Causal Inference with Latent Variables

Justin Grimmer

Professor
Department of Political Science
Stanford University
Joint work with Christian Fong, Assistant Professor, University of Michigan

June 20, 2019

Causal Inference with Latent Variables

- Causal Inference With Text-Based Treatments

Causal Inference with Latent Variables

- Causal Inference With Text-Based Treatments
  - Vignette Experiments, Messaging Experiments, Advertising, ...

Causal Inference with Latent Variables

- Causal Inference With Text-Based Treatments
  - Vignette Experiments, Messaging Experiments, Advertising, ...
- Goal: measure causal effect of a latent feature

Causal Inference with Latent Variables

- Causal Inference With Text-Based Treatments
  - Vignette Experiments, Messaging Experiments, Advertising, ...
- Goal: measure causal effect of a latent feature
- Problem: cannot directly randomize that latent feature

Causal Inference with Latent Variables

- Causal Inference With Text-Based Treatments
  - Vignette Experiments, Messaging Experiments, Advertising, ...
- Goal: measure causal effect of a latent feature
- Problem: cannot directly randomize that latent feature
- Causal inference of text usually relies upon an exclusion restriction

Causal Inference with Latent Variables

- Causal Inference With Text-Based Treatments
  - Vignette Experiments, Messaging Experiments, Advertising, ...
- Goal: measure causal effect of a latent feature
- Problem: cannot directly randomize that latent feature
- Causal inference of text usually relies upon an exclusion restriction
- Even in vignette experiments:

Causal Inference with Latent Variables

- Causal Inference With Text-Based Treatments
  - Vignette Experiments, Messaging Experiments, Advertising, ...
- Goal: measure causal effect of a latent feature
- Problem: cannot directly randomize that latent feature
- Causal inference of text usually relies upon an exclusion restriction
- Even in vignette experiments:
  1) Effect may depend on “background" text

Causal Inference with Latent Variables

- Causal Inference With Text-Based Treatments
  - Vignette Experiments, Messaging Experiments, Advertising, ...
- Goal: measure causal effect of a latent feature
- Problem: cannot directly randomize that latent feature
- Causal inference of text usually relies upon an exclusion restriction
- Even in vignette experiments:
  1) Effect may depend on "background" text
  2) Difficult to vary just latent treatment of interest

Causal Inference with Latent Variables

- Causal Inference With Text-Based Treatments
  - Vignette Experiments, Messaging Experiments, Advertising, ...
- Goal: measure causal effect of a latent feature
- Problem: cannot directly randomize that latent feature
- Causal inference of text usually relies upon an exclusion restriction
- Even in vignette experiments:
  1) Effect may depend on “background" text
  2) Difficult to vary just latent treatment of interest
- If measuring treatments and inferring effects: Fundamental Problem of Causal Inference with Latent Variables (FPCILV) leaves estimators undefined

Causal Inference with Latent Variables

- Causal Inference With Text-Based Treatments
  - Vignette Experiments, Messaging Experiments, Advertising, ...
- Goal: measure causal effect of a latent feature
- Problem: cannot directly randomize that latent feature
- Causal inference of text usually relies upon an exclusion restriction
- Even in vignette experiments:
  1) Effect may depend on “background" text
  2) Difficult to vary just latent treatment of interest
- If measuring treatments and inferring effects: Fundamental Problem of Causal Inference with Latent Variables (FPCILV) leaves estimators undefined ,biased

Causal Inference with Latent Variables

- Causal Inference With Text-Based Treatments
  - Vignette Experiments, Messaging Experiments, Advertising, ...
- Goal: measure causal effect of a latent feature
- Problem: cannot directly randomize that latent feature
- Causal inference of text usually relies upon an exclusion restriction
- Even in vignette experiments:
  1) Effect may depend on “background" text
  2) Difficult to vary just latent treatment of interest
- If measuring treatments and inferring effects: Fundamental Problem of Causal Inference with Latent Variables (FPCILV) leaves estimators undefined ,biased, and new risks of fishing or overfitting

Causal Inference with Latent Variables (Fong and Grimmer 2019)

Experimental framework to address issues:

Causal Inference with Latent Variables (Fong and Grimmer 2019)

Experimental framework to address issues:

1) Train/Test split, rather than (or in addition to) a Pre-Analysis Plan (Fafchamps, Labonne 2017; Athey and Wagner 2015; Chernozhukov et al 2017; Cranmer and Desmaris 2016, Egami et al 2019; Fong and Grimmer 2016)

Causal Inference with Latent Variables (Fong and Grimmer 2019)

Experimental framework to address issues:

1) Train/Test split, rather than (or in addition to) a Pre-Analysis Plan (Fafchamps, Labonne 2017; Athey and Wagner 2015; Chernozhukov et al 2017; Cranmer and Desmaris 2016, Egami et al 2019; Fong and Grimmer 2016)
2) Multiple vignettes per treatment

Causal Inference with Latent Variables (Fong and Grimmer 2019)

Experimental framework to address issues:

1) Train/Test split, rather than (or in addition to) a Pre-Analysis Plan (Fafchamps, Labonne 2017; Athey and Wagner 2015; Chernozhukov et al 2017; Cranmer and Desmaris 2016, Egami et al 2019; Fong and Grimmer 2016)
2) Multiple vignettes per treatment
3) Enable simultaneous estimation of treatments and their effect

Causal Inference with Latent Variables (Fong and Grimmer 2019)

Experimental framework to address issues:

1) Train/Test split, rather than (or in addition to) a Pre-Analysis Plan (Fafchamps, Labonne 2017; Athey and Wagner 2015; Chernozhukov et al 2017; Cranmer and Desmaris 2016, Egami et al 2019; Fong and Grimmer 2016)
2) Multiple vignettes per treatment
3) Enable simultaneous estimation of treatments and their effect

Apply framework to infer public reaction to Trump tweets

An Interdisciplinary Interest in Causal Inferences with Text

An Interdisciplinary Interest in Causal Inferences with Text

- An astonishing avalanche of new data and decreasing costs of experiments

An Interdisciplinary Interest in Causal Inferences with Text

- An astonishing avalanche of new data and decreasing costs of experiments
- New methods adapted as a tool for measurement.

