Exploratory and Confirmatory Causal Inference for High Dimensional Interventions

Justin Grimmer
(Drawing from joint work with Christian Fong, Molly Roberts, and Brandon Stewart)

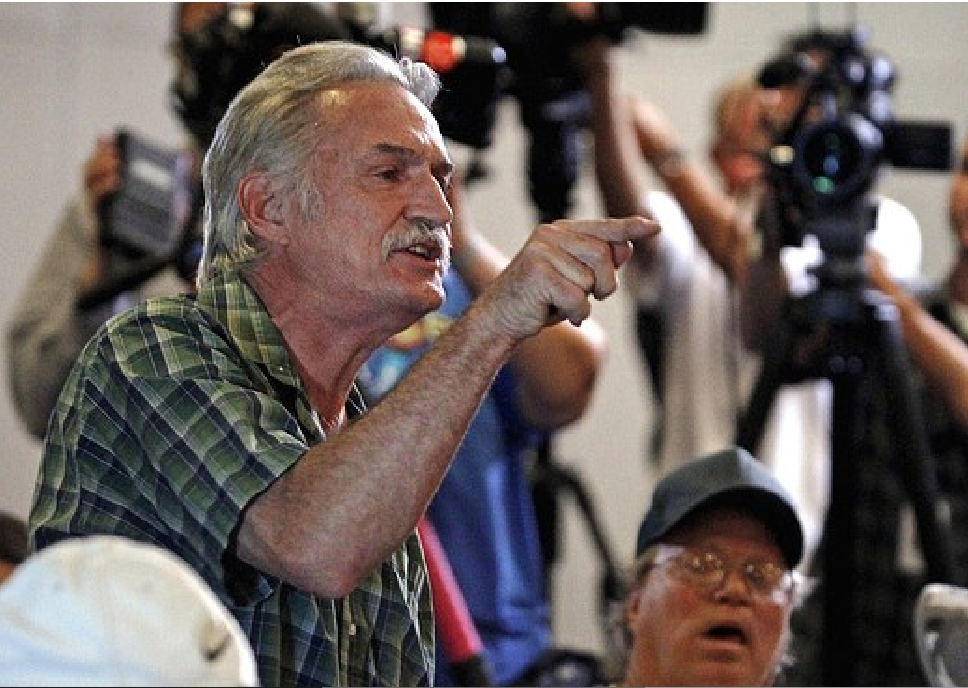
Stanford University $\rightsquigarrow$ University of Chicago

Causal Inference with Text Treatments

- How does X affect Y?

- How does X affect Y?
- X is often high-dimensional $\rightsquigarrow$ encoded in natural language


‘Brutal Anti-Cruz Attack Ad
Just 30 Seconds Of Candidate’s Photo
Displayed Without Any Text, Voiceover, Music’
The Onion




RESIST
RESIST
RESIST
RESIST


USA
Shelby Hall
Engineering and Computing Sciences


DANGER
RUN
WITH
HAND

Causal effect of text?

(Discover) Causal effect of text?

Candidate Backgrounds: Running Example for Paper

How do voters evaluate candidates?

- Candidates have policy positions and partisan affiliation
- Also have lives before elected office

What biographical facts affect voter evaluations?

Survey Experiment

Current Approaches: Text-Based Survey Experiments

Document 1: He earned his Juris Doctor in 1997 from Yale Law School, where he operated free legal clinics for low-income residents of New Haven, Connecticut.

Document 2: He served in South Vietnam from 1970 to 1971 during the Vietnam War in the Army Rangers' 75th Ranger Regiment, attached to the 173rd Airborne Brigade. He participated in 24 helicopter assaults...

Current Approaches: Text-Based Survey Experiments

Document 1: He earned his Juris Doctor in 1997 from Yale Law School, where he operated free legal clinics for low-income residents of New Haven, Connecticut.

Document 2: He served in South Vietnam from 1970 to 1971 during the Vietnam War in the Army Rangers' 75th Ranger Regiment, attached to the 173rd Airborne Brigade. He participated in 24 helicopter assaults...

