Making Causal Critiques

Day 4 - How much are we Learning?

Jonathan Phillips

January 31, 2019

How much are we Learning?

- Everything we have discussed so far has been about the **accuracy** of a causal claim

How much are we Learning?

- Everything we have discussed so far has been about the **accuracy** of a causal claim
- But not every study is as valuable to political science

How much are we Learning?

- Everything we have discussed so far has been about the **accuracy** of a causal claim
- But not every study is as valuable to political science
- We *learn* more from some studies than from others

How much are we Learning?

- Everything we have discussed so far has been about the **accuracy** of a causal claim
- But not every study is as valuable to political science
- We *learn* more from some studies than from others
    1. Reliability of the claim

How much are we Learning?

- Everything we have discussed so far has been about the **accuracy** of a causal claim
- But not every study is as valuable to political science
- We *learn* more from some studies than from others
    1. Reliability of the claim
    2. Reproducibility of the claim

How much are we Learning?

- Everything we have discussed so far has been about the **accuracy** of a causal claim
- But not every study is as valuable to political science
- We *learn* more from some studies than from others
    1. Reliability of the claim
    2. Reproducibility of the claim
    3. Scope (*generalizability*) of the claim

Robustness

- For simplicity, we publish a paper with a ’final’ result

Robustness

- For simplicity, we publish a paper with a 'final' result
  - 1% extra GDP growth increases the President's chance of re-election by 5%

Robustness

- For simplicity, we publish a paper with a ’final’ result
  - 1% extra GDP growth increases the President’s chance of re-election by 5%
- But how **confident** are we in these figures?

Robustness

- For simplicity, we publish a paper with a ’final’ result
  - 1% extra GDP growth increases the President’s chance of re-election by 5%
- But how **confident** are we in these figures?
- Good studies include estimates of uncertainty

Robustness

- For simplicity, we publish a paper with a ’final’ result
  - 1% extra GDP growth increases the President’s chance of re-election by 5%
- But how **confident** are we in these figures?
- Good studies include estimates of uncertainty
  - 1% extra GDP growth increases the President’s chance of re-election by 5% with a standard deviation of 0.2%

Robustness

- For simplicity, we publish a paper with a ’final’ result
  - 1% extra GDP growth increases the President’s chance of re-election by 5%
- But how **confident** are we in these figures?
- Good studies include estimates of uncertainty
  - 1% extra GDP growth increases the President’s chance of re-election by 5% with a standard deviation of 0.2%
- But these confidence intervals are usually for a *single* methodology and a fixed set of assumptions

Robustness

- What if our assumptions were wrong?
  pause
- How much would our results change if we used a different methodology?

Robustness

- What if our assumptions were wrong?
  pause
- How much would our results change if we used a different methodology?
  - Including different controls

Robustness

- What if our assumptions were wrong?
  pause
- How much would our results change if we used a different methodology?
  - Including different controls
  - Including alternative measures of the variables

Robustness

- What if our assumptions were wrong?
  pause
- How much would our results change if we used a different methodology?
  - Including different controls
  - Including alternative measures of the variables
  - Including or excluding outliers

Robustness

- What if our assumptions were wrong?
  pause
- How much would our results change if we used a different methodology?
  - Including different controls
  - Including alternative measures of the variables
  - Including or excluding outliers
  - Including a different functional form for the regression

Robustness

- What if our assumptions were wrong?
  pause
- How much would our results change if we used a different methodology?
  - Including different controls
  - Including alternative measures of the variables
  - Including or excluding outliers
  - Including a different functional form for the regression
- If we can change all these things and still get the same answers, our result is **reliable** and **robust**

Robustness

- For example, Michalpoulos and Papaioannou (2013) show that more centralized pre-colonial societies in Africa have more economic activity today

Robustness

- For example, Michalpoulos and Papaioannou (2013) show that more centralized pre-colonial societies in Africa have more economic activity today
- Robustness tests include:
  - Extra controls for disease, land, natural resources
  - Alternative model for spatial autocorrelation
  - Country fixed effects to focus only on within-country variation
  - Comparing only neighbouring societies
  - Alternative codings of centralized pre-colonial societies
  - Alternative measures of economic activity (nightlights etc.)
  - Different units of analysis - grid squares instead of ethnic territories

Robustness

- Robustness tests help avoid **researcher bias**

Robustness

- Robustness tests help avoid **researcher bias**
  - Running 200 models with different covariates