An Interdisciplinary Interest in Causal Inferences with Text

- An astonishing avalanche of new data and decreasing costs of experiments
- New methods adapted as a tool for measurement.
- Allows us to use rich, high-dimensional data, one example: text

An Interdisciplinary Interest in Causal Inferences with Text

- An astonishing avalanche of new data and decreasing costs of experiments
- New methods adapted as a tool for measurement.
- Allows us to use rich, high-dimensional data, one example: text
- An interdisciplinary moment which includes: computer science, statistics, digital humanities and social sciences

An Interdisciplinary Interest in Causal Inferences with Text

- An astonishing avalanche of new data and decreasing costs of experiments
- New methods adapted as a tool for measurement.
- Allows us to use rich, high-dimensional data, one example: text
- An interdisciplinary moment which includes: computer science, statistics, digital humanities and social sciences
- Allows us to discover new patterns and find new concepts

An Interdisciplinary Interest in Causal Inferences with Text

- An astonishing avalanche of new data and decreasing costs of experiments
- New methods adapted as a tool for measurement.
- Allows us to use rich, high-dimensional data, one example: text
- An interdisciplinary moment which includes: computer science, statistics, digital humanities and social sciences
- Allows us to discover new patterns and find new concepts
- Increasing interest in using text-based measures for causal inference

An Interdisciplinary Interest in Causal Inferences with Text

An Interdisciplinary Interest in Causal Inferences with Text

Increasing prevelance of experiments, increasing concerns about their validity

An Interdisciplinary Interest in Causal Inferences with Text

Increasing prevelance of experiments, increasing concerns about their validity

- Concerns about credibility

An Interdisciplinary Interest in Causal Inferences with Text

Increasing prevelance of experiments, increasing concerns about their validity

- Concerns about credibility
- Concerns about extrapolation

An Interdisciplinary Interest in Causal Inferences with Text

Increasing prevelance of experiments, increasing concerns about their validity

- Concerns about credibility
- Concerns about extrapolation
- Concerns about attribution, interpretation, and equilibrium (Bueno de Mesquita and Tyson, 2019)

What features of Trump's rhetoric cause a reaction? $\rightsquigarrow$ Our framework will blend A/B Tests with more traditional vignette experiments

Donald J. Trump
@realDonaldTrump

Following

Little Adam Schiff, who is desperate to run for higher office, is one of the biggest liars and leakers in Washington, right up there with Comey, Warner, Brennan and Clapper! Adam leaves closed committee hearings to illegally leak confidential information. Must be stopped!

4:39 AM - 5 Feb 2018

31,930 Retweets **99,706** Likes

48K 32K 100K

Tweet 1:

Why would Kim Jong-un insult me by calling me “old," when I would NEVER call him “short and fat?" Oh well, I try so hard to be his friend-and maybe someday that will happen!

Tweet 2:

Steve Bannon will be a tough and smart new voice at @BreitbartNews...maybe even better than ever before. Fake News needs the competition!

Tweet 1:

```
Why would Kim Jong-un insult me by calling me “old,"
when I would NEVER call him “short and fat?" Oh
well, I try so hard to be his friend-and maybe
someday that will happen!
```

Tweet 2:

```
Steve Bannon will be a tough and smart new voice at
@BreitbartNews...maybe even better than ever before.
Fake News needs the competition!
```

Observe difference in evaluations of biographies

Tweet 1:

`Why would Kim Jong-un insult me by calling me “old," when I would NEVER call him “short and fat?" Oh well, I try so hard to be his friend-and maybe someday that will happen!`

Tweet 2:

`Steve Bannon will be a tough and smart new voice at @BreitbartNews...maybe even better than ever before. Fake News needs the competition!`

Observe difference in evaluations of biographies $\rightsquigarrow$ Difficult to generalize underlying features (treatments) that drive response

Tweet 1:

Steve Bannon will be a tough and smart new voice at @BreitbartNews...maybe even better than ever before. Fake News needs the competition!

Tweet $1'$:

Steve Bannon will be a tough and smart new voice at @BreitbartNews...maybe even better than ever before. ~~Fake~~ News needs the competition!

Tweet 1:

`Steve Bannon will be a tough and smart new voice at @BreitbartNews...maybe even better than ever before. Fake News needs the competition!`

Tweet $1'$:

Steve Bannon will be a tough and smart new voice at @BreitbartNews...maybe even better than ever before. ~~Fake~~ News needs the competition!

Randomly assign $1, 1'$ and assess response $\rightsquigarrow$ are we interested in effect of one word?

Tweet 1:

`Steve Bannon will be a tough and smart new voice at @BreitbartNews...maybe even better than ever before. Fake News needs the competition!`

Tweet $1'$:

`Steve Bannon will be a tough and smart new voice at @BreitbartNews...maybe even better than ever before.` ~~`Fake`~~ `News needs the competition!`

Randomly assign 1, $1'$ and assess response $\rightsquigarrow$ are we interested in effect of one word?

Effect will be dependent on "background" text

Tweet 1:

Negotiations on DACA have begun. Republicans want to make a deal and Democrats say they want to make a deal. Wouldn't it be great if we could finally, after so many years, solve the DACA puzzle. This will be our last chance, there will never be another opportunity! March 5th.

Tweet 2:

Negotiations on DACA have begun. Republicans want to make a deal and Democrats say they want to make a deal. Wouldn't it be great if we could finally, after so many years, solve the DACA puzzle. I will use my office to negotiate a fair deal for the dreamers! March 5th.

Tweet 1:

Negotiations on DACA have begun. Republicans want to make a deal and Democrats say they want to make a deal. Wouldn't it be great if we could finally, after so many years, solve the DACA puzzle. This will be our last chance, there will never be another opportunity! March 5th.

Tweet 2:

Negotiations on DACA have begun. Republicans want to make a deal and Democrats say they want to make a deal. Wouldn't it be great if we could finally, after so many years, solve the DACA puzzle. I will use my office to negotiate a fair deal for the dreamers! March 5th.

Latent Representation (Codebook) $\rightsquigarrow$ true whether hand coded, supervised, or unsupervised

Tweet 1:

Negotiations on DACA have begun. Republicans want to make a deal and Democrats say they want to make a deal. Wouldn't it be great if we could finally, after so many years, solve the DACA puzzle. This will be our last chance, there will never be another opportunity! March 5th.

Tweet 2:

Negotiations on DACA have begun. Republicans want to make a deal and Democrats say they want to make a deal. Wouldn't it be great if we could finally, after so many years, solve the DACA puzzle. I will use my office to negotiate a fair deal for the dreamers! March 5th.

