Text as Data

Justin Grimmer

Associate Professor
Department of Political Science
University of Chicago

August 24th, 2017

Discovery and Measurement

What is the research process? (Grimmer, Roberts, and Stewart 2017)

1) Discovery: a hypothesis or view of the world
2) Measurement according to some organization
3) Causal Inference: effect of some intervention

Text as data methods assist at each stage of research process

Causal Inference

- Text as Data + Social Science

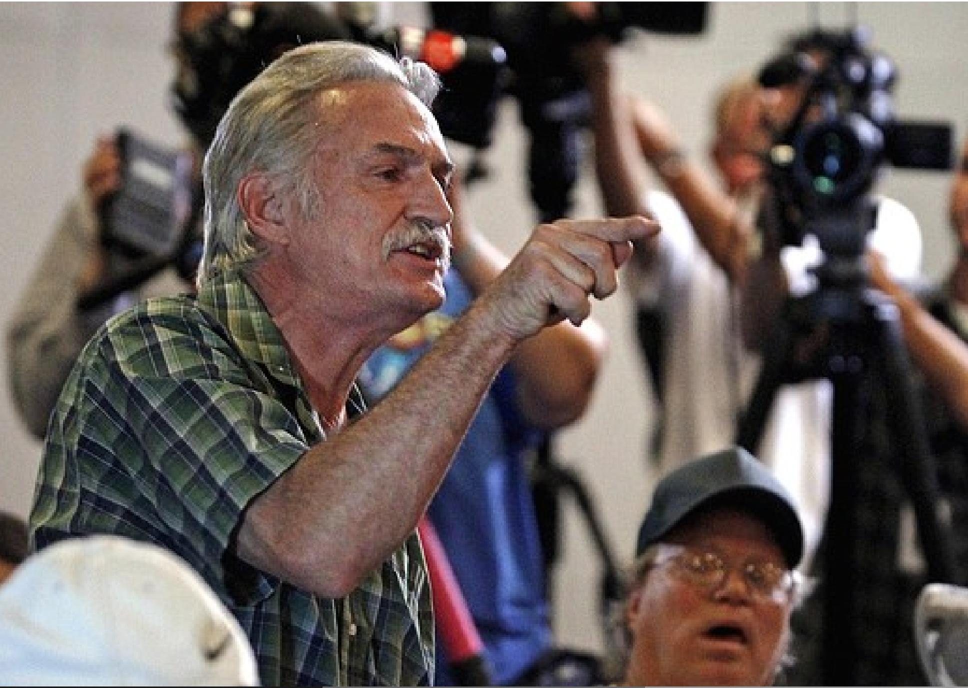
- Text as Data + Social Science
- Discovery: clustering, topic models,...

- Text as Data + Social Science
- Discovery: clustering, topic models,...
- Measurement: topic models, supervised classification, ...

- Text as Data + Social Science
- Discovery: clustering, topic models,...
- Measurement: topic models, supervised classification, ...
- Causal Inference

- Text as Data + Social Science
- Discovery: clustering, topic models,...
- Measurement: topic models, supervised classification, ...
- Causal Inference

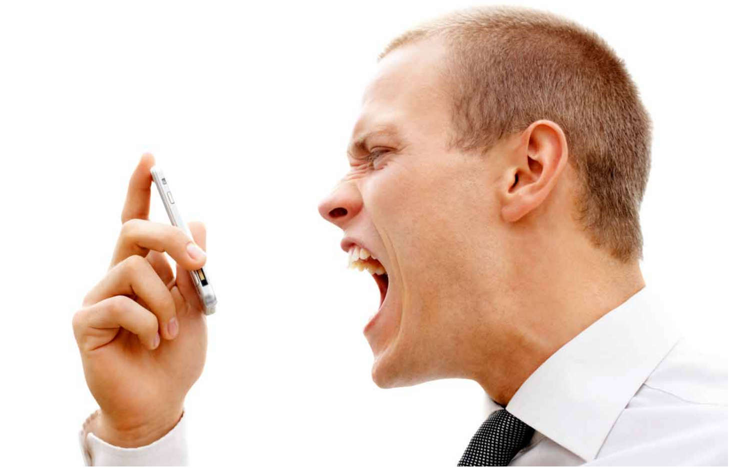
Discovery and Estimation




RESIST
RESIST
RESIST


USA
Shelby Hall
Engineering and Computing Sciences


DANGER
RUN
WITH




cfpb

Causal Inference in Text

Text as Intervention & Text as Response

1) Causal inference: latent representation of texts (g function to find latent features)
2) Discovery of features + Estimating effects$\rightsquigarrow$ train/test split

Which consumer complaints lead to a timely response?