Robustness

- Robustness tests help avoid **researcher bias**
  - Running 200 models with different covariates
  - Only reporting one that is significant

Robustness

- Robustness tests help avoid **researcher bias**
  - Running 200 models with different covariates
  - Only reporting one that is significant
  - But even if there was **no causal effect** in the data, *by chance* we would expect 10 models to produce significant effects

Reproducibility

1. If we take the same data and apply the same method, do we get the same result?

Reproducibility

1. If we take the same data and apply the same method, do we get the same result?
   - Often, no! Only 35% replication rate in Brazilian political science journals (Avelino and Desposato 2018)

Reproducibility

1. If we take the same data and apply the same method, do we get the same result?
    - Often, no! Only 35% replication rate in Brazilian political science journals (Avelino and Desposato 2018)
    - And that's for the papers where we have access to the data and code

Reproducibility

1. If we take the same data and apply the same method, do we get the same result?
   - Often, no! Only 35% replication rate in Brazilian political science journals (Avelino and Desposato 2018)
   - And that’s for the papers where we have access to the data and code
2. If we take **another** sample of data and apply the same method, do we get the same result?
   - Very rarely done

Reproducibility

- A big problem for reproducibility is **publication bias**

Reproducibility

- A big problem for reproducibility is **publication bias**
  - Lots of researchers perform lots of studies

Reproducibility

- A big problem for reproducibility is **publication bias**
  - Lots of researchers perform lots of studies
  - Some find positive results, some negative, many ’null’ findings

Reproducibility

- A big problem for reproducibility is **publication bias**
  - Lots of researchers perform lots of studies
  - Some find positive results, some negative, many ’null’ findings
  - But journals want readers, and readers like positive results

Reproducibility

- A big problem for reproducibility is **publication bias**
  - Lots of researchers perform lots of studies
  - Some find positive results, some negative, many ’null’ findings
  - But journals want readers, and readers like positive results
  - So only the positive results get published

Reproducibility

- A big problem for reproducibility is **publication bias**
  - Lots of researchers perform lots of studies
  - Some find positive results, some negative, many ’null’ findings
  - But journals want readers, and readers like positive results
  - So only the positive results get published
- If you’re reading a paper, think of the ten other papers you’re *not* reading that tried the same thing and found no effect

Reproducibility

- Publication bias is a **huge** problem

Reproducibility

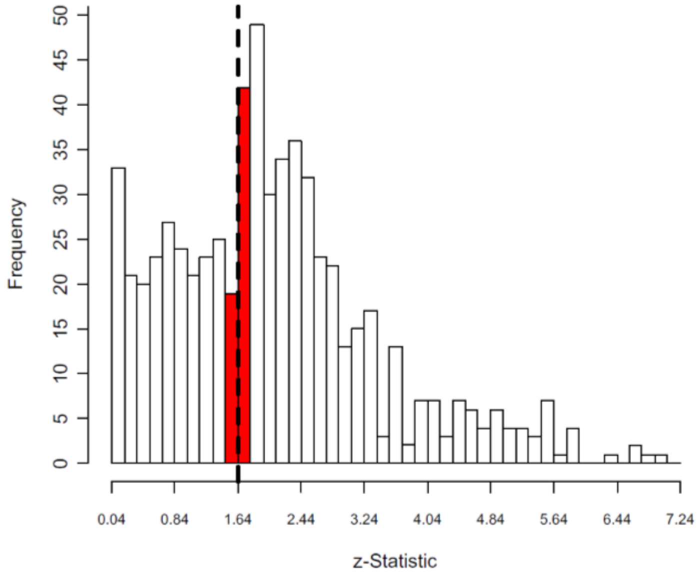
- Publication bias is a **huge** problem
- Compare the frquency of results in APSR and AJPS just above and below the 1.96 test statistic (for 5% significance)

Reproducibility

- Publication bias is a **huge** problem
- Compare the frquency of results in APSR and AJPS just above and below the 1.96 test statistic (for 5% significance)
- Many more values just below the threshold

Reproducibility

- Publication bias is a **huge** problem
- Compare the frquency of results in APSR and AJPS just above and below the 1.96 test statistic (for 5% significance)
- Many more values just below the threshold
- Less than 1 in 32 billion chance this happened by chance!