Observe difference in evaluations of biographies

Current Approaches: Text-Based Survey Experiments

Document 1: He earned his Juris Doctor in 1997 from Yale Law School, where he operated free legal clinics for low-income residents of New Haven, Connecticut.

Document 2: He served in South Vietnam from 1970 to 1971 during the Vietnam War in the Army Rangers' 75th Ranger Regiment, attached to the 173rd Airborne Brigade. He participated in 24 helicopter assaults...

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

Current Approaches: Text-Based Survey Experiments

Condition 1:

[] majored in economics and political science at Miami University in Oxford, Ohio, where he became interested in the writings of Friedrich Hayek, Ludwig von Mises, and Milton Friedman...

Condition 2:

[] majored in economics and political science at Miami University in Oxford, Ohio, where he became interested in the writings of Karl Marx, Howard Zinn, and Noam Chomsky...

Current Approaches: Text-Based Survey Experiments

Condition 1:

```
[] majored in economics and political science at
Miami University in Oxford, Ohio, where he became
interested in the writings of Friedrich Hayek,
Ludwig von Mises, and Milton Friedman...
```

Condition 2:

```
[] majored in economics and political science at
Miami University in Oxford, Ohio, where he became
interested in the writings of Karl Marx, Howard
Zinn, and Noam Chomsky...
```

Design isolates the feature of the text (treatment) that drives the response

Current Approaches: Text-Based Survey Experiments

Condition 1:

`[] majored in economics and political science at Miami University in Oxford, Ohio, where he became interested in the writings of Friedrich Hayek, Ludwig von Mises, and Milton Friedman...`

Condition 2:

`[] majored in economics and political science at Miami University in Oxford, Ohio, where he became interested in the writings of Karl Marx, Howard Zinn, and Noam Chomsky...`

Design isolates the feature of the text (treatment) that drives the response $\rightsquigarrow$ What if you don't know interesting features (treatments) of text?

Current Approaches to Text-Based Survey Experiments

- "Interesting" treatments must be known in advance

Current Approaches to Text-Based Survey Experiments

- "Interesting" treatments must be known in advance
  - Occurs frequently in population of texts

Current Approaches to Text-Based Survey Experiments

- "Interesting" treatments must be known in advance
  - Occurs frequently in population of texts
  - Large effect on individuals' decisions

Current Approaches to Text-Based Survey Experiments

- "Interesting" treatments must be known in advance
  - Occurs frequently in population of texts
  - Large effect on individuals' decisions
- We invert current approach: we have population of texts and want to discover treatments

Current Approaches to Text-Based Survey Experiments

- "Interesting" treatments must be known in advance
  - Occurs frequently in population of texts
  - Large effect on individuals' decisions
- We invert current approach: we have population of texts and want to discover treatments

Automatically discover treatments

$+$

Estimate marginal effects

Three Key Steps

Three Key Steps

1) Prove: interventions at text level identify marginal effects of underlying treatments (Average Marginal Component Effect (AMCE) is identified)

Three Key Steps

1) Prove: interventions at text level identify marginal effects of underlying treatments (Average Marginal Component Effect (AMCE) is identified)

2) Method for discovering features (treatments)

Three Key Steps

1) Prove: interventions at text level identify marginal effects of underlying 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 AMCE

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

Identifying the AMCE

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

Identifying the AMCE

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

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

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

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

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

- An individual sees a text ($\boldsymbol{X}_{d[i]}$: text seen by i)
- Function: 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

Proposition 1

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_{d[i]}$: text seen by i)
- Function: 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:

Proposition 1

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_{d[i]}$: text seen by i)
- Function: 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: $Y_i(\boldsymbol{X}_N) = Y_i(\boldsymbol{X}_{d[i]})$)