Complaint A:
“I have been cheated by Wells Fargo! They were to set me up on an interest free payment plan, and I trusted them to do that. However, they set me up on a payment plan that took me way beyond the interest free date...Wells Fargo really sucks! I will avoid doing business with them in the future."

Complaint B:
“My name is XXXX XXXX. I am a Wells Fargo account holder. Wells Fargo illegally withdrew money from my account without notice or explanation"

Complaint A:
"I have been cheated by Wells Fargo! They were to set me up on an interest free payment plan, and I trusted them to do that. However, they set me up on a payment plan that took me way beyond the interest free date...Wells Fargo really sucks! I will avoid doing business with them in the future."

Complaint B:
"My name is XXXX XXXX. I am a Wells Fargo account holder. Wells Fargo illegally withdrew money from my account without notice or explanation"

Random assign A/B and assess response $\rightsquigarrow$ what about the complaint makes it better?

Complaint A:
"I have been cheated by Wells Fargo! They were to set me up on an interest free payment plan, and I trusted them to do that. However, they set me up on a payment plan that took me way beyond the interest free date...Wells Fargo really sucks! I will avoid doing business with them in the future."

Complaint A$'$:
"I have been cheated by Wells Fargo! They were to set me up on an interest free ~~payment~~ plan, and I trusted them to do that. However, they set me up on a ~~payment~~ plan that took me way beyond the interest free date...Wells Fargo really sucks! I will avoid doing business with them in the future."

Complaint A:
"I have been cheated by Wells Fargo! They were to set me up on an interest free payment plan, and I trusted them to do that. However, they set me up on a payment plan that took me way beyond the interest free date...Wells Fargo really sucks! I will avoid doing business with them in the future."

Complaint A':
"I have been cheated by Wells Fargo! They were to set me up on an interest free ~~payment~~ plan, and I trusted them to do that. However, they set me up on a ~~payment~~ plan that took me way beyond the interest free date...Wells Fargo really sucks! I will avoid doing business with them in the future."

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

Complaint A (Treatment 1):
"I have been cheated by Wells Fargo! They were to set me up on an interest free payment plan, and I trusted them to do that. However, they set me up on a payment plan that took me way beyond the interest free date...Wells Fargo really sucks! I will avoid doing business with them in the future."

Complaint $A^{'}$ (Treatment 0):
"I have been cheated by Wells Fargo! They were to set me up on an interest free payment plan, and I trusted them to do that. However, they set me up on a payment plan that took me way beyond the interest free date...I understand mistakes happen, I hope Wells Fargo can help improve their procedures in the future."

Complaint A (Treatment 1):
"I have been cheated by Wells Fargo! They were to set me up on an interest free payment plan, and I trusted them to do that. However, they set me up on a payment plan that took me way beyond the interest free date...Wells Fargo really sucks! I will avoid doing business with them in the future."

Complaint A' (Treatment 0):
"I have been cheated by Wells Fargo! They were to set me up on an interest free payment plan, and I trusted them to do that. However, they set me up on a payment plan that took me way beyond the interest free date...I understand mistakes happen, I hope Wells Fargo can help improve their procedures in the future."

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

Text-Based Intervention

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

Text-Based Intervention

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

Text-Based Intervention

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

Automatically discover treatments

$+$

Estimate marginal effects

Three Key Steps

Three Key Steps

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

Three Key Steps

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

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, holding other treatments constant

Identifying the Marginal Effects of Latent Treatments

- 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 Marginal Effects of Latent Treatments

- 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 Marginal Effects of Latent Treatments

- 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 Marginal Effects of Latent Treatments

- 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 Marginal Effects of Latent Treatments

- 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 Marginal Effects of Latent Treatments

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

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

Conjoint With Discovered Treatments

Identifying the Marginal Effects of Latent Treatments

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

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

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: 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: 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: 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: $Y_i(\boldsymbol{X}) = Y_i(\boldsymbol{X}_i)$)

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: 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: $Y_i(\boldsymbol{X}) = Y_i(\boldsymbol{X}_i)$)
2) Random assignment of texts ($Y_i(\boldsymbol{X}_i) \perp\!\!\!\perp \boldsymbol{X}_i$ for all i)

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: text seen by i)
- Function: 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: $Y_i(\boldsymbol{X}) = Y_i(\boldsymbol{X}_i)$)
2) Random assignment of texts ($Y_i(\boldsymbol{X}_i) \perp\!\!\!\perp \boldsymbol{X}_i$ for all i)
3) Sufficiency: For all $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ such that $g(\boldsymbol{X}) = g(\boldsymbol{X}^{'})$ then $E[Y_i(g(\boldsymbol{X}))] = E[Y_i(g(\boldsymbol{X}^{'}))]$.