Many ways for different words to deliver same latent treatment

Tweet 1:

Negotiations on DACA have begun. Republicans want to make a deal and Democrats say they want to make a deal. Wouldn't it be great if we could finally, after so many years, solve the DACA puzzle. This will be our last chance, there will never be another opportunity! March 5th.

Tweet 2:

Negotiations on DACA have begun. Republicans want to make a deal and Democrats say they want to make a deal. Wouldn't it be great if we could finally, after so many years, solve the DACA puzzle. I will use my office to negotiate a fair deal for the dreamers! March 5th.

May change more than the theoretical treatment of interest

What is a Social Scientist To Do?

Issues to consider:

What is a Social Scientist To Do?

Issues to consider:

1) Cannot randomize latent treatment directly

What is a Social Scientist To Do?

Issues to consider:

1) Cannot randomize latent treatment directly
2) Requires "codebook" to map from text $\rightsquigarrow$ treatment

What is a Social Scientist To Do?

Issues to consider:

1) Cannot randomize latent treatment directly
2) Requires "codebook" to map from text $\rightsquigarrow$ treatment
3) Effect could depend on background text

What is a Social Scientist To Do?

Issues to consider:

1) Cannot randomize latent treatment directly
2) Requires "codebook" to map from text $\rightsquigarrow$ treatment
3) Effect could depend on background text
4) More than treatment of interest could change

What is a Social Scientist To Do?

Issues to consider:

1) Cannot randomize latent treatment directly
2) Requires "codebook" to map from text $\rightsquigarrow$ treatment
3) Effect could depend on background text
4) More than treatment of interest could change

In addition to usual credibility concerns

What is a Social Scientist To Do?

Specify a pre-analysis plan?

Specify a pre-analysis plan? It could:

Specify a pre-analysis plan? It could:

- Define codebook/latent treatments

Specify a pre-analysis plan? It could:

- Define codebook/latent treatments
- Procedure for determining vignettes

Specify a pre-analysis plan? It could:

- Define codebook/latent treatments
- Procedure for determining vignettes
- Procedure for minimizing (estimating) background text influence

What is a Social Scientist To Do?

Specify a pre-analysis plan? It could:

- Define codebook/latent treatments
- Procedure for determining vignettes
- Procedure for minimizing (estimating) background text influence
- Limit credibility concerns

Pre-Analysis Plan (PAP)

Prior to analysis: declare plan

1) Declare all hypotheses
2) Specify all coding rules
3) Declare all statistical procedures
4) Commit to reporting all results

Output: Algorithm that takes new data and produces results

After analysis: report in line with declaration

1) Report all hypotheses
2) Report all results

This is a useful framework! $\rightsquigarrow$ But what actually happens?

Grimmer, Lal, and Renshon (In Progress)

Plans $\rightsquigarrow$ Papers

Papers $\rightsquigarrow$ Plans

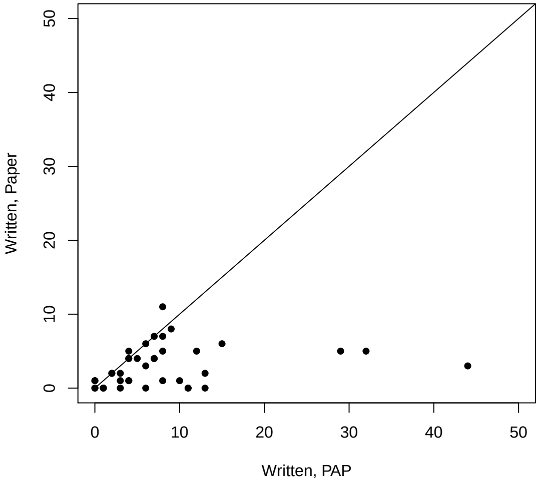
Analyzing Pre-Analysis Plans

Evidence in Governance And Politics: EGAP

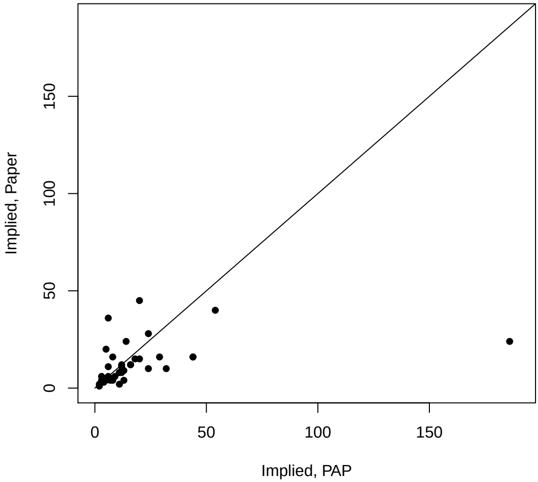
`http://egap.org/design-registrations`

- Pre-registration data:
    - Meta-data: Title, link, Year, Authors
    - Content: # Written Hypotheses, # Statistical Hypotheses
- Paper data:
    - Meta data: Journal, title, link
    - Content: # Written Hypotheses, # Statistical Hypotheses, # References to PAP

Comparing Written Hypotheses in PAP and Papers



Comparing Hypotheses Implied from Statistics



Paper$\rightsquigarrow$ PAP

- 31: Top journal articles in PS (APSR / AJPS / JOP / PA / QJPS) and Economics (AER / QJE / AEJs / JPubE / JDevE) mention PAPs
- 26/31 : claimed to have used preanalysis plans
- 12/26: linked to the PAP (11/26 of these worked; 1 article linked to a pdf now defunct personal website)
- Average number of mentions of PAPs is 2.7, modal article mentions the preanalysis plan in the text and footnote.

How to Think About PAPs and Credibility?

Experiments in a sequence $\rightsquigarrow$ market incentives for replication

Train/Test Split

Train/Test Split

1) Explicitly set aside a (random) training set for discovery.
2) Use this Training data to (1) estimate codebook, (2) determine any data errors, (3) assess assumption needed for experiment
3) Test data: apply model from training data

Train/Test Split

1) Explicitly set aside a (random) training set for discovery.
2) Use this Training data to (1) estimate codebook, (2) determine any data errors, (3) assess assumption needed for experiment
3) Test data: apply model from training data

Concerns:

Train/Test Split

1) Explicitly set aside a (random) training set for discovery.
2) Use this Training data to (1) estimate codebook, (2) determine any data errors, (3) assess assumption needed for experiment
3) Test data: apply model from training data

Concerns:

- Power?