Frequency
0
5
10
15
20
25
30
35
40
45
50
0.04
0.84
1.64
2.44
3.24
4.04
4.84
5.64
6.44
7.24
z-Statistic

Reproducibility

- One solution is **Pre-registration**

Reproducibility

- One solution is **Pre-registration**
  - Submit your study design to a website - what you will analyse and how

Reproducibility

- One solution is **Pre-registration**
  - Submit your study design to a website - what you will analyse and how
  - Everyone knows who is researching what, and if they published or not

Reproducibility

- One solution is **Pre-registration**
  - Submit your study design to a website - what you will analyse and how
  - Everyone knows who is researching what, and if they published or not
  - Researchers are also less tempted to 'pick' their preferred analysis after seeing the data

Reproducibility

- One solution is **Pre-registration**
  - Submit your study design to a website - what you will analyse and how
  - Everyone knows who is researching what, and if they published or not
  - Researchers are also less tempted to ’pick’ their preferred analysis after seeing the data
  - Eg. EGAP Pre-Registration

Generalizability

- But even if studies are robust and reproducible, **how much** are we learning?

Generalizability

- But even if studies are robust and reproducible, **how much** are we learning?
- We can learn very little even from a precise, bias-free study:

Generalizability

- But even if studies are robust and reproducible, **how much** are we learning?
- We can learn very little even from a precise, bias-free study:
  - IgNobel Prize
  - "Suicide rates are linked to the amount of country music played on the radio"
  - "Is using voodoo dolls effective?"
  - "Why do old men have big ears?"
  - "How exposure to a crocodile encourages people to gamble"

Generalizability

- **Internal Validity**
  - Are the conclusions of the study accurate *within* the sample?

Generalizability

- **Internal Validity**
  - Are the conclusions of the study accurate *within* the sample?
  - Are the assumptions valid, is our causal effect biased?

Generalizability

- **Internal Validity**
  - Are the conclusions of the study accurate *within* the sample?
  - Are the assumptions valid, is our causal effect biased?
  - Is the conclusion reliable if we use slightly different assumptions?

Generalizability

- **Internal Validity**
  - Are the conclusions of the study accurate *within* the sample?
  - Are the assumptions valid, is our causal effect biased?
  - Is the conclusion reliable if we use slightly different assumptions?
- **External Validity**
  - How far can the results 'travel' outside of the study sample?

Generalizability

- **Internal Validity**
  - Are the conclusions of the study accurate *within* the sample?
  - Are the assumptions valid, is our causal effect biased?
  - Is the conclusion reliable if we use slightly different assumptions?
- **External Validity**
  - How far can the results ’travel’ outside of the study sample?
    1. Does the study reflect a wider population?

Generalizability

- **Internal Validity**
  - Are the conclusions of the study accurate *within* the sample?
  - Are the assumptions valid, is our causal effect biased?
  - Is the conclusion reliable if we use slightly different assumptions?
- **External Validity**
  - How far can the results 'travel' outside of the study sample?
    1. Does the study reflect a wider population?
    2. How big, representative and interesting is that wider population?

Generalizability

- For example, Chattopadhyay and Duflo (2004) argue that women leaders invest more in education using data from an experiment in 265 villages in two states in India (West Bengal and Rajasthan)

Generalizability

- For example, Chattopadhyay and Duflo (2004) argue that women leaders invest more in education using data from an experiment in 265 villages in two states in India (West Bengal and Rajasthan)
- But does the conclusion apply to:

Generalizability

- For example, Chattopadhyay and Duflo (2004) argue that women leaders invest more in education using data from an experiment in 265 villages in two states in India (West Bengal and Rajasthan)
- But does the conclusion apply to:
    1. 265 different villages?

Generalizability

- For example, Chattopadhyay and Duflo (2004) argue that women leaders invest more in education using data from an experiment in 265 villages in two states in India (West Bengal and Rajasthan)
- But does the conclusion apply to:
  1. 265 different villages?
  2. Different states?

Generalizability

- For example, Chattopadhyay and Duflo (2004) argue that women leaders invest more in education using data from an experiment in 265 villages in two states in India (West Bengal and Rajasthan)
- But does the conclusion apply to:
  1. 265 different villages?
  2. Different states?
  3. Different countries?

Generalizability

- For example, Chattopadhyay and Duflo (2004) argue that women leaders invest more in education using data from an experiment in 265 villages in two states in India (West Bengal and Rajasthan)
- But does the conclusion apply to:
  1. 265 different villages?
  2. Different states?
  3. Different countries?
  4. Different years?