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: 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: $Y_i(\boldsymbol{X}) = Y_i(\boldsymbol{X}_i)$)
2) Random assignment of texts ($Y_i(\boldsymbol{X}_i) \perp\!\!\!\perp \boldsymbol{X}_i$ for all i)
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 $\boldsymbol{Z}_i \in$ Range $g(\cdot)$)

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: 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: $Y_i(\boldsymbol{X}) = Y_i(\boldsymbol{X}_i)$)
2) Random assignment of texts ($Y_i(\boldsymbol{X}_i) \perp\!\!\!\perp \boldsymbol{X}_i$ for all i)
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 $\boldsymbol{Z}_i \in$ 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) (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 technical issues with using entire sample (Analyst-induced SUTVA violations)

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 technical issues with using entire sample (Analyst-induced SUTVA violations)
    b) Ensure we avoid "p-hacking" (false discovery)

Formal Definition of Analyst Induced SUTVA

Define g as dependent on other data:

Formal Definition of Analyst Induced SUTVA

Define g as dependent on other data:

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X}))$$

Formal Definition of Analyst Induced SUTVA

Define g as dependent on other data:

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X}))$$

Rerandomize ($\boldsymbol{X}^{'}$), but fix $\boldsymbol{X}_i$

Formal Definition of Analyst Induced SUTVA

Define g as dependent on other data:

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X}))$$

Rerandomize ($\boldsymbol{X}^{'}$), but fix $\boldsymbol{X}_i$

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X}^{'}))$$

Formal Definition of Analyst Induced SUTVA

Define g as dependent on other data:

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X}))$$

Rerandomize ($\boldsymbol{X}^{'}$), but fix $\boldsymbol{X}_i$

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X}^{'}))$$

analyst-level SUTVA holds if

Formal Definition of Analyst Induced SUTVA

Define g as dependent on other data:

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X}))$$

Rerandomize ($\boldsymbol{X}^{'}$), but fix $\boldsymbol{X}_i$

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X}^{'}))$$

analyst-level SUTVA holds if

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X})) = g(\boldsymbol{X}_i, Y(\boldsymbol{X}^{'}))$$

Formal Definition of Analyst Induced SUTVA

Define g as dependent on other data:

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X}))$$

Rerandomize ($\boldsymbol{X}^{'}$), but fix $\boldsymbol{X}_i$

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X}^{'}))$$

analyst-level SUTVA holds if

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X})) = g(\boldsymbol{X}_i, Y(\boldsymbol{X}^{'}))$$

Requires:

Formal Definition of Analyst Induced SUTVA

Define g as dependent on other data:

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X}))$$

Rerandomize $(\boldsymbol{X}^{'})$, but fix $\boldsymbol{X}_i$

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X}^{'}))$$

analyst-level SUTVA holds if

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X})) = g(\boldsymbol{X}_i, Y(\boldsymbol{X}^{'}))$$

Requires:

1) Category stability

Formal Definition of Analyst Induced SUTVA

Define g as dependent on other data:

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X}))$$

Rerandomize $(\boldsymbol{X}^{'})$, but fix $\boldsymbol{X}_i$

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X}^{'}))$$

analyst-level SUTVA holds if

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X})) = g(\boldsymbol{X}_i, Y(\boldsymbol{X}^{'}))$$

Requires:

1) Category stability $\rightsquigarrow$ same categories (retrained)
2) Classification Stability

Formal Definition of Analyst Induced SUTVA

Define g as dependent on other data:

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X}))$$

Rerandomize ($\boldsymbol{X}^{'}$), but fix $\boldsymbol{X}_i$

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X}^{'}))$$

analyst-level SUTVA holds if

$$g(\boldsymbol{X}_i, Y(\boldsymbol{X})) = g(\boldsymbol{X}_i, Y(\boldsymbol{X}^{'}))$$

Requires:

1) Category stability $\rightsquigarrow$ same categories (retrained)
2) Classification Stability $\rightsquigarrow$ same text leads to same label (retrained)

Train/Test split ensures both hold.

Why Train-Test Splits?

Why Train-Test Splits?

Train-Test splits are on the rise in causal inference (e.g. Athey and Wager, Chernozhukov et al.) and political science (Cranmer and Desmarais).

Why Train-Test Splits?

Train-Test splits are on the rise in causal inference (e.g. Athey and Wager, Chernozhukov et al.) and political science (Cranmer and Desmarais).

We like that they build in discovery and connect with a general theory of experiments.

Why Train-Test Splits?

Train-Test splits are on the rise in causal inference (e.g. Athey and Wager, Chernozhukov et al.) and political science (Cranmer and Desmarais).

We like that they build in discovery and connect with a general theory of experiments.

Why not Pre-Analysis Plans (PAP)?