Train/Test Split

1) Explicitly set aside a (random) training set for discovery.
2) Use this Training data to (1) estimate codebook, (2) determine any data errors, (3) assess assumption needed for experiment
3) Test data: apply model from training data

Concerns:

- Power?
- Single very expensive experiment?

Train/Test Split

1) Explicitly set aside a (random) training set for discovery.
2) Use this Training data to (1) estimate codebook, (2) determine any data errors, (3) assess assumption needed for experiment
3) Test data: apply model from training data

Concerns:

- Power?
- Single very expensive experiment?
- Gaming the splits?

A New Framework for Text-Based Experiments

- Assume "Interesting" treatments (coding) must be known in advance

A New Framework for Text-Based Experiments

- Assume "Interesting" treatments (coding) must be known in advance
- Discovery of treatments may (often/usually) happen after viewing data

A New Framework for Text-Based Experiments

- Assume “Interesting” treatments (coding) must be known in advance
- Discovery of treatments may (often/usually) happen after viewing data
- Explicit discovery phase in research

Three Key Steps to our Framework

Three Key Steps to our Framework

1) Theory: conditions to identify marginal effects of latent treatments (Average Marginal Component Effect (AMCE) is identified)

Three Key Steps to our Framework

1) Theory: conditions to identify marginal effects of latent treatments (Average Marginal Component Effect (AMCE) is identified)

2) Method for discovering features (treatments)

Three Key Steps to our Framework

1) Theory: conditions to identify marginal effects of latent treatments (Average Marginal Component Effect (AMCE) is identified)
2) Method for discovering features (treatments)
3) Method for estimating marginal effect for discovered features (treatments)

Identifying the Marginal Effects of Latent Treatments

- Average Marginal Component Effect (AMCE): Isolate effect of one treatment, averaging over other treatments

Identifying the Marginal Effects of Latent Treatments

- Average Marginal Component Effect (AMCE): Isolate effect of one treatment, averaging over other treatments
- Let $\boldsymbol{Z}_i$ be i's binary feature vector
- Ex: $\boldsymbol{Z}_i = (0, 0, 1, 1, 0)$

Identifying the Marginal Effects of Latent Treatments

- Average Marginal Component Effect (AMCE): Isolate effect of one treatment, averaging over other treatments
- Let $\boldsymbol{Z}_i$ be i's binary feature vector
- Ex: $\boldsymbol{Z}_i = (0, 0, 1, 1, 0)$

$$\text{AMCE}_k = \int_{\boldsymbol{Z}_{-k}} \mathbb{E}\left[Y(Z_k = 1, \boldsymbol{Z}_{-k}) - Y(Z_k = 0, \boldsymbol{Z}_{-k})\right] m(\boldsymbol{Z}_{-k}) d\boldsymbol{Z}_{-k}$$

Identifying the Marginal Effects of Latent Treatments

- Average Marginal Component Effect (AMCE): Isolate effect of one treatment, averaging over other treatments
- Let $\boldsymbol{Z}_i$ be i's binary feature vector
- Ex: $\boldsymbol{Z}_i = (0, 0, 1, 1, 0)$

$$\text{AMCE}_k = \int_{\boldsymbol{Z}_{-k}} \mathbb{E}\left[Y(Z_k = 1, \boldsymbol{Z}_{-k}) - Y(Z_k = 0, \boldsymbol{Z}_{-k})\right] m(\boldsymbol{Z}_{-k}) d\boldsymbol{Z}_{-k}$$

Identifying the Marginal Effects of Latent Treatments

- Average Marginal Component Effect (AMCE): Isolate effect of one treatment, averaging over other treatments

- Let $\boldsymbol{Z}_i$ be i's binary feature vector

- Ex: $\boldsymbol{Z}_i = (0, 0, 1, 1, 0)$

$$\text{AMCE}_k = \int_{\boldsymbol{Z}_{-k}} \mathbb{E}\left[Y(Z_k = 1, \boldsymbol{Z}_{-k}) - Y(Z_k = 0, \boldsymbol{Z}_{-k})\right] m(\boldsymbol{Z}_{-k}) d\boldsymbol{Z}_{-k}$$

Identifying the Marginal Effects of Latent Treatments

- Average Marginal Component Effect (AMCE): Isolate effect of one treatment, averaging over other treatments
- Let $\boldsymbol{Z}_i$ be i's binary feature vector
- Ex: $\boldsymbol{Z}_i = (0, 0, 1, 1, 0)$

$$\text{AMCE}_k = \int_{\boldsymbol{Z}_{-k}} \mathbb{E}\left[Y(Z_k = 1, \boldsymbol{Z}_{-k}) - Y(Z_k = 0, \boldsymbol{Z}_{-k})\right] m(\boldsymbol{Z}_{-k}) d\boldsymbol{Z}_{-k}$$

Identifying the Marginal Effects of Latent Treatments

- Average Marginal Component Effect (AMCE): Isolate effect of one treatment, averaging over other treatments
- Let $\boldsymbol{Z}_i$ be i's binary feature vector
- Ex: $\boldsymbol{Z}_i = (0, 0, 1, 1, 0)$

$$\text{AMCE}_k = \int_{\boldsymbol{Z}_{-k}} \mathbb{E}\left[Y(Z_k = 1, \boldsymbol{Z}_{-k}) - Y(Z_k = 0, \boldsymbol{Z}_{-k})\right] m(\boldsymbol{Z}_{-k}) d\boldsymbol{Z}_{-k}$$

Conjoint With Discovered Treatments

Identifying the Marginal Effects of Latent Treatments

- Average Marginal Component Effect (AMCE): Isolate effect of one treatment, averaging over other treatments
- Let $\boldsymbol{Z}_i$ be i's binary feature vector
- Ex: $\boldsymbol{Z}_i = (0, 0, 1, 1, 0)$

$$\text{AMCE}_k = \int_{\boldsymbol{Z}_{-k}} \mathbb{E}\left[Y(Z_k = 1, \boldsymbol{Z}_{-k}) - Y(Z_k = 0, \boldsymbol{Z}_{-k})\right] m(\boldsymbol{Z}_{-k}) d\boldsymbol{Z}_{-k}$$