Generalizability

- Most studies are designed with generalizability in mind:

Generalizability

- Most studies are designed with generalizability in mind:
    - Representative Samples are drawn from a target population

Generalizability

- Most studies are designed with generalizability in mind:
  - Representative Samples are drawn from a target population
  - We use **statistical inference** to extend our conclusions from the sample to the population

Generalizability

- Most studies are designed with generalizability in mind:
  - Representative Samples are drawn from a target population
  - We use **statistical inference** to extend our conclusions from the sample to the population
    - Note this only works if we know all the units (hidden tribes etc.)

Generalizability

- Most studies are designed with generalizability in mind:
  - Representative Samples are drawn from a target population
  - We use **statistical inference** to extend our conclusions from the sample to the population
    - Note this only works if we know all the units (hidden tribes etc.)
  - But Chattopadhyay and Duflo (2004) was not a representative sample of villages

Generalizability

- Most studies are designed with generalizability in mind:
  - Representative Samples are drawn from a target population
  - We use **statistical inference** to extend our conclusions from the sample to the population
    - Note this only works if we know all the units (hidden tribes etc.)
  - But Chattopadhyay and Duflo (2004) was not a representative sample of villages
  - Their widely-cited paper *only* applies to Birbhum and Udaipur districts

Generalizability

- Most studies are designed with generalizability in mind:
  - Representative Samples are drawn from a target population
  - We use **statistical inference** to extend our conclusions from the sample to the population
    - Note this only works if we know all the units (hidden tribes etc.)
  - But Chattopadhyay and Duflo (2004) was not a representative sample of villages
  - Their widely-cited paper *only* applies to Birbhum and Udaipur districts
  - We have no evidence of how women leaders govern elsewhere in India or the world

Generalizability

- Specific causal research designs also restrict the scope of our findings

Generalizability

- Specific causal research designs also restrict the scope of our findings
  - Precisely because we had to restrict our sample to find appropriate counterfactuals

Generalizability

- Specific causal research designs also restrict the scope of our findings
  - Precisely because we had to restrict our sample to find appropriate counterfactuals
  - The new comparisons are often less representative or interesting

Generalizability

- Specific causal research designs also restrict the scope of our findings
  - Precisely because we had to restrict our sample to find appropriate counterfactuals
  - The new comparisons are often less representative or interesting
- Instead of an **Average Treatment Effect (ATE)** they represent a **Local Average Treatment Effect (LATE)**

Generalizability

- Specific causal research designs also restrict the scope of our findings
  - Precisely because we had to restrict our sample to find appropriate counterfactuals
  - The new comparisons are often less representative or interesting
- Instead of an **Average Treatment Effect (ATE)** they represent a **Local Average Treatment Effect (LATE)**
  - A treatment effect applicable only to those units who were affected by the ’random’ part of treatment: **compliers**

Field Experiments

- Implementation is limited to a small sample, often non-representative

Field Experiments

- Implementation is limited to a small sample, often non-representative
  - Due to costs, consent

Field Experiments

- Implementation is limited to a small sample, often non-representative
    - Due to costs, consent
- And the findings *only* apply to that sample

Field Experiments

- Implementation is limited to a small sample, often non-representative
  - Due to costs, consent
- And the findings *only* apply to that sample
- Or maybe only to a sub-group of that sample

Field Experiments

- **External Validity** in Field Experiments:

Field Experiments

- **External Validity** in Field Experiments:
  - What **theory** are we testing? We can't accumulate knowledge without theory. The causal mechanisms are still a black box.

Field Experiments

- **External Validity** in Field Experiments:
  - What **theory** are we testing? We can't accumulate knowledge without theory. The causal mechanisms are still a black box.
  - Limited **portability** of findings - context matters for the treatment effects:

Field Experiments

- **External Validity** in Field Experiments:
  - What **theory** are we testing? We can't accumulate knowledge without theory. The causal mechanisms are still a black box.
  - Limited **portability** of findings - context matters for the treatment effects:
    - Eg. CCTs improve child health *only* where clinics are available, people are sufficiently educated, etc.