Why Train-Test Splits?

Train-Test splits are on the rise in causal inference (e.g. Athey and Wager, Chernozhukov et al.) and political science (Cranmer and Desmarais).

We like that they build in discovery and connect with a general theory of experiments.

Why not Pre-Analysis Plans (PAP)?

- When possible, we advocate sequential experiments.

Why Train-Test Splits?

Train-Test splits are on the rise in causal inference (e.g. Athey and Wager, Chernozhukov et al.) and political science (Cranmer and Desmarais).

We like that they build in discovery and connect with a general theory of experiments.

Why not Pre-Analysis Plans (PAP)?

- When possible, we advocate sequential experiments.
- How would we design experiments to ensure we replicate?

Why Train-Test Splits?

Train-Test splits are on the rise in causal inference (e.g. Athey and Wager, Chernozhukov et al.) and political science (Cranmer and Desmarais).

We like that they build in discovery and connect with a general theory of experiments.

Why not Pre-Analysis Plans (PAP)?

- When possible, we advocate sequential experiments.
- How would we design experiments to ensure we replicate?
  Train-test Split: it simulates a 'fresh' test and let's us learn from the first round data.

Why Train-Test Splits?

Train-Test splits are on the rise in causal inference (e.g. Athey and Wager, Chernozhukov et al.) and political science (Cranmer and Desmarais).

We like that they build in discovery and connect with a general theory of experiments.

Why not Pre-Analysis Plans (PAP)?

- When possible, we advocate sequential experiments.
- How would we design experiments to ensure we replicate?
  Train-test Split: it simulates a 'fresh' test and let's us learn from the first round data.
- regulation $\rightsquigarrow$ Pre-analysis plans;

Why Train-Test Splits?

Train-Test splits are on the rise in causal inference (e.g. Athey and Wager, Chernozhukov et al.) and political science (Cranmer and Desmarais).

We like that they build in discovery and connect with a general theory of experiments.

Why not Pre-Analysis Plans (PAP)?

- When possible, we advocate sequential experiments.
- How would we design experiments to ensure we replicate?
  Train-test Split: it simulates a 'fresh' test and let's us learn from the first round data.
- regulation $\rightsquigarrow$ Pre-analysis plans;
  incentives $\rightsquigarrow$ Train-Test splits

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)

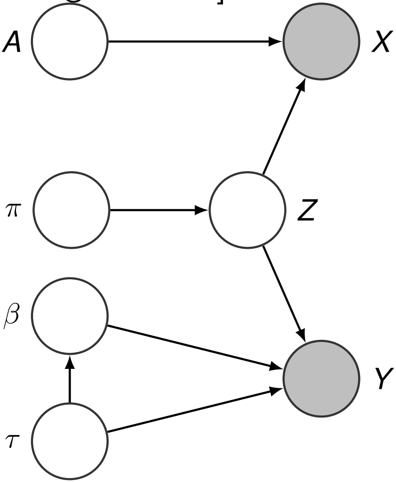
Discovering Interesting Treatments

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

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

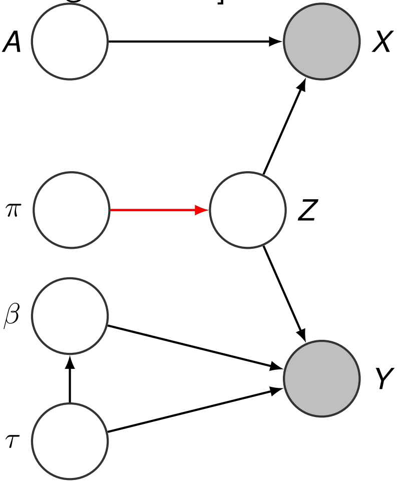
- Document Creation:

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

- 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

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

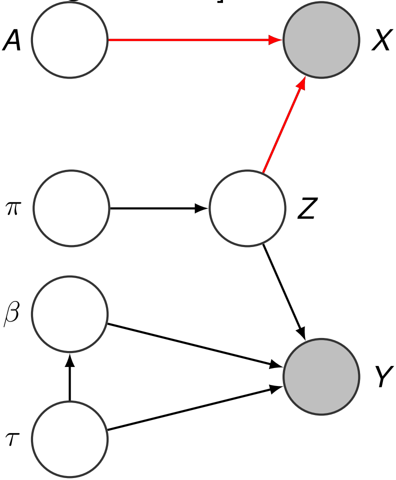
- Document Creation:

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

- Response:

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

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



- 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:

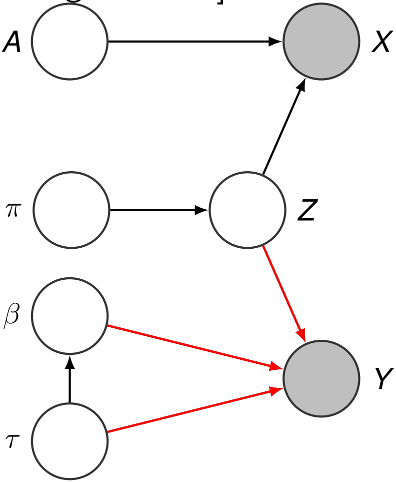
$$\boldsymbol{X}_i \sim \text{MVN}(\boldsymbol{Z}_i \boldsymbol{A}, \sigma_X^2 I_D)$$
$$\boldsymbol{A}_k \sim \text{MVN}(\boldsymbol{0}, \sigma_A^2 I_D)$$

- Response:

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

Text and response depend on latent treatments

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

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

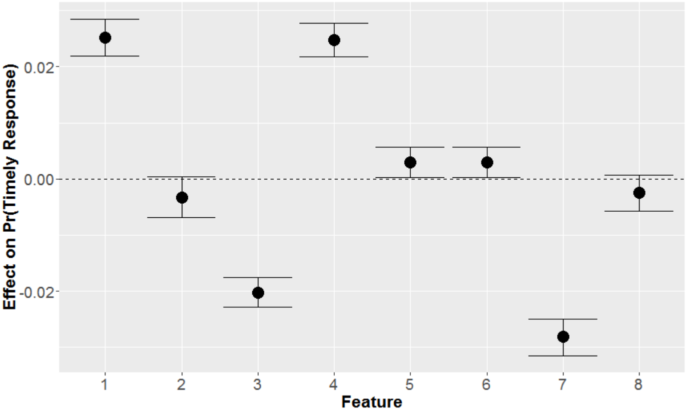
Consumer Financial Protection Bureau: Timely Response?

- 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



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

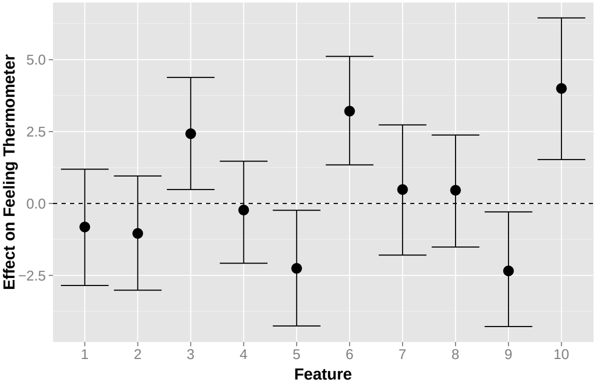
- Observe text
- Feeling thermometer rating: 0-100

1,886 participants, 5,303 responses

2,651 training, 2,652 test

Candiate Biographies on Wikipedia: Results

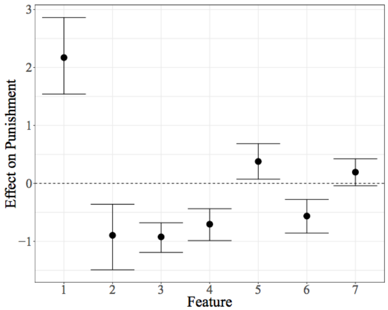
Treatment	Keywords
3	director, university, received, president, phd, policy
5	elected, house, democratic, seat
6	united_states, military, combat, rank
9	law, school_law, law_school, juris_doctor, student
10	war, enlisted, united_states, assigned, army



Two Examples from Jane Esberg (2017)

1.1 Spain Political Prisoner Trials

Text data: Random sample of 1,800 (out of 3,900) Spanish Tribunal of Public Order criminal summaries (1964-1975).
Outcome: Trial decision (punishment in years).



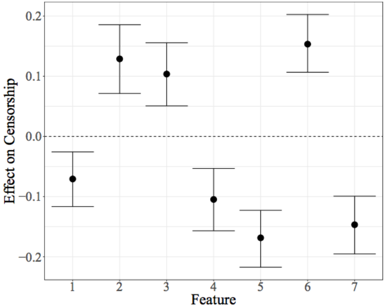
Feature			
1	2	3	4
crime	delinquent	drunk	accusation
adult	francoist	boss	disorder
responsible	youth	disturbance	pretend
author	policy	alcohol	financial
legal	subversive	yelling	traffic
5	6	7	
communist	gun	illicit	
party	possession	licence	
propaganda	pistol	revolver	
organization	fire	milimeter	
violence	munition	belonging	

Figure 1: *Trial summary features and effect on punishment in Spain*

Two Examples from Jane Esberg (2017)