Conjoint With Discovered Treatments (or) Discover Features that Drive Response in A/B Test

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: text seen by i)
- Function (assume known for now): text $\rightsquigarrow$ treatments in text ($\boldsymbol{Z}_i \equiv g(\boldsymbol{X}_i)$)

  $\boldsymbol{Z}_i$ is a low-dimensional rep of $\boldsymbol{X}_i$, describing treatments

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: text seen by i)
- Function (assume known for now): text $\rightsquigarrow$ treatments in text ($\boldsymbol{Z}_i \equiv g(\boldsymbol{X}_i)$)

  $\boldsymbol{Z}_i$ is a low-dimensional rep of $\boldsymbol{X}_i$, describing treatments

Assume:

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: text seen by i)
- Function (assume known for now): text $\rightsquigarrow$ treatments in text ($\boldsymbol{Z}_i \equiv g(\boldsymbol{X}_i)$)

  $\boldsymbol{Z}_i$ is a low-dimensional rep of $\boldsymbol{X}_i$, describing treatments

Assume:

1) No "spillover" (SUTVA, Rubin 1986: $\tilde{Y}_i(\boldsymbol{X}) = \tilde{Y}_i(\boldsymbol{X}_i)$)

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: text seen by i)
- Function (assume known for now): text $\rightsquigarrow$ treatments in text ($\boldsymbol{Z}_i \equiv g(\boldsymbol{X}_i)$)

  $\boldsymbol{Z}_i$ is a low-dimensional rep of $\boldsymbol{X}_i$, describing treatments

Assume:

1) No "spillover" (SUTVA, Rubin 1986: $\tilde{Y}_i(\boldsymbol{X}) = \tilde{Y}_i(\boldsymbol{X}_i)$)
2) Random assignment of texts ($\tilde{Y}_i(\boldsymbol{X}_i) \perp\!\!\!\perp \boldsymbol{X}_i$ for all i)

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: text seen by i)
- Function (assume known for now): text $\rightsquigarrow$ treatments in text ($\boldsymbol{Z}_i \equiv g(\boldsymbol{X}_i)$)

  $\boldsymbol{Z}_i$ is a low-dimensional rep of $\boldsymbol{X}_i$, describing treatments

Assume:

1) No "spillover" (SUTVA, Rubin 1986: $\tilde{Y}_i(\boldsymbol{X}) = \tilde{Y}_i(\boldsymbol{X}_i)$)
2) Random assignment of texts ($\tilde{Y}_i(\boldsymbol{X}_i) \perp\!\!\!\perp \boldsymbol{X}_i$ for all i)
3) Exclusion Restriction: Treatment of interest g is independent of other text features (Grimmer and Fong 2019)

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: text seen by i)
- Function (assume known for now): text $\rightsquigarrow$ treatments in text ($\boldsymbol{Z}_i \equiv g(\boldsymbol{X}_i)$)

  $\boldsymbol{Z}_i$ is a low-dimensional rep of $\boldsymbol{X}_i$, describing treatments

Assume:

1) No "spillover" (SUTVA, Rubin 1986: $\tilde{Y}_i(\boldsymbol{X}) = \tilde{Y}_i(\boldsymbol{X}_i)$)
2) Random assignment of texts ($\tilde{Y}_i(\boldsymbol{X}_i) \perp\!\!\!\perp \boldsymbol{X}_i$ for all i)
3) Exclusion Restriction: Treatment of interest g is independent of other text features (Grimmer and Fong 2019)
4) Common support: all combinations of treatments have non-zero probability ($f(\boldsymbol{Z}_i) > 0$ for all $\boldsymbol{Z}_i \in$ Range $g(\cdot)$)

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: text seen by i)
- Function (assume known for now): text $\rightsquigarrow$ treatments in text ($\boldsymbol{Z}_i \equiv g(\boldsymbol{X}_i)$)

  $\boldsymbol{Z}_i$ is a low-dimensional rep of $\boldsymbol{X}_i$, describing treatments

Assume:

1) No "spillover" (SUTVA, Rubin 1986: $\tilde{Y}_i(\boldsymbol{X}) = \tilde{Y}_i(\boldsymbol{X}_i)$)
2) Random assignment of texts ($\tilde{Y}_i(\boldsymbol{X}_i) \perp\!\!\!\perp \boldsymbol{X}_i$ for all i)
3) Exclusion Restriction: Treatment of interest g is independent of other text features (Grimmer and Fong 2019)
4) Common support: all combinations of treatments have non-zero probability ($f(\boldsymbol{Z}_i) > 0$ for all $\boldsymbol{Z}_i \in$ Range $g(\cdot)$)

Proposition 1

Assumptions 1-4 are sufficient to identify the $AMCE_k$ for arbitrary k.

Applying Framework to Vignette Experiments

Standard vignette experiment

Applying Framework to Vignette Experiments

Standard vignette experiment $\rightsquigarrow$ one vignette per latent treatment

Applying Framework to Vignette Experiments

Standard vignette experiment $\rightsquigarrow$ one vignette per latent treatment

Two additional assumptions for experiment to identify AMCE:

Applying Framework to Vignette Experiments

Standard vignette experiment $\rightsquigarrow$ one vignette per latent treatment

Two additional assumptions for experiment to identify AMCE:

1) Latent treatment's effects do not depend on "background" text

Applying Framework to Vignette Experiments

Standard vignette experiment $\rightsquigarrow$ one vignette per latent treatment

Two additional assumptions for experiment to identify AMCE:

1) Latent treatment's effects do not depend on "background" text
2) Latent treatments are not aliased

Applying Framework to Vignette Experiments

Standard vignette experiment $\rightsquigarrow$ one vignette per latent treatment

Two additional assumptions for experiment to identify AMCE:

1) Latent treatment's effects do not depend on "background" text
2) Latent treatments are not aliased

Fong and Grimmer (2019): Many vignettes per treatment enable sensitivity analysis

Discovering Treatments and Estimating Marginal Effects

Discovery of Treatments from Text Corpora

Discovery of Treatments from Text Corpora

1) (Assume) Randomly assign texts, $\boldsymbol{X}_i$, to respondents

Discovery of Treatments from Text Corpora

1) (Assume) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent

Discovery of Treatments from Text Corpora

1) (Assume) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) (Randomly) divide texts and responses into training and test set

Discovery of Treatments from Text Corpora

1) (Assume) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) (Randomly) divide texts and responses into training and test set
    a) Avoid identification issues

Discovery of Treatments from Text Corpora

1) (Assume) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) (Randomly) divide texts and responses into training and test set
    a) Avoid identification issues
    b) Help avoid overfitting

What Can Go Wrong with the Fundamental Problem of Causal Inference with Latent Variables?

