to the age of 36 (the omitted baseline category).

We can repeat this procedure for other attributes, such as gender, education, etc.

Alternatively, we can simply run one big regression for all attributes:

\[
\text{rating}_{ijk} = \theta_0 + \theta_1 \text{[age}_{ijk} = 75] + \theta_2 \text{[age}_{ijk} = 68] + \theta_3 \text{[age}_{ijk} = 60] + \theta_4 \text{[age}_{ijk} = 52] + \gamma_1 \text{[gender}_{ijk} = \text{female}] + \delta_1 \text{[race}_{ijk} = \text{black}] + \cdots + \epsilon_{ijk}.
\]
Suppose we are interested in the AMCE of candidate age on ratings.

To obtain nonparametric estimates, we run the following linear regression:

\[
\text{rating}_{ijk} = \theta_0 + \theta_1[\text{age}_{ijk} = 75] + \theta_2[\text{age}_{ijk} = 68] + \theta_3[\text{age}_{ijk} = 60] \\
+ \theta_4[\text{age}_{ijk} = 52] + \theta_5[\text{age}_{ijk} = 45] + \epsilon_{ijk},
\]

where \([\text{age}_{ijk} = XX]\) indicates a dummy variable for age XX.
Suppose we are interested in the AMCE of candidate age on ratings.

To obtain nonparametric estimates, we run the following linear regression:

\[
\text{rating}_{ijk} = \theta_0 + \theta_1[\text{age}_{ijk} = 75] + \theta_2[\text{age}_{ijk} = 68] + \theta_3[\text{age}_{ijk} = 60] \\
+ \theta_4[\text{age}_{ijk} = 52] + \theta_5[\text{age}_{ijk} = 45] + \varepsilon_{ijk},
\]

where \([\text{age}_{ijk} = XX]\) indicates a dummy variable for age XX.

Then, \(\hat{\theta}_1, \hat{\theta}_2, \text{ etc.}\), are the estimated AMCE for ages 75, 68, etc. compared to the age of 36 (the omitted baseline category).
Example: Effect of Candidate Age

Suppose we are interested in the AMCE of candidate age on ratings.

To obtain nonparametric estimates, we run the following linear regression:

\[ \text{rating}_{ijk} = \theta_0 + \theta_1[\text{age}_{ijk} = 75] + \theta_2[\text{age}_{ijk} = 68] + \theta_3[\text{age}_{ijk} = 60] + \theta_4[\text{age}_{ijk} = 52] + \theta_5[\text{age}_{ijk} = 45] + \epsilon_{ijk}, \]

where \([\text{age}_{ijk} = XX]\) indicates a dummy variable for age XX.

Then, \(\hat{\theta}_1, \hat{\theta}_2, \text{etc.},\) are the estimated AMCE for ages 75, 68, etc. compared to the age of 36 (the omitted baseline category).

We can repeat this procedure for other attributes, such as gender, education, etc.
Example: Effect of Candidate Age

Suppose we are interested in the AMCE of candidate age on ratings.

To obtain nonparametric estimates, we run the following linear regression:

\[
\text{rating}_{ijk} = \theta_0 + \theta_1[\text{age}_{ijk} = 75] + \theta_2[\text{age}_{ijk} = 68] + \theta_3[\text{age}_{ijk} = 60] \\
+ \theta_4[\text{age}_{ijk} = 52] + \theta_5[\text{age}_{ijk} = 45] + \varepsilon_{ijk},
\]

where \([\text{age}_{ijk} = XX]\) indicates a dummy variable for age XX.

Then, \(\hat{\theta}_1, \hat{\theta}_2, \) etc., are the estimated AMCE for ages 75, 68, etc. compared to the age of 36 (the omitted baseline category).

We can repeat this procedure for other attributes, such as gender, education, etc.

Alternatively, we can simply run one big regression for all attributes:

\[
\text{rating}_{ijk} = \theta_0 + \theta_1[\text{age}_{ijk} = 75] + \theta_2[\text{age}_{ijk} = 68] + \theta_3[\text{age}_{ijk} = 60] \\
+ \theta_4[\text{age}_{ijk} = 52] + \theta_5[\text{age}_{ijk} = 45] \\
+ \gamma_1[\text{gender}_{ijk} = \text{female}] + \delta_1[\text{race}_{ijk} = \text{black}] + \cdots + \varepsilon_{ijk}.
\]
Ex ante restrictions can be put on attribute combinations under the conditionally independent randomization:

\[ T_{ijkl} \perp \{ T^S_{ijk}, T_{i[-j]k} \} \mid T^R_{ijk} \text{ for all } i, j, k, \]

where

\[
\left\{
\begin{align*}
T^R_{ijk} &= \text{a subvector of } T_{ijk[-l]} \text{ involved in restriction} \\
T^S_{ijk} &= \text{the rest of } T_{ijk[-l]}
\end{align*}
\right.
\]
Ex ante restrictions can be put on attribute combinations under the conditionally independent randomization:

\[ T_{ijkl} \perp \perp \{ T_{ijk}^{S}, T_{i[-j]k} \} \mid T_{ijk}^{R} \text{ for all } i, j, k, \]

where

\[
\begin{align*}
T_{ijk}^{R} &= \text{a subvector of } T_{ijk[-l]} \text{ involved in restriction} \\
T_{ijk}^{S} &= \text{the rest of } T_{ijk[-l]}
\end{align*}
\]

In this case, the subclassification estimator is unbiased:

\[
\hat{\pi}_l(t_1, t_0, p(t)) = \sum_{t^{R} \in T^{R}} \left\{ \frac{\sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{ijk} 1\{T_{ijkl} = t_1, T_{ijk}^{R} = t^{R}\}}{n_{1t^{R}}} - \frac{\sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{ijk} 1\{T_{ijkl} = t_0, T_{ijk}^{R} = t^{R}\}}{n_{0t^{R}}} \right\} \frac{\Pr(T_{ijk}^{R} = t^{R})}{n_{tt^{R}}},
\]

where \( n_{tt^{R}} = \text{the (known) number of profiles for which } (T_{ijkl}, T_{ijk}^{R}) = (t, t^{R}) \).
Ex ante restrictions can be put on attribute combinations under the conditionally independent randomization:

\[ T_{ijkl} \perp \perp \{ T_{i\{j\}k}, T_{i[-j]k} \} \mid T_{ijk} \text{ for all } i, j, k, \]

where

\[
\begin{align*}
T^R_{ijk} &= \text{a subvector of } T_{ijk[-l]} \text{ involved in restriction} \\
T^S_{ijk} &= \text{the rest of } T_{ijk[-l]}
\end{align*}
\]

In this case, the subclassification estimator is unbiased:

\[
\hat{\pi}_l(t_1, t_0, p(t)) = \sum_{t^R \in T^R} \left\{ \frac{\sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{ijk} 1\{T_{ijkl} = t_1, T^R_{ijk} = t^R\}}{n_{1t^R}} - \frac{\sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{ijk} 1\{T_{ijkl} = t_0, T^R_{ijk} = t^R\}}{n_{0t^R}} \right\} \Pr(T^R_{ijk} = t^R),
\]

where \( n_{tt^R} = \text{the (known) number of profiles for which } (T_{ijkl}, T^R_{ijk}) = (t, t^R) \)

Can be computed by regression as a linear combination of coefficients
Ex ante restrictions can be put on attribute combinations under the conditionally independent randomization:

\[ T_{ijkl} \perp \perp \{ T_{ijk}^S, T_{i[-j]k} \} \mid T_{ijk}^R \text{ for all } i, j, k, \]

where

\[
\begin{cases}
T_{ijk}^R &= \text{a subvector of } T_{ijk[-l]} \text{ involved in restriction} \\
T_{ijk}^S &= \text{the rest of } T_{ijk[-l]}
\end{cases}
\]

In this case, the subclassification estimator is unbiased:

\[
\hat{\pi}_l(t_1, t_0, p(t)) = \sum_{t^R \in T^R} \left\{ \frac{\sum_{i=1}^N \sum_{j=1}^J \sum_{k=1}^K Y_{ijk} \mathbf{1}\{T_{ijkl} = t_1, T_{ijk}^R = t^R\} n_{1t^R}}{n_{0t^R}} - \frac{\sum_{i=1}^N \sum_{j=1}^J \sum_{k=1}^K Y_{ijk} \mathbf{1}\{T_{ijkl} = t_0, T_{ijk}^R = t^R\} n_{0t^R}}{n_{0t^R}} \right\} \Pr(T_{ijk}^R = t^R),
\]

where \( n_{tt^R} \) is the (known) number of profiles for which \((T_{ijkl}, T_{ijk}^R) = (t, t^R)\)

Can be computed by regression as a linear combination of coefficients

Interaction effects can be estimated using the same idea
Now, let’s estimate the AMCE of immigrant education on choice.