1.2 Chile Movie Censorship

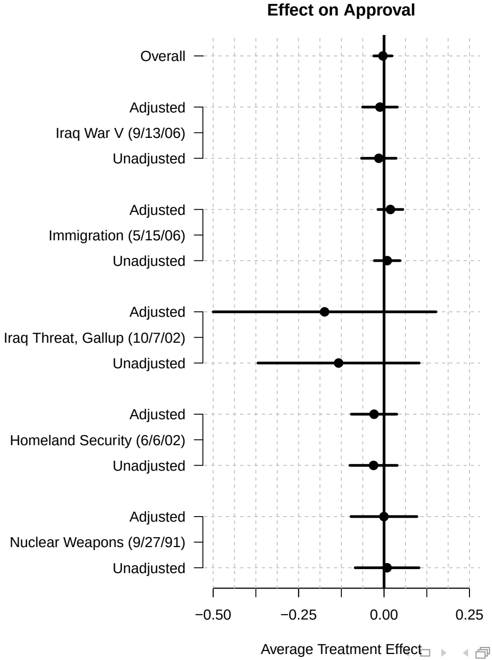
Text data: IMDb keywords for the 6,000 movies reviewed under Chile's dictatorship.
Outcome: Whether a film was banned (0/1).



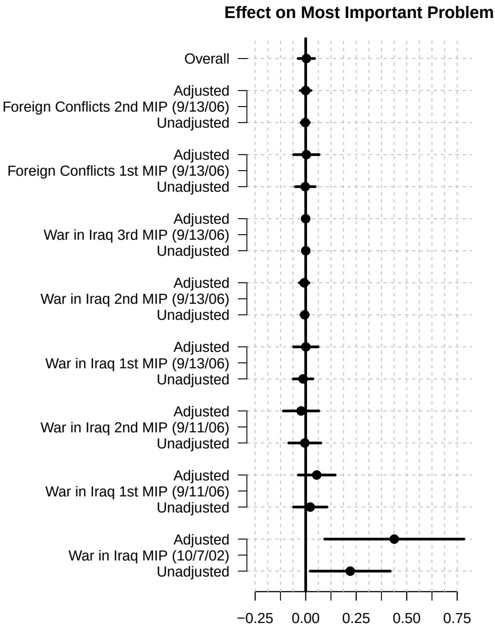
Feature			
1	2	3	4
singing	lust	death	rifle
photograph	cleavage	blood	cowboy
drinking	nudity	cruelty	gunfighter
relationship	erotic	corpse	saloon
tears	mini skirt	knife	battle
5	6	7	
hero	voyeur	tough guy	
showdown	undressing	quick draw	
fistfight	peeping tom	explosion	
martial arts	scantily clad	warrior	
ambush	nudity	lone	

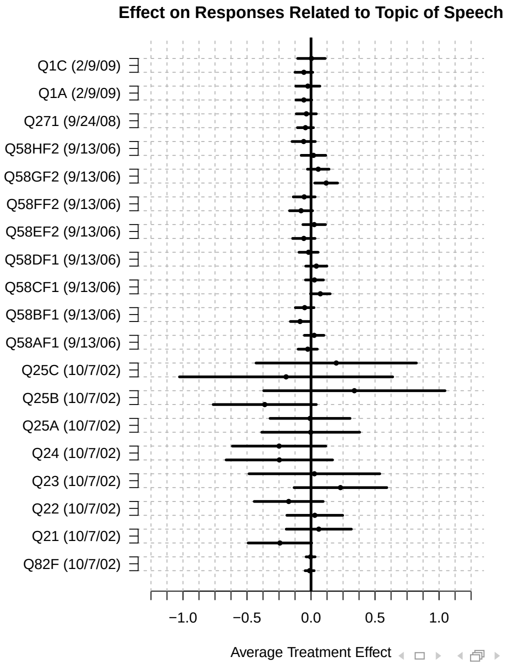
Figure 2: *Movie features and effect on censorship in Chile*

How do presidents "going public" affect public opinion?