Field Experiments

- **External Validity** in Field Experiments:
  - What **theory** are we testing? We can't accumulate knowledge without theory. The causal mechanisms are still a black box.
  - Limited **portability** of findings - context matters for the treatment effects:
    - Eg. CCTs improve child health *only* where clinics are available, people are sufficiently educated, etc.
  - How much do the results depend on researcher oversight?

Lab Experiments

- Problems generalizing from the lab:

Lab Experiments

- Problems generalizing from the lab:
  - **Hawthorne effect**: Lab context influences behaviour, social desirability bias

Lab Experiments

- Problems generalizing from the lab:
  - **Hawthorne effect**: Lab context influences behaviour, social desirability bias
  - **Context effects**: The real-world always provides more information, more history

Lab Experiments

- Problems generalizing from the lab:
  - **Hawthorne effect**: Lab context influences behaviour, social desirability bias
  - **Context effects**: The real-world always provides more information, more history
  - **Process effects**: People care *how* decisions are made

Lab Experiments

- Problems generalizing from the lab:
  - **Hawthorne effect**: Lab context influences behaviour, social desirability bias
  - **Context effects**: The real-world always provides more information, more history
  - **Process effects**: People care *how* decisions are made
  - **Selection effects**: Actors in specific roles are rarely representative samples, 'WEIRD' or pro-social lab subjects

Lab Experiments

- The lab differs from the field:

Lab Experiments

- The lab differs from the field:
  - The stakes
  - The norms
  - The degree of scrutiny (Levitt and List 2006, “You tip more when you’re on a date”)
  - The sample of individuals
  - The degree of anonymity

Lab Experiments

- Many studies find more cooperation in the lab than in the real world

Lab Experiments

- Many studies find more cooperation in the lab than in the real world
  - Scrutiny increases cooperation

Lab Experiments

- Many studies find more cooperation in the lab than in the real world
  - Scrutiny increases cooperation
  - Anonymity reduces cooperation

Lab Experiments

- Many studies find more cooperation in the lab than in the real world
  - Scrutiny increases cooperation
  - Anonymity reduces cooperation
  - That's interesting in itself! We can manipulate the degree of scrutiny/anonymity etc.

Conjoint Survey Experiments

- Hainmueller et al 2013 - How do attitudes to immigrants depend on immigrant characteristics?

Conjoint Survey Experiments

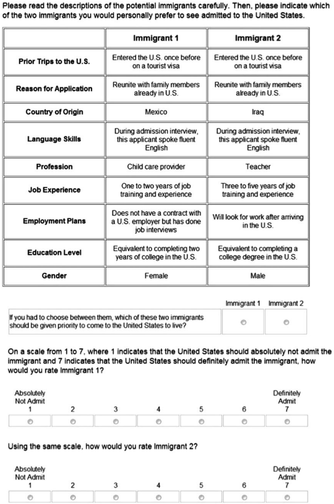
- Hainmueller et al 2013 - How do attitudes to immigrants depend on immigrant characteristics?
- Vary education, profession, language, gender, national origin, etc.

Conjoint Survey Experiments

- Hainmueller et al 2013 - How do attitudes to immigrants depend on immigrant characteristics?
- Vary education, profession, language, gender, national origin, etc.
- Profiles
  - Attributes
    - Values

Conjoint Survey Experiments

- Hainmueller et al 2013 - How do attitudes to immigrants depend on immigrant characteristics?
- Vary education, profession, language, gender, national origin, etc.
- Profiles
  - Attributes
    - Values
- Randomize attribute order to prevent bias



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?	◎	◎

On a scale from 1 to 7, where 1 indicates that the United States should absolutely not admit the immigrant and 7 indicates that the United States should definitely admit the immigrant, how would you rate Immigrant 1?

Absolutely Not Admit 1	2	3	4	5	6	Definitely Admit 7
◎	◎	◎	◎	◎	◎	◎

Using the same scale, how would you rate Immigrant 2?

Absolutely Not Admit 1	2	3	4	5	6	Definitely Admit 7
◎	◎	◎	◎	◎	◎	◎

Fig. 1 Experimental design: Immigration conjoint. This figure illustrates the experimental design for the conjoint analysis that examines immigrant admission to the United States.

choice outcomes hereafter. Second, in "rating-based conjoint analysis," respondents give a numerical rating to each profile which represents their degree of preference for the profile. This format is preferred by some analysts who contend that such ratings provide more direct, finely grained information about respondents' preferences. We call this latter type of outcome a *rating outcome*.