For simplicity, assume that we have two education levels (low and high) and three occupation categories (O1, O2, and O3), and that O3 is only allowed for immigrants with high education. Our regression is now:

\[
\text{chosen}_{ijk} = \theta_0 + \theta_1 \text{high}_{ijk} + \theta_2 \text{O2}_{ijk} + \theta_3 \text{O3}_{ijk} + \theta_4 (\text{high}_{ijk} \cdot \text{O2}_{ijk}) + \theta_5 (\text{high}_{ijk} \cdot \text{O3}_{ijk}) + \epsilon_{ijk}
\]

and the estimated AMCE for high vs. low education is

\[
0.5 \times \hat{\theta}_1 + 0.5 \times (\hat{\theta}_1 + \hat{\theta}_4) = \hat{\theta}_1 + 0.5 \times \hat{\theta}_4
\]

Note that the education effect for O3 (\(\hat{\theta}_1 + \hat{\theta}_5\)) is excluded from the estimate, because O3 cannot have low education.

Again, we can either repeat this procedure or run a big regression to obtain estimates for other attributes.
Example: Effect of Immigrant Education

Now, let’s estimate the AMCE of immigrant education on choice.

For simplicity, assume that we have two education levels (low and high) and three occupation categories (O1, O2 and O3), and that O3 is only allowed for immigrants with high education.

\[
\text{chosen}_{ijk} = \theta_0 + \theta_1 \text{high}_{ijk} + \theta_2 \text{O2}_{ijk} + \theta_3 \text{O3}_{ijk} + \theta_4 (\text{high}_{ijk} \cdot \text{O2}_{ijk}) + \theta_5 (\text{high}_{ijk} \cdot \text{O3}_{ijk}) + \epsilon_{ijk}
\]

and the estimated AMCE for high vs. low education is

\[
0.5 \times \hat{\theta}_1 + 0.5 \times (\hat{\theta}_1 + \hat{\theta}_4) = \hat{\theta}_1 + 0.5 \times \hat{\theta}_4
\]

Note that the education effect for O3 (\(\hat{\theta}_1 + \hat{\theta}_5\)) is excluded from the estimate, because O3 cannot have low education.

Again, we can either repeat this procedure or run a big regression to obtain estimates for other attributes.
Example: Effect of Immigrant Education

Now, let’s estimate the AMCE of immigrant education on choice.

For simplicity, assume that we have two education levels (low and high) and three occupation categories (O1, O2 and O3), and that O3 is only allowed for immigrants with high education.

Our regression is now:

\[
\text{chosen}_{ijk} = \theta_0 + \theta_1 \text{high}_{ijk} + \theta_2 \text{O2}_{ijk} + \theta_3 \text{O3}_{ijk} + \theta_4 (\text{high}_{ijk} \cdot \text{O2}_{ijk}) + \theta_5 (\text{high}_{ijk} \cdot \text{O3}_{ijk}) + \varepsilon_{ijk}
\]

and the estimated AMCE for high vs. low education is

\[
0.5 \times \hat{\theta}_1 + 0.5 \times (\hat{\theta}_1 + \hat{\theta}_4) = \hat{\theta}_1 + 0.5 \times \hat{\theta}_4
\]

Note that the education effect for O3 (\(\hat{\theta}_1 + \hat{\theta}_5\)) is excluded from the estimate, because O3 cannot have low education.
Now, let’s estimate the AMCE of immigrant education on choice.

For simplicity, assume that we have two education levels (low and high) and three occupation categories (O1, O2 and O3), and that O3 is only allowed for immigrants with high education.

Our regression is now:

\[
\text{chosen}_{ijk} = \theta_0 + \theta_1 \text{high}_{ijk} + \theta_2 \text{O2}_{ijk} + \theta_3 \text{O3}_{ijk} \\
+ \theta_4 (\text{high}_{ijk} \cdot \text{O2}_{ijk}) + \theta_5 (\text{high}_{ijk} \cdot \text{O3}_{ijk}) + \varepsilon_{ijk}
\]

and the estimated AMCE for high vs. low education is

\[
0.5 \times \hat{\theta}_1 + 0.5 \times (\hat{\theta}_1 + \hat{\theta}_4) = \hat{\theta}_1 + 0.5 \times \hat{\theta}_4
\]

Note that the education effect for O3 (\(\hat{\theta}_1 + \hat{\theta}_5\)) is excluded from the estimate, because O3 cannot have low education.