What Can Go Wrong with the Fundamental Problem of Causal Inference with Latent Variables?

	Treated	Control
Person 1	Candidate Morals	Taxes
Person 2	Candidate Morals	Taxes
Person 3	Polarization	Immigration
Person 4	Polarization	Immigration

What Can Go Wrong with the Fundamental Problem of Causal Inference with Latent Variables?

	Treated	Control
Person 1	Candidate Morals	Taxes
Person 2	Candidate Morals	Taxes
Person 3	Polarization	Immigration
Person 4	Polarization	Immigration

- $\boldsymbol{T}_1 = (1, 1, 0, 0) \rightsquigarrow$ Categories: Candidate Morals, Immigration

What Can Go Wrong with the Fundamental Problem of Causal Inference with Latent Variables?

	Treated	Control
Person 1	Candidate Morals	Taxes
Person 2	Candidate Morals	Taxes
Person 3	Polarization	Immigration
Person 4	Polarization	Immigration

- $\boldsymbol{T}_1 = (1, 1, 0, 0) \rightsquigarrow$ Categories: Candidate Morals, Immigration
- $\boldsymbol{T}_2 = (1, 0, 1, 0) \rightsquigarrow$ Categories: Candidate Morals, Taxes, Polarization, Immigration

What Can Go Wrong with the Fundamental Problem of Causal Inference with Latent Variables?

	Treated	Control
Person 1	Candidate Morals	Taxes
Person 2	Candidate Morals	Taxes
Person 3	Polarization	Immigration
Person 4	Polarization	Immigration

- $\boldsymbol{T}_1 = (1, 1, 0, 0) \rightsquigarrow$ Categories: Candidate Morals, Immigration
- $\boldsymbol{T}_2 = (1, 0, 1, 0) \rightsquigarrow$ Categories: Candidate Morals, Taxes, Polarization, Immigration

Cannot compare categories from $\boldsymbol{T}_1$ and $\boldsymbol{T}_2 \rightsquigarrow$ properties of estimator (bias, consistency) not defined!

What Can Go Wrong with the Fundamental Problem of Causal Inference with Latent Variables?

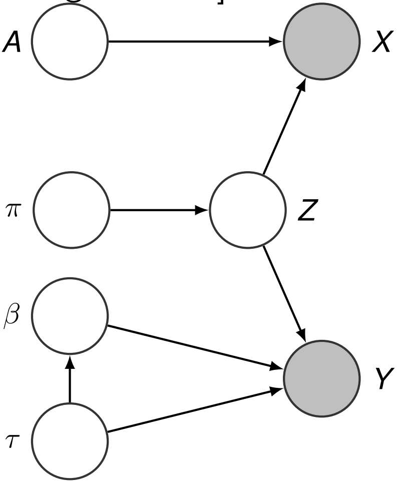
	Treated	Control
Person 1	Candidate Morals	Taxes
Person 2	Candidate Morals	Taxes
Person 3	Polarization	Immigration
Person 4	Polarization	Immigration

- $\boldsymbol{T}_1 = (1, 1, 0, 0) \rightsquigarrow$ Categories: Candidate Morals, Immigration
- $\boldsymbol{T}_2 = (1, 0, 1, 0) \rightsquigarrow$ Categories: Candidate Morals, Taxes, Polarization, Immigration

Cannot compare categories from $\boldsymbol{T}_1$ and $\boldsymbol{T}_2 \rightsquigarrow$ properties of estimator (bias, consistency) not defined!
Focus on instances where specific g of interest is discovered $\rightsquigarrow$ Biased, inconsistent estimator unless discovery unrelated to effect

Discovery method for a g

The Supervised Indian Buffet Process (sIBP, distinct [though related] to Quadrianto et al 2013)



Text and response depend on latent treatments

- Treatment assignment

$$z_{i,k} \sim \text{Bernoulli}(\pi_k)$$
$$\pi_k \sim \prod_{m=1}^{k} \eta_m$$
$$\eta_m \sim \text{Beta}(\alpha, 1)$$

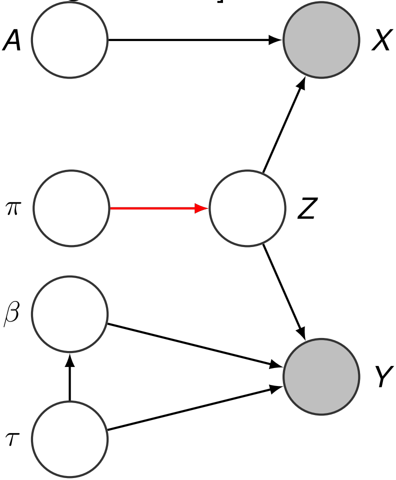
- Document Creation:

$$\boldsymbol{X}_i \sim \text{MVN}(\boldsymbol{Z}_i \boldsymbol{A}, \sigma_X^2 I_D)$$
$$\boldsymbol{A}_k \sim \text{MVN}(\boldsymbol{0}, \sigma_A^2 I_D)$$

- Response:

$$Y_i \sim \text{MVN}(Z_i \beta, \tau^{-1})$$
$$\beta|\tau \sim \text{MVN}(\boldsymbol{0}, \tau^{-1} I_K)$$
$$\tau \sim \text{Gamma}(a, b)$$

The Supervised Indian Buffet Process (sIBP, distinct [though related] to Quadrianto et al 2013)



Text and response depend on latent treatments

- Treatment assignment

$$z_{i,k} \sim \text{Bernoulli}(\pi_k)$$
$$\pi_k \sim \prod_{m=1}^{k} \eta_m$$
$$\eta_m \sim \text{Beta}(\alpha, 1)$$