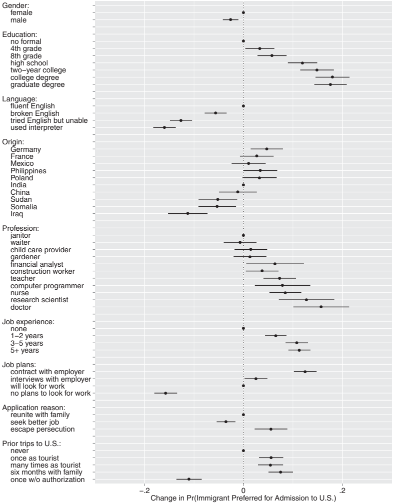


Fig. 3 Effects of immigrant attributes on preference for admission. This plot shows estimates of the effects of the randomly assigned immigrant attributes on the probability of being preferred for admission to the United States. Estimates are based on the regression estimators with clustered standard errors; bars represent 95% confidence intervals. The points without horizontal bars denote the attribute value that is the reference category for each attribute.

Conjoint Survey Experiments

- How realistic are the responses?

Conjoint Survey Experiments

- How realistic are the responses?
  - Not a **behavioural** measure; nothing ’at stake’

Conjoint Survey Experiments

- How realistic are the responses?
  - Not a **behavioural** measure; nothing ’at stake’
  - Social desirability bias

Conjoint Survey Experiments

- How realistic are the responses?
    - Not a **behavioural** measure; nothing ’at stake’
    - Social desirability bias
    - Not like the real-world

Conjoint Survey Experiments

- How realistic are the responses?
  - Not a **behavioural** measure; nothing ’at stake’
  - Social desirability bias
  - Not like the real-world
- Hainmueller et al 2014 - compare conjoint responses to a Swiss referendum

Conjoint Survey Experiments

- How realistic are the responses?
  - Not a **behavioural** measure; nothing ’at stake’
  - Social desirability bias
  - Not like the real-world
- Hainmueller et al 2014 - compare conjoint responses to a Swiss referendum
- Citizens voted on specific naturalization applicants (Really!)

Figure S11: Effects of Applicant Attributes on Opposition to Naturalization Request (Unweighted Survey Sample)

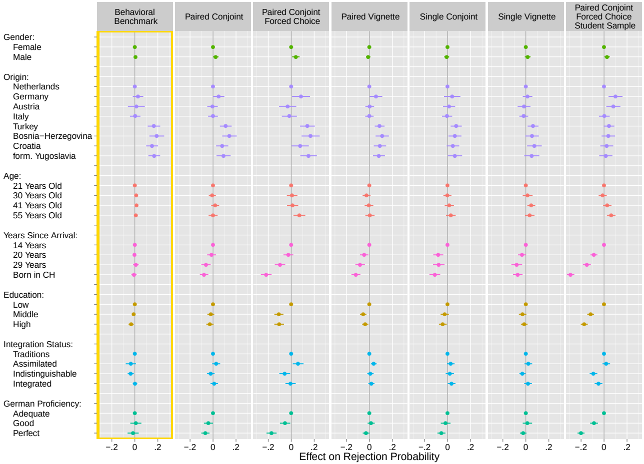


Figure shows point estimates (dots) and corresponding, cluster-robust 95 % confidence intervals (horizontal lines) from ordinary least squares regressions. The dots on the zero line without confidence intervals denote the reference category for each applicant attribute.

Conjoint Survey Experiments

- Marginal effects are quite similar

Conjoint Survey Experiments

- Marginal effects are quite similar
- But note the conjoint method still hugely under-estimated the overall rejection rate
- 21% versus 37% in reality

Regression Discontinuity

- The LATE estimate is for those people who were so close to the discontinuity that whether they were treated or not is basically random

Regression Discontinuity

- The LATE estimate is for those people who were so close to the discontinuity that whether they were treated or not is basically random
  - Even though those cases are rare (eg. tied elections)

Regression Discontinuity

- The LATE estimate is for those people who were so close to the discontinuity that whether they were treated or not is basically random
  - Even though those cases are rare (eg. tied elections)
  - Even though we use data from a lot more people to estimate the LATE

Regression Discontinuity

- The LATE estimate is for those people who were so close to the discontinuity that whether they were treated or not is basically random
  - Even though those cases are rare (eg. tied elections)
  - Even though we use data from a lot more people to estimate the LATE
- Do we care about those people at the discontinuity?