Proposition 1

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_{d[i]}$: text seen by i)
- Function: 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: $Y_i(\boldsymbol{X}_N) = Y_i(\boldsymbol{X}_{d[i]})$)
2) Randomization or selection on observables:
   a) $Y_i(x) \perp\!\!\!\perp \boldsymbol{X}_{d[i]}$ and $\Pr(\boldsymbol{X}_{d[i]} = x) > 0$ for all $x \in \mathcal{X}$, or

Proposition 1

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_{d[i]}$: text seen by i)
- Function: 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: $Y_i(\boldsymbol{X}_N) = Y_i(\boldsymbol{X}_{d[i]})$)
2) Randomization or selection on observables:
   a) $Y_i(x) \perp\!\!\!\perp \boldsymbol{X}_{d[i]}$ and $\Pr(\boldsymbol{X}_{d[i]} = x) > 0$ for all $x \in \mathcal{X}$, or
   b) There exists a set of observed covariates, $\boldsymbol{C}$ such that $Y_i(x) \perp\!\!\!\perp \boldsymbol{X}_{d[i]} | \boldsymbol{C}_i$ and $\Pr(\boldsymbol{X}_{d[i]} = x | \boldsymbol{C}_i) > 0$ for all $x \in \mathcal{X}$.

Proposition 1

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_{d[i]}$: text seen by i)
- Function: 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: $Y_i(\boldsymbol{X}_N) = Y_i(\boldsymbol{X}_{d[i]})$)
2) Randomization or selection on observables:
   a) $Y_i(x) \perp\!\!\!\perp \boldsymbol{X}_{d[i]}$ and $\Pr(\boldsymbol{X}_{d[i]} = x) > 0$ for all $x \in \mathcal{X}$, or
3) Sufficiency: For all $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ such that $g(\boldsymbol{X}) = g(\boldsymbol{X}^{'})$ then $E[Y_i(g(\boldsymbol{X}))] = E[Y_i(g(\boldsymbol{X}^{'}))]$.

Proposition 1

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_{d[i]}$: text seen by i)
- Function: 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: $Y_i(\boldsymbol{X}_N) = Y_i(\boldsymbol{X}_{d[i]})$)
2) Randomization or selection on observables:
   a) $Y_i(x) \perp\!\!\!\perp \boldsymbol{X}_{d[i]}$ and $\Pr(\boldsymbol{X}_{d[i]} = x) > 0$ for all $x \in \mathcal{X}$, or
3) Sufficiency: For all $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ such that $g(\boldsymbol{X}) = g(\boldsymbol{X}^{'})$ then $E[Y_i(g(\boldsymbol{X}))] = E[Y_i(g(\boldsymbol{X}^{'}))]$.
4) Common support: all combinations of treatments have non-zero probability ($f(\boldsymbol{Z}_i) > 0$ for all $\boldsymbol{Z}_i \in \text{Range } g(\cdot)$)

Proposition 1

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_{d[i]}$: text seen by i)
- Function: 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: $Y_i(\boldsymbol{X}_N) = Y_i(\boldsymbol{X}_{d[i]})$)
2) Randomization or selection on observables:
   a) $Y_i(x) \perp\!\!\!\perp \boldsymbol{X}_{d[i]}$ and $\Pr(\boldsymbol{X}_{d[i]} = x) > 0$ for all $x \in \mathcal{X}$, or
3) Sufficiency: For all $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ such that $g(\boldsymbol{X}) = g(\boldsymbol{X}^{'})$ then $E[Y_i(g(\boldsymbol{X}))] = E[Y_i(g(\boldsymbol{X}^{'}))]$.
4) Common support: all combinations of treatments have non-zero probability ($f(\boldsymbol{Z}_i) > 0$ for all $\boldsymbol{Z}_i \in \text{Range } g(\cdot)$)

Proposition 1

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

Discovering Treatments and Estimating Marginal Effects

Discovery of Treatments from Text Corpora

Discovery of Treatments from Text Corpora

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

Discovery of Treatments from Text Corpora

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

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) (Randomly) divide texts and responses into training and test set