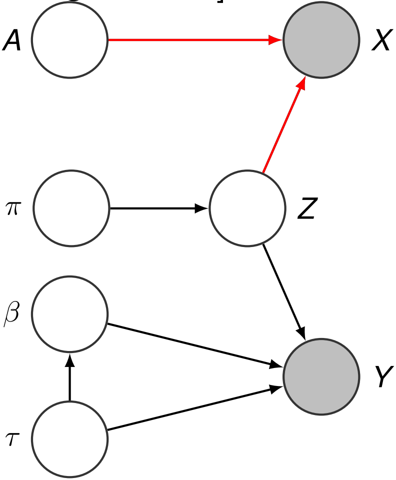
- Document Creation:

$$\boldsymbol{X}_i \sim \text{MVN}(\boldsymbol{Z}_i \boldsymbol{A}, \sigma_X^2 I_D)$$
$$\boldsymbol{A}_k \sim \text{MVN}(\mathbf{0}, \sigma_A^2 I_D)$$

- Response:

$$Y_i \sim \text{MVN}(Z_i \boldsymbol{\beta}, \tau^{-1})$$
$$\boldsymbol{\beta}|\tau \sim \text{MVN}(\mathbf{0}, \tau^{-1} I_K)$$
$$\tau \sim \text{Gamma}(a, b)$$

The Supervised Indian Buffet Process (sIBP, distinct [though related] to Quadrianto et al 2013)



Text and response depend on latent treatments

- Treatment assignment

$$
\begin{aligned}
z_{i,k} &\sim \text{Bernoulli}(\pi_k) \\
\pi_k &\sim \prod_{m=1}^{k} \eta_m \\
\eta_m &\sim \text{Beta}\,(\alpha, 1)
\end{aligned}
$$

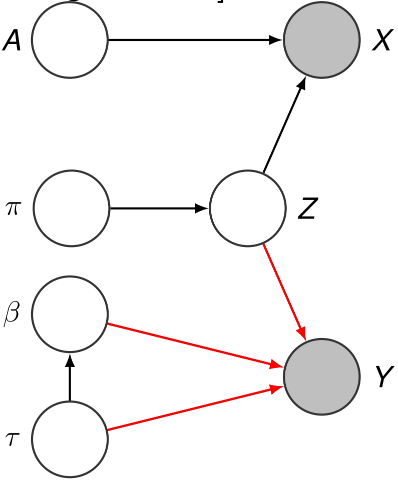
- Document Creation:

$$
\begin{aligned}
\boldsymbol{X}_i &\sim \text{MVN}(\boldsymbol{Z}_i \boldsymbol{A}, \sigma_X^2 I_D) \\
\boldsymbol{A}_k &\sim \text{MVN}(\boldsymbol{0}, \sigma_A^2 I_D)
\end{aligned}
$$

- Response:

$$
\begin{aligned}
Y_i &\sim \text{MVN}(Z_i \boldsymbol{\beta}, \tau^{-1}) \\
\boldsymbol{\beta}|\tau &\sim \text{MVN}(\boldsymbol{0}, \tau^{-1} I_K) \\
\tau &\sim \text{Gamma}(a, b)
\end{aligned}
$$

The Supervised Indian Buffet Process (sIBP, distinct [though related] to Quadrianto et al 2013)



Text and response depend on latent treatments

- Treatment assignment

$$
\begin{aligned}
z_{i,k} &\sim \text{Bernoulli}(\pi_k) \\
\pi_k &\sim \prod_{m=1}^{k} \eta_m \\
\eta_m &\sim \text{Beta}\,(\alpha, 1)
\end{aligned}
$$

- Document Creation:

$$
\begin{aligned}
\boldsymbol{X}_i &\sim \text{MVN}(\boldsymbol{Z}_i \boldsymbol{A}, \sigma_X^2 I_D) \\
\boldsymbol{A}_k &\sim \text{MVN}(\boldsymbol{0}, \sigma_A^2 I_D)
\end{aligned}
$$

- Response:

$$
\begin{aligned}
Y_i &\sim \text{MVN}(Z_i \boldsymbol{\beta}, \tau^{-1}) \\
\boldsymbol{\beta}|\tau &\sim \text{MVN}(\boldsymbol{0}, \tau^{-1} I_K) \\
\tau &\sim \text{Gamma}(a, b)
\end{aligned}
$$

Discovery of Treatments from Text Corpora

1) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) Divide texts and responses into training and test set
4) In training set: Discover mapping from texts to treatments

Discovery of Treatments from Text Corpora

1) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) Divide texts and responses into training and test set
4) In training set: Discover mapping from texts to treatments
    a) Apply supervised Indian Buffet Process (sIBP) to documents and responses to infer latent treatments in texts

Discovery of Treatments from Text Corpora

1) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) Divide texts and responses into training and test set
4) In training set: Discover mapping from texts to treatments
    a) Apply supervised Indian Buffet Process (sIBP) to documents and responses to infer latent treatments in texts
    b) Model selection via nonparametric process, quantitative fit, and qualitative assessment

Discovery of Treatments from Text Corpora

1) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) Divide texts and responses into training and test set
4) In training set: Discover mapping from texts to treatments
    a) Apply supervised Indian Buffet Process (sIBP) to documents and responses to infer latent treatments in texts
    b) Model selection via nonparametric process, quantitative fit, and qualitative assessment
5) In test set: infer treatments and measure their effect

Discovery of Treatments from Text Corpora

1) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) Divide texts and responses into training and test set
4) In training set: Discover mapping from texts to treatments
    a) Apply supervised Indian Buffet Process (sIBP) to documents and responses to infer latent treatments in texts
    b) Model selection via nonparametric process, quantitative fit, and qualitative assessment
5) In test set: infer treatments and measure their effect
    a) Use sIBP trained on training set to infer latent treatments on test set documents (without conditioning on test set responses)

Discovery of Treatments from Text Corpora

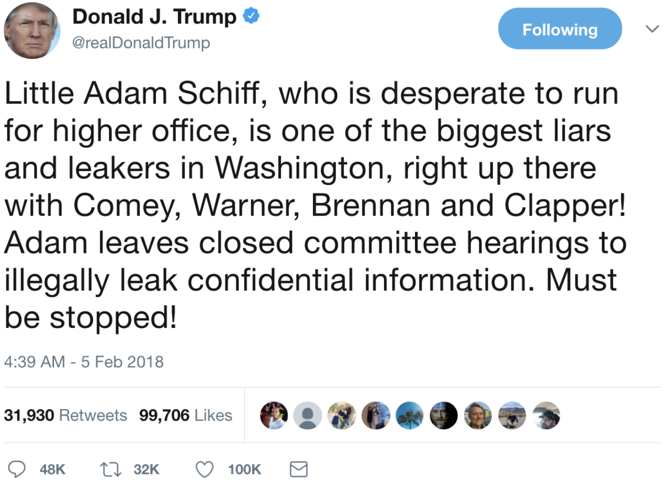
1) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) Divide texts and responses into training and test set
4) In training set: Discover mapping from texts to treatments
    a) Apply supervised Indian Buffet Process (sIBP) to documents and responses to infer latent treatments in texts
    b) Model selection via nonparametric process, quantitative fit, and qualitative assessment
5) In test set: infer treatments and measure their effect
    a) Use sIBP trained on training set to infer latent treatments on test set documents (without conditioning on test set responses)
    b) Estimate effect of treatments with regression, with a bootstrap procedure to estimate uncertainty