Regression Discontinuity

- The LATE estimate is for those people who were so close to the discontinuity that whether they were treated or not is basically random
  - Even though those cases are rare (eg. tied elections)
  - Even though we use data from a lot more people to estimate the LATE
- Do we care about those people at the discontinuity?
  - It depends on our research/policy question

Regression Discontinuity

- The LATE estimate is for those people who were so close to the discontinuity that whether they were treated or not is basically random
    - Even though those cases are rare (eg. tied elections)
    - Even though we use data from a lot more people to estimate the LATE
- Do we care about those people at the discontinuity?
    - It depends on our research/policy question
    - A trade-off between representativeness and accuracy of our estimates

Regression Discontinuity

- Titiunik et al (2011)

Regression Discontinuity

- Titiunik et al (2011)
  - -6% incumbency effect

Regression Discontinuity

- Titiunik et al (2011)
  - -6% incumbency effect
  - But this does **not** mean that there is a negative incumbency effect in most Brazilian municipalities

Regression Discontinuity

- Titiunik et al (2011)
  - -6% incumbency effect
  - But this does **not** mean that there is a negative incumbency effect in most Brazilian municipalities
  - Only about 500 out of 5,570 municipalities had 'close' elections (within +/-3%)

Regression Discontinuity

- Titiunik et al (2011)
  - -6% incumbency effect
  - But this does **not** mean that there is a negative incumbency effect in most Brazilian municipalities
  - Only about 500 out of 5,570 municipalities had 'close' elections (within +/-3%)
  - Those municipalities were more urban, southern and wealthy than the rest

Regression Discontinuity

- Titiunik et al (2011)
  - -6% incumbency effect
  - But this does **not** mean that there is a negative incumbency effect in most Brazilian municipalities
  - Only about 500 out of 5,570 municipalities had ’close’ elections (within +/-3%)
  - Those municipalities were more urban, southern and wealthy than the rest
  - We do not learn anything about places where the result was a landslide (70-80%)

Regression Discontinuity

- Titiunik et al (2011)
  - -6% incumbency effect
  - But this does **not** mean that there is a negative incumbency effect in most Brazilian municipalities
  - Only about 500 out of 5,570 municipalities had ’close’ elections (within +/-3%)
  - Those municipalities were more urban, southern and wealthy than the rest
  - We do not learn anything about places where the result was a landslide (70-80%)
    - But these are the places where incumbents probably benefitted a lot!

Regression Discontinuity

- Similarly, geographic regression discontinuities only tells us the effect of living on one side of the border *for people who live by the border*

Regression Discontinuity

- Similarly, geographic regression discontinuities only tells us the effect of living on one side of the border *for people who live by the border*
  - But who chooses to live by a border? People who like rural areas, migrants etc.

Regression Discontinuity

- Similarly, geographic regression discontinuities only tells us the effect of living on one side of the border *for people who live by the border*
  - But who chooses to live by a border? People who like rural areas, migrants etc.
  - Self-selection bias has come back!

Instrumental Variables

- Instrumental Variables also estimate LATE
  - A causal effect estimate for **compliers**, units that received treatment *because of variation in the instrument*
  - "Better LATE than never"
- Compliers
- Always-takers
- Never-takers
- Defiers

Instrumental Variables

- Critique of **Opportunism** (Deaton 2009):

Instrumental Variables

- Critique of **Opportunism** (Deaton 2009):
  - If we use ’convenient’ instruments, our causal effect and complier population are out of our control and might not be interesting

Instrumental Variables

- Critique of **Opportunism** (Deaton 2009):
  - If we use ’convenient’ instruments, our causal effect and complier population are out of our control and might not be interesting
  - A risk of chasing impressive research designs instead of asking important questions

Learning in Political Science

- So how much can we learn?

Learning in Political Science

- So how much can we learn?
  - We have to make careful judgments based on internal and external validity

Learning in Political Science

- So how much can we learn?
    - We have to make careful judgments based on internal and external validity
    - Ideally combining multiple methodologies to compare low-bias low-generalizibility evidence with high-bias high-generalizability evidence

Learning in Political Science

- So how much can we learn?
  - We have to make careful judgments based on internal and external validity
  - Ideally combining multiple methodologies to compare low-bias low-generalizibility evidence with high-bias high-generalizability evidence
  - Some topics maybe we simply cannot learn very much.