Again, we can either repeat this procedure or run a big regression to obtain estimates for other attributes.
Standard OLS variance estimates will be substantially biased because of the two sources of dependency:

1. Choice outcomes within choice tasks are strongly negatively correlated (e.g. \( \text{Corr}(Y_{i1k}, Y_{i2k}) = -1 \) when \( J = 2 \))

2. Potential outcomes are positively correlated within respondents
Variance Estimation

- Standard OLS variance estimates will be substantially biased because of the two sources of dependency:
  1. Choice outcomes within choice tasks are strongly negatively correlated (e.g. $\text{Corr}(Y_{i1k}, Y_{i2k}) = -1$ when $J = 2$)
  2. Potential outcomes are positively correlated within respondents

- To address these, we can use
  1. Correction for within-respondent clustering
  2. Block bootstrap
Standard OLS variance estimates will be substantially biased because of the two sources of dependency:

1. Choice outcomes within choice tasks are strongly negatively correlated (e.g. $\text{Corr}(Y_{i1k}, Y_{i2k}) = -1$ when $J = 2$)

2. Potential outcomes are positively correlated within respondents

To address these, we can use

1. Correction for within-respondent clustering
2. Block bootstrap

Both are consistent as $N \to \infty$ (but not $J$ or $K$)
Variance Estimation

- Standard OLS variance estimates will be substantially biased because of the two sources of dependency:
  1. Choice outcomes within choice tasks are strongly negatively correlated (e.g. \( \text{Corr}(Y_{i1k}, Y_{i2k}) = -1 \) when \( J = 2 \))
  2. Potential outcomes are positively correlated within respondents

- To address these, we can use
  1. Correction for within-respondent clustering
  2. Block bootstrap

  Both are consistent as \( N \rightarrow \infty \) (but not \( J \) or \( K \))

- \( N \) is typically large in a survey experiment, so using a cluster-robust standard error will be appropriate
Immigrant Experiment: AMCE

Change in Pr(Immigrant Preferred for Admission to U.S.)

Gender:
- female
- male

Education:
- no formal
- 4th grade
- 8th grade
- high school
- two-year college
- college degree
- graduate degree

Language:
- fluent English
- broken English
- tried English but unable to use
- used interpreter

Origin:
- Germany
- France
- Mexico
- Philippines
- Poland
- India
- China
- Sudan
- Somalia
- Iraq

Profession:
- janitor
- waiter
- child care provider
- gardener
- financial analyst
- construction worker
- teacher
- computer programmer
- nurse
- research scientist
- doctor

Job experience:
- none
- 1–2 years
- 3–5 years
- 5+ years

Job plans:
- contract with employer
- interviews with employer
- will look for work
- no plans to look for work

Application reason:
- reunite with family
- seek better job
- escape persecution

Prior trips to U.S.:
- never
- once as tourist
- many times as tourist
- six months with family
- once w/o authorization

Teppie Yamamoto (MIT)  Conjoint Analysis  April 5, 2013  27 / 30
Interaction between Immigrant Attributes

**Immigrant Has Contract with an Employer**

- Gender: female, male
- Education: no formal, 4th grade, 8th grade, high school, two-year college, college degree, graduate degree
- Language: fluent English, broken English, tried English but unable, used interpreter
- Origin: Germany, France, Mexico, Philippines, Poland, India, China, Sudan, Somalia, Iraq
- Profession: janitor, waiter, child care provider, gardener, financial analyst, construction worker, teacher, computer programmer, nurse, research scientist, doctor
- Job experience: none, 1-2 years, 3-5 years, 5+ years
- Application reason: reunite with family, seek better job, escape persecution
- Prior trips to U.S.: never, once as tourist, many times as tourist, six months with family, once w/o authorization

**Immigrant Has No Plans to Look for Work**

- Change in Pr(Immigrant Preferred for Admission to U.S.)