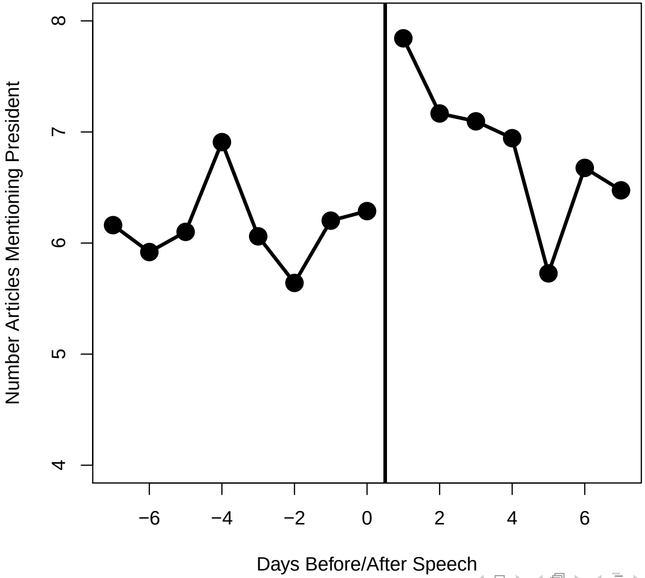


Effect on Approval
Overall
Adjusted
Iraq War V (9/13/06)
Unadjusted
Adjusted
Immigration (5/15/06)
Unadjusted
Adjusted
Iraq Threat, Gallup (10/7/02)
Unadjusted
Adjusted
Homeland Security (6/6/02)
Unadjusted
Adjusted
Nuclear Weapons (9/27/91)
Unadjusted
−0.50
−0.25
0.00
0.25
Average Treatment Effect


Effect on Most Important Problem
Overall
Adjusted
Foreign Conflicts 2nd MIP (9/13/06)
Unadjusted
Adjusted
Foreign Conflicts 1st MIP (9/13/06)
Unadjusted
Adjusted
War in Iraq 3rd MIP (9/13/06)
Unadjusted
Adjusted
War in Iraq 2nd MIP (9/13/06)
Unadjusted
Adjusted
War in Iraq 1st MIP (9/13/06)
Unadjusted
Adjusted
War in Iraq 2nd MIP (9/11/06)
Unadjusted
Adjusted
War in Iraq 1st MIP (9/11/06)
Unadjusted
Adjusted
War in Iraq MIP (10/7/02)
Unadjusted
−0.25
0.00
0.25
0.50
0.75


Effect on Responses Related to Topic of Speech
Q1C (2/9/09)
Q1A (2/9/09)
Q271 (9/24/08)
Q58HF2 (9/13/06)
Q58GF2 (9/13/06)
Q58FF2 (9/13/06)
Q58EF2 (9/13/06)
Q58DF1 (9/13/06)
Q58CF1 (9/13/06)
Q58BF1 (9/13/06)
Q58AF1 (9/13/06)
Q25C (10/7/02)
Q25B (10/7/02)
Q25A (10/7/02)
Q24 (10/7/02)
Q23 (10/7/02)
Q22 (10/7/02)
Q21 (10/7/02)
Q82F (10/7/02)
−1.0
−0.5
0.0
0.5
1.0
Average Treatment Effect

How do presidents "going public" affect ~~public opinion~~ the media agenda?


Number Articles Mentioning President
4
5
6
7
8
−6
−4
−2
0
2
4
6
Days Before/After Speech

1) (Assume) random assignment of treatments

1) (Assume) random assignment of treatments

2) Obtain text based response $\boldsymbol{Y}_i(T_i)$

1) (Assume) random assignment of treatments
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$

Function g now uncovers latent features of response: map from text to small number of categories

1) (Assume) random assignment of treatments
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$

Function $\boldsymbol{g}$ now uncovers latent features of response: map from text to small number of categories

$$\text{ATE}_k = \text{E}[g(\boldsymbol{Y}(1))_k - g(\boldsymbol{Y}(0))_k]$$

Discovering (Estimating) Dependent Variable

- Numerous options to discover: (hand coding, supervised models, unsupervised models, mixture)

Discovering (Estimating) Dependent Variable

- Numerous options to discover: (hand coding, supervised models, unsupervised models, mixture)
- All have same worries: (1) Analyst Induced SUTVA violation (2) Fishing

Discovering (Estimating) Dependent Variable

- Numerous options to discover: (hand coding, supervised models, unsupervised models, mixture)
- All have same worries: (1) Analyst Induced SUTVA violation (2) Fishing

Train/Test Split

1) (Assume) random assignment of treatments

2) Obtain text based response $\boldsymbol{Y}_i(T_i)$

1) (Assume) random assignment of treatments
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$
3) Randomly split response and text into train/test split

1) (Assume) random assignment of treatments
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$
3) Randomly split response and text into train/test split
4) In training set: discover latent dependent variables

1) (Assume) random assignment of treatments
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$
3) Randomly split response and text into train/test split
4) In training set: discover latent dependent variables
    a) Apply Structural Topic Model (Roberts, Stewart, and Airoldi 2017)

1) (Assume) random assignment of treatments
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$
3) Randomly split response and text into train/test split
4) In training set: discover latent dependent variables
    a) Apply Structural Topic Model (Roberts, Stewart, and Airoldi 2017)
    b) Make final model pick based on quantitative model fit and exploration