Trump Tweets

`YouGov`: survey response to trump tweets

Trump Tweets

YouGov: survey response to trump tweets


Donald J. Trump
@realDonaldTrump
Following

Little Adam Schiff, who is desperate to run for higher office, is one of the biggest liars and leakers in Washington, right up there with Comey, Warner, Brennan and Clapper! Adam leaves closed committee hearings to illegally leak confidential information. Must be stopped!

4:39 AM - 5 Feb 2018

31,930 Retweets 99,706 Likes

48K 32K 100K

Trump Tweets

`YouGov`: survey response to trump tweets

- Survey Equal # Republicans, Democrats, Independents: read Trump tweet + evaluate (`Great, Good, OK, Bad, Terrible`)

Trump Tweets

YouGov: survey response to trump tweets

- Survey Equal # Republicans, Democrats, Independents: read Trump tweet + evaluate (Great, Good, OK, Bad, Terrible)
- Aggregate, create scale $[-200, 200]$

Trump Tweets

YouGov: survey response to trump tweets

- Survey Equal # Republicans, Democrats, Independents: read Trump tweet + evaluate (Great, Good, OK, Bad, Terrible)
- Aggregate, create scale $[-200, 200]$
- Modify sIBP: Shared treatments, discover heterogeneous effects

Trump Tweets

YouGov: survey response to trump tweets

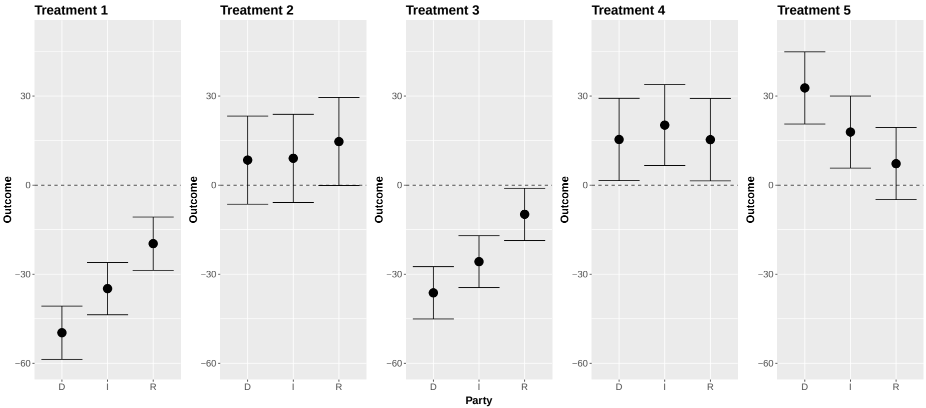
- Survey Equal # Republicans, Democrats, Independents: read Trump tweet + evaluate (Great, Good, OK, Bad, Terrible)
- Aggregate, create scale $[-200, 200]$
- Modify sIBP: Shared treatments, discover heterogeneous effects
- Train (66%), Test (33%), Clustered by tweet

Trump Tweets

YouGov: survey response to trump tweets

- Survey Equal # Republicans, Democrats, Independents: read Trump tweet + evaluate (Great, Good, OK, Bad, Terrible)
- Aggregate, create scale $[-200, 200]$
- Modify sIBP: Shared treatments, discover heterogeneous effects
- Train (66%), Test (33%), Clustered by tweet

Treatment 1	Treatment 2	Treatment 3	Treatment 4	Treatment 5
fake	cuts	obamacare	flotus	prime
news	strange	senators	behalf	minister
media	tax	repeal	anthem	korea
cnn	luther	healthcare	melania	north
election	stock	replace	nfl	stock
story	market	republican	flag	market
nbc	alabama	vote	prayers	china
stories	reform	republicans	bless	executive
hillary	record	senate	ready	prayers
clinton	high	north	players	order


Treatment 1
Treatment 2
Treatment 3
Treatment 4
Treatment 5
Outcome
30
0
−30
−60
D
I
R
Party

Sensitivity Analysis: analogous to residual plot in linear regression

Sensitivity Analysis: analogous to residual plot in linear regression
R Package: `textEffect`

Conclusions and Future Directions

- *Text as Data: How to Make Social Science Inferences Using Language* (Grimmer, Roberts, and Stewart In progress)

Conclusions and Future Directions

- *Text as Data: How to Make Social Science Inferences Using Language* (Grimmer, Roberts, and Stewart In progress)
  - Sequential (inductive) approach to social science

Conclusions and Future Directions

- *Text as Data: How to Make Social Science Inferences Using Language* (Grimmer, Roberts, and Stewart In progress)
    - Sequential (inductive) approach to social science
    - Build and refine theory with successive experiments

Conclusions and Future Directions

- *Text as Data: How to Make Social Science Inferences Using Language* (Grimmer, Roberts, and Stewart In progress)
  - Sequential (inductive) approach to social science
  - Build and refine theory with successive experiments
- Redesign of Survey Experiments, Observational Studies

Conclusions and Future Directions

- *Text as Data: How to Make Social Science Inferences Using Language* (Grimmer, Roberts, and Stewart In progress)
  - Sequential (inductive) approach to social science
  - Build and refine theory with successive experiments
- Redesign of Survey Experiments, Observational Studies
- Caution when combining machine learning and causal inference methods

Conclusions and Future Directions

- *Text as Data: How to Make Social Science Inferences Using Language* (Grimmer, Roberts, and Stewart In progress)
  - Sequential (inductive) approach to social science
  - Build and refine theory with successive experiments
- Redesign of Survey Experiments, Observational Studies
- Caution when combining machine learning and causal inference methods