---

Teppei Yamamoto (MIT)
Conjoint Analysis
April 5, 2013 29 / 30
Concluding Remarks

- Conjoint analysis as a tool to identify components of treatment effect
- Compare multiple causal effects on the same scale in one study
- Evaluation of multiple causal theories
- Practical advantages: Enhanced realism, suppression of social desirability bias
- Proposed estimators do not rely on functional-form assumptions, especially no-interaction assumptions
- Wide range of applicability
<table>
<thead>
<tr>
<th>Country</th>
<th>Number</th>
<th>% of Immigrants</th>
<th>% with Some Coll.</th>
<th>% with BA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexico</td>
<td>26,693</td>
<td>0.243</td>
<td>0.170</td>
<td>0.061</td>
</tr>
<tr>
<td>Somalia</td>
<td>450</td>
<td>0.004</td>
<td>0.262</td>
<td>0.076</td>
</tr>
<tr>
<td>Iraq</td>
<td>426</td>
<td>0.004</td>
<td>0.498</td>
<td>0.270</td>
</tr>
<tr>
<td>Sudan</td>
<td>216</td>
<td>0.002</td>
<td>0.532</td>
<td>0.278</td>
</tr>
<tr>
<td>China</td>
<td>3,875</td>
<td>0.035</td>
<td>0.558</td>
<td>0.427</td>
</tr>
<tr>
<td>Poland</td>
<td>1,077</td>
<td>0.010</td>
<td>0.564</td>
<td>0.341</td>
</tr>
<tr>
<td>Germany</td>
<td>3,015</td>
<td>0.027</td>
<td>0.667</td>
<td>0.369</td>
</tr>
<tr>
<td>Philippines</td>
<td>5,577</td>
<td>0.051</td>
<td>0.709</td>
<td>0.443</td>
</tr>
<tr>
<td>France</td>
<td>531</td>
<td>0.005</td>
<td>0.727</td>
<td>0.463</td>
</tr>
<tr>
<td>India</td>
<td>4,806</td>
<td>0.044</td>
<td>0.840</td>
<td>0.760</td>
</tr>
</tbody>
</table>

**Table:** This table reports estimates obtained from the Current Population Surveys from September 2011 through March 2012. In total, these surveys had 1,060,286 respondents, 109,763 of whom were immigrants who provided their levels of education.
### Table: Results of eight-cluster implementation of Latent Dirichlet Allocation

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>immigrant</td>
<td>illegally</td>
<td>family</td>
<td>persecution</td>
<td>experience</td>
<td>look</td>
<td>contract</td>
<td>english</td>
</tr>
<tr>
<td>2</td>
<td>society</td>
<td>enter</td>
<td>reunite</td>
<td>escape</td>
<td>education</td>
<td>plans</td>
<td>employer</td>
<td>speaks</td>
</tr>
<tr>
<td>3</td>
<td>chose</td>
<td>country</td>
<td>education</td>
<td>escaping</td>
<td>job</td>
<td>educated</td>
<td>degree</td>
<td>fluent</td>
</tr>
<tr>
<td>4</td>
<td>able</td>
<td>entered</td>
<td>looking</td>
<td>seeking</td>
<td>training</td>
<td>experience</td>
<td>college</td>
<td>speak</td>
</tr>
<tr>
<td>5</td>
<td>contribute</td>
<td>tried</td>
<td>person</td>
<td>experience</td>
<td>lined</td>
<td>time</td>
<td>graduate</td>
<td>broken</td>
</tr>
<tr>
<td>6</td>
<td>profession</td>
<td>educated</td>
<td>united</td>
<td>society</td>
<td>level</td>
<td>job</td>
<td>immigrant</td>
<td>spoke</td>
</tr>
<tr>
<td>7</td>
<td>educational</td>
<td>authorization</td>
<td>support</td>
<td>trying</td>
<td>formal</td>
<td>speaking</td>
<td>applicant</td>
<td>teacher</td>
</tr>
<tr>
<td>8</td>
<td>skills</td>
<td>reason</td>
<td>shes</td>
<td>person</td>
<td>schooling</td>
<td>field</td>
<td>equivalent</td>
<td>fluently</td>
</tr>
<tr>
<td>9</td>
<td>people</td>
<td>legal</td>
<td>system</td>
<td>political</td>
<td>teacher</td>
<td>legally</td>
<td>lined</td>
<td>application</td>
</tr>
<tr>
<td>10</td>
<td>chance</td>
<td>doctor</td>
<td>research</td>
<td>religious</td>
<td>useful</td>
<td>planning</td>
<td>job</td>
<td>care</td>
</tr>
<tr>
<td>11</td>
<td>language</td>
<td>law</td>
<td>probably</td>
<td>political</td>
<td>looking</td>
<td>qualified</td>
<td>doesnt</td>
<td>child</td>
</tr>
<tr>
<td>12</td>
<td>seek</td>
<td>didnt</td>
<td>desire</td>
<td>help</td>
<td>society</td>
<td>choice</td>
<td>experience</td>
<td>makes</td>
</tr>
<tr>
<td>13</td>
<td>education</td>
<td>immigrant</td>
<td>reunited</td>
<td>education</td>
<td>slightly</td>
<td>nurses</td>
<td>live</td>
<td>able</td>
</tr>
<tr>
<td>14</td>
<td>background</td>
<td>breaking</td>
<td>urgent</td>
<td>profession</td>
<td>hes</td>
<td>seek</td>
<td>looking</td>
<td>skill</td>
</tr>
<tr>
<td>15</td>
<td>employment</td>
<td>previously</td>
<td>asylum</td>
<td>religious</td>
<td>valuable</td>
<td>easier</td>
<td>illegal</td>
<td>reuniting</td>
</tr>
<tr>
<td>16</td>
<td>america</td>
<td>teacher</td>
<td>looks</td>
<td>lined</td>
<td>willing</td>
<td>shes</td>
<td>illegal</td>
<td>community</td>
</tr>
<tr>
<td>17</td>
<td>level</td>
<td>hasnt</td>
<td>demand</td>
<td>priority</td>
<td>programmer</td>
<td>finding</td>
<td>employment</td>
<td>field</td>
</tr>
<tr>
<td>18</td>
<td>worker</td>
<td>valid</td>
<td>simply</td>
<td>nurses</td>
<td>highly</td>
<td>highly</td>
<td>learn</td>
<td>set</td>
</tr>
<tr>
<td>19</td>
<td>skilled</td>
<td>past</td>
<td>somalia</td>
<td>people</td>
<td>professional</td>
<td>applicant</td>
<td>nurse</td>
<td>little</td>
</tr>
<tr>
<td>20</td>
<td>doctor</td>
<td>rules</td>
<td>smarter</td>
<td>skilled</td>
<td>looks</td>
<td>jobs</td>
<td>family</td>
<td>language</td>
</tr>
</tbody>
</table>
Example: Immigration Conjoint Analysis