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) (Randomly) divide texts and responses into training and test set
    a) Avoid technical issues with using entire sample (Analyst-induced SUTVA violations)

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) (Randomly) divide texts and responses into training and test set
    a) Avoid technical issues with using entire sample (Analyst-induced SUTVA violations)
    b) Ensure we avoid “p-hacking" (false discovery)

Discovering Interesting Treatments

Discovering function from texts to treatments $g()$

Discovering Interesting Treatments

Discovering function from texts to treatments $g()$

- Use both documents and responses to discover the function

Discovering Interesting Treatments

Discovering function from texts to treatments $g()$

- Use both documents and responses to discover the function

- Topic and Supervised Topic models workhorse text models (Blei, Ng, and Jordan 2003; Blei and McAuliffe, 2007)

Discovering Interesting Treatments

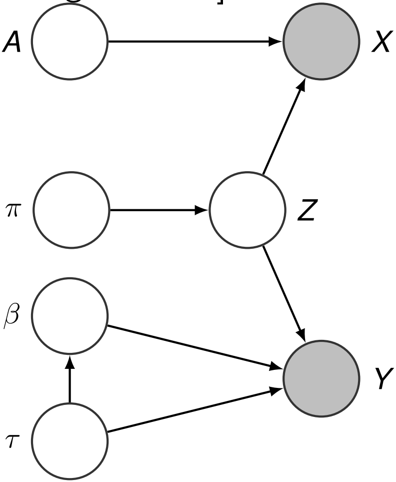
Discovering function from texts to treatments $g()$

- Use both documents and responses to discover the function

- Topic and Supervised Topic models workhorse text models (Blei, Ng, and Jordan 2003; Blei and McAuliffe, 2007)

Treatments on simplex imply marginalization impossible $\rightsquigarrow$ increase in one category implies decrease in other category

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



Text and response depend on latent treatments

- Treatment assignment

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

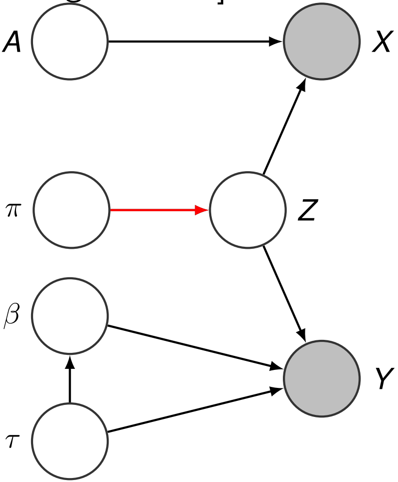
- Document Creation:

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

- Response:

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

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



Text and response depend on latent treatments

- Treatment assignment

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

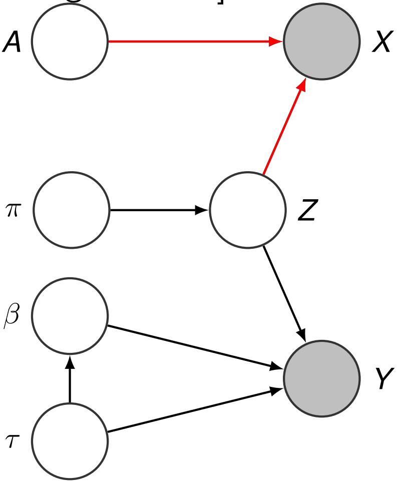
- Document Creation:

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

- Response:

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

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



Text and response depend on latent treatments

- Treatment assignment

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

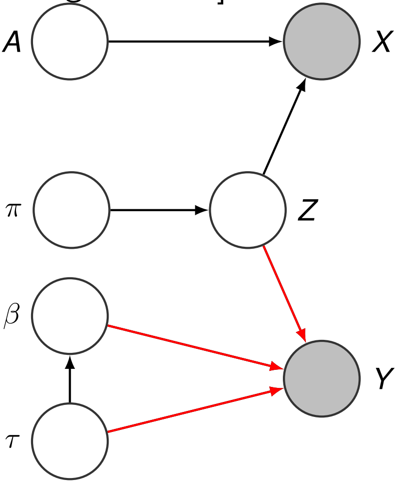
- Document Creation:

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

- Response:

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

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:

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

- Response:

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

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

Candidate Biographies on Wikipedia: Setup

Barbara Mikulski $\rightsquigarrow$ Barbara Schumacher

```
Schumacher was born and raised in the Highlandtown
neighborhood of East Baltimore, the eldest of the
three daughters of Christine Eleanor (nee Kutz) and
William Schumacher.  Her parents were both of Polish
descent; her immigrant great-grandparents had owned
a bakery in Baltimore.  During her high school years
at the Institute of Notre Dame, she worked in her
parents' grocery store...
```

Candidate Biographies on Wikipedia: Setup

Barbara Mikulski $\rightsquigarrow$ Barbara Schumacher

`Schumacher was born and raised in the Highlandtown neighborhood of East Baltimore, the eldest of the three daughters of Christine Eleanor (nee Kutz) and William Schumacher. Her parents were both of Polish descent; her immigrant great-grandparents had owned a bakery in Baltimore. During her high school years at the Institute of Notre Dame, she worked in her parents' grocery store...`

Protocol: For each respondent sees up to 3 texts from the corpus of > 2200 biographies

Candidate Biographies on Wikipedia: Setup

Barbara Mikulski $\rightsquigarrow$ Barbara Schumacher

`Schumacher was born and raised in the Highlandtown neighborhood of East Baltimore, the eldest of the three daughters of Christine Eleanor (nee Kutz) and William Schumacher. Her parents were both of Polish descent; her immigrant great-grandparents had owned a bakery in Baltimore. During her high school years at the Institute of Notre Dame, she worked in her parents' grocery store...`

Protocol: For each respondent sees up to 3 texts from the corpus of > 2200 biographies

- Observe text

Candidate Biographies on Wikipedia: Setup

Barbara Mikulski $\rightsquigarrow$ Barbara Schumacher

`Schumacher was born and raised in the Highlandtown neighborhood of East Baltimore, the eldest of the three daughters of Christine Eleanor (nee Kutz) and William Schumacher. Her parents were both of Polish descent; her immigrant great-grandparents had owned a bakery in Baltimore. During her high school years at the Institute of Notre Dame, she worked in her parents' grocery store...`

Protocol: For each respondent sees up to 3 texts from the corpus of > 2200 biographies

- Observe text
- Feeling thermometer rating: 0-100

Candidate Biographies on Wikipedia: Setup

Barbara Mikulski $\rightsquigarrow$ Barbara Schumacher

`Schumacher was born and raised in the Highlandtown neighborhood of East Baltimore, the eldest of the three daughters of Christine Eleanor (nee Kutz) and William Schumacher. Her parents were both of Polish descent; her immigrant great-grandparents had owned a bakery in Baltimore. During her high school years at the Institute of Notre Dame, she worked in her parents' grocery store...`

Protocol: For each respondent sees up to 3 texts from the corpus of > 2200 biographies

- Observe text
- Feeling thermometer rating: 0-100

1,886 participants, 5,303 responses

Candidate Biographies on Wikipedia: Setup

Barbara Mikulski $\rightsquigarrow$ Barbara Schumacher

```
Schumacher was born and raised in the Highlandtown
neighborhood of East Baltimore, the eldest of the
three daughters of Christine Eleanor (nee Kutz) and
William Schumacher.  Her parents were both of Polish
descent; her immigrant great-grandparents had owned
a bakery in Baltimore.  During her high school years
at the Institute of Notre Dame, she worked in her
parents' grocery store...
```

Protocol: For each respondent sees up to 3 texts from the corpus of > 2200 biographies