- Gender: female, male
- Education: no formal, 4th grade, 8th grade, high school, two-year college, college degree, graduate degree
- Language: fluent English, broken English, unable to use interpreter
- Origin: Germany, France, Mexico, Philippines, Poland, India, China, Sudan, Somalia, Iraq
- Profession: janitor, waiter, child care provider, gardener, financial analyst, construction worker, teacher, computer programmer, nurse, research scientist, doctor
- Job experience: none, 1-2 years, 3-5 years, 5+ years
- Job plans: contract with employer, interviews with employer, will look for work, no plans to look for work
- Application reason: reunite with family, seek better job, escape persecution
- Prior trips to U.S.: never, once as tourist, six months with family, once w/o authorization

Change in Pr(Immigrant Preferred for Admission to U.S.)
- once w/o authorization
- six months with family
- many times as tourist
- once as tourist
- never

Prior trips to U.S.: never

Pairing No: 1
Pairing No: 2
Pairing No: 3
Pairing No: 4
Pairing No: 5
Choice Probabilities

% of Voters Supporting Admission of Immigrant to U.S.

- Gender: Female
  - Education: graduate degree
  - Language: fluent English
  - Origin: Germany
  - Profession: research scientist
  - Job experience: 3–5 years
  - Job plans: contract with employer
  - Application reason: seek better job
  - Prior trips to U.S.: six months with family
  - Percentile: 99

- Gender: Female
  - Education: college degree
  - Language: fluent English
  - Origin: Mexico
  - Profession: teacher
  - Job experience: 1–2 years
  - Job plans: interviews with employer
  - Application reason: seek better job
  - Prior trips to U.S.: many times as tourist
  - Percentile: 75

- Gender: Female
  - Education: high school
  - Language: broken English
  - Origin: India
  - Profession: child care provider
  - Job experience: none
  - Job plans: interviews with employer
  - Application reason: seek better job
  - Prior trips to U.S.: once as tourist
  - Percentile: 50

- Gender: Female
  - Education: 8th grade
  - Language: tried English but unable
  - Origin: Sudan
  - Profession: construction worker
  - Job experience: none
  - Job plans: interviews with employer
  - Application reason: seek better job
  - Prior trips to U.S.: never
  - Percentile: 25

- Gender: Female
  - Education: 4th grade
  - Language: used interpreter
  - Origin: Iraq
  - Profession: janitor
  - Job experience: none
  - Job plans: no plans to look for work
  - Application reason: seek better job
  - Prior trips to U.S.: once w/o authorization
  - Percentile: 1

Teppei Yamamoto (MIT)