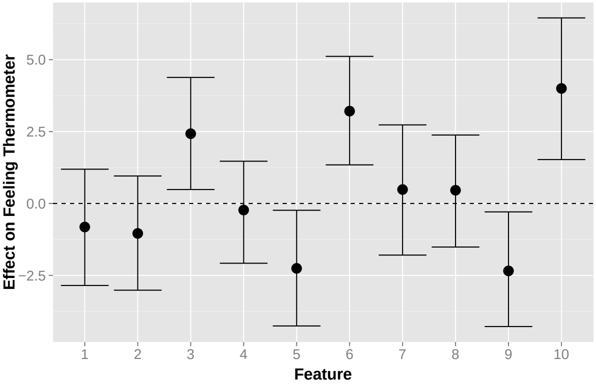
- Observe text
- Feeling thermometer rating: 0-100

1,886 participants, 5,303 responses

2,651 training, 2,652 test

Candiate Biographies on Wikipedia: Results

Treatment	Keywords
3	director, university, received, president, phd, policy
5	elected, house, democratic, seat
6	united_states, military, combat, rank
9	law, school_law, law_school, juris_doctor, student
10	war, enlisted, united_states, assigned, army



Consumer Financial Protection Bureau: Timely Response?

Consumer Financial Protection Bureau: Timely Response?

- Consumers log complaint about financial products

Consumer Financial Protection Bureau: Timely Response?

- Consumers log complaint about financial products

"The service representative was harsh and not listening to my questions. Attempting to collect on a debt I thought was in a grace period ...They were aggressive and unwilling to hear it."

Consumer Financial Protection Bureau: Timely Response?

- Consumers log complaint about financial products
- CFBP data: provides text of complaint and whether resolved "promptly"

Consumer Financial Protection Bureau: Timely Response?

- Consumers log complaint about financial products
- CFBP data: provides text of complaint and whether resolved “promptly"
- Plausible selection on observables: texts only what bureau has

Consumer Financial Protection Bureau: Timely Response?

- Consumers log complaint about financial products
- CFBP data: provides text of complaint and whether resolved "promptly"
- Plausible selection on observables: texts only what bureau has
- 113,424 total complaints

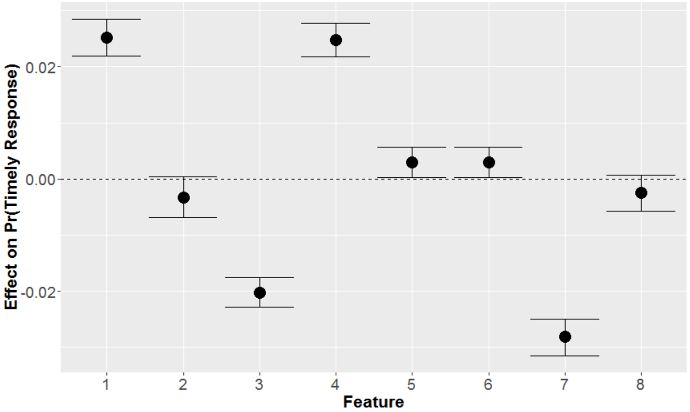
Consumer Financial Protection Bureau: Timely Response?

- Consumers log complaint about financial products
- CFBP data: provides text of complaint and whether resolved "promptly"
- Plausible selection on observables: texts only what bureau has
- 113,424 total complaints
- Train on 10%, Test on 90%

Consumer Financial Protection Bureau: Timely Response?

- Consumers log complaint about financial products
- CFBP data: provides text of complaint and whether resolved "promptly"
- Plausible selection on observables: texts only what bureau has
- 113,424 total complaints
- Train on 10%, Test on 90%

Treatment	Keywords
1	payment, card, debt , xxx , payment , loan
3	amount, call, account, time, pay, modification
4	interest, branch, number, xxxx_xxxx, told, house
7	month, credit_card, collection, received, called, loan_modificat



Conclusions and Future Directions

- Empirical Test of assumptions (sufficiency); Clarification of risks (Analyst SUTVA, Causal Effects with LDA)
- New treatment discovery methods
- Application to non-text settings (images, voting records)
- Text as Treatment ; Text as Outcome (STM, Roberts et al); Text as Outcome and Treatment