1) (Assume) random assignment of treatments
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$
3) Randomly split response and text into train/test split
4) In training set: discover latent dependent variables
    a) Apply Structural Topic Model (Roberts, Stewart, and Airoldi 2017)
    b) Make final model pick based on quantitative model fit and exploration
5) In test set:

1) (Assume) random assignment of treatments
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$
3) Randomly split response and text into train/test split
4) In training set: discover latent dependent variables
    a) Apply Structural Topic Model (Roberts, Stewart, and Airoldi 2017)
    b) Make final model pick based on quantitative model fit and exploration
5) In test set:
    a) Infer dependent variables (using newly available updates to `STM` software (Roberts, Stewart, and Tingley 2017))

1) (Assume) random assignment of treatments
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$
3) Randomly split response and text into train/test split
4) In training set: discover latent dependent variables
    a) Apply Structural Topic Model (Roberts, Stewart, and Airoldi 2017)
    b) Make final model pick based on quantitative model fit and exploration
5) In test set:
    a) Infer dependent variables (using newly available updates to `STM` software (Roberts, Stewart, and Tingley 2017))
    b) Estimate effect of treatments on topic prevalence across categories

A President's effect on newspaper agenda

A President's effect on newspaper agenda

- Response: newspaper articles mentioning `president` in 10 highest circulation papers, two-week window around speech

A President's effect on newspaper agenda

- Response: newspaper articles mentioning `president` in 10 highest circulation papers, two-week window around speech
- Treatment: Number of days before/after speech article was published

A President's effect on newspaper agenda

- Response: newspaper articles mentioning `president` in 10 highest circulation papers, two-week window around speech
- Treatment: Number of days before/after speech article was published
- 159,217 articles

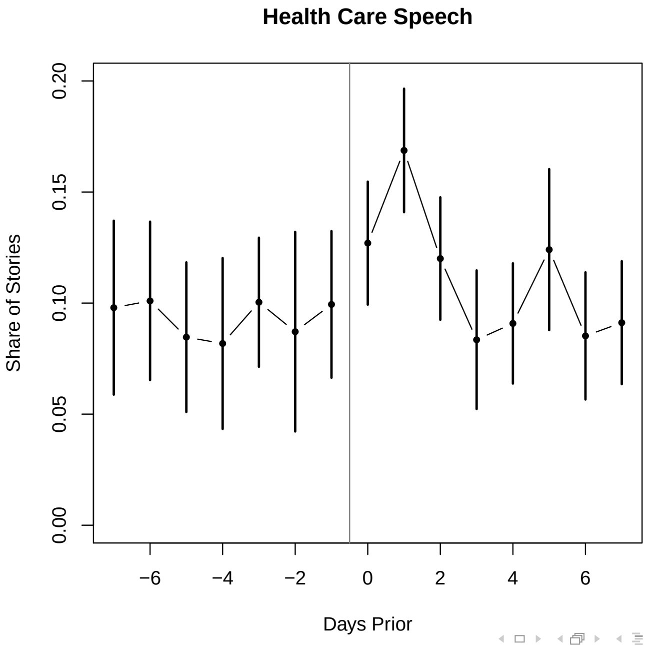
A President's effect on newspaper agenda

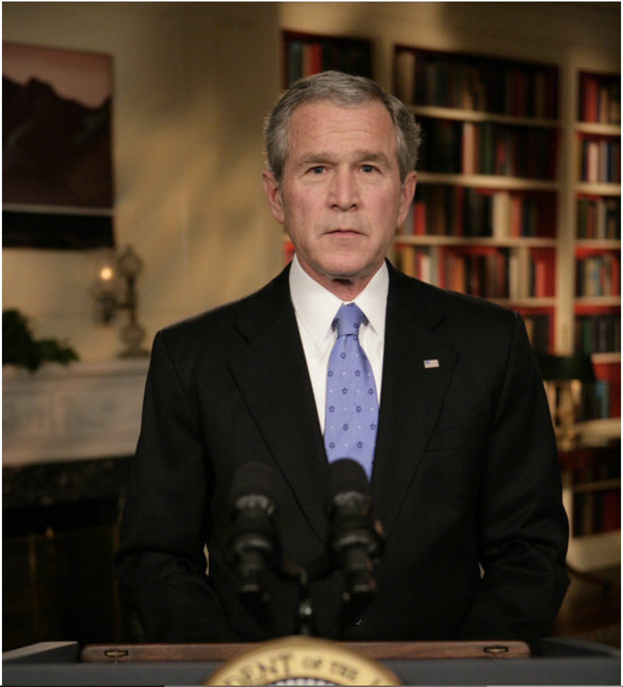
- Response: newspaper articles mentioning `president` in 10 highest circulation papers, two-week window around speech
- Treatment: Number of days before/after speech article was published
- 159,217 articles
- Train: 10%, Test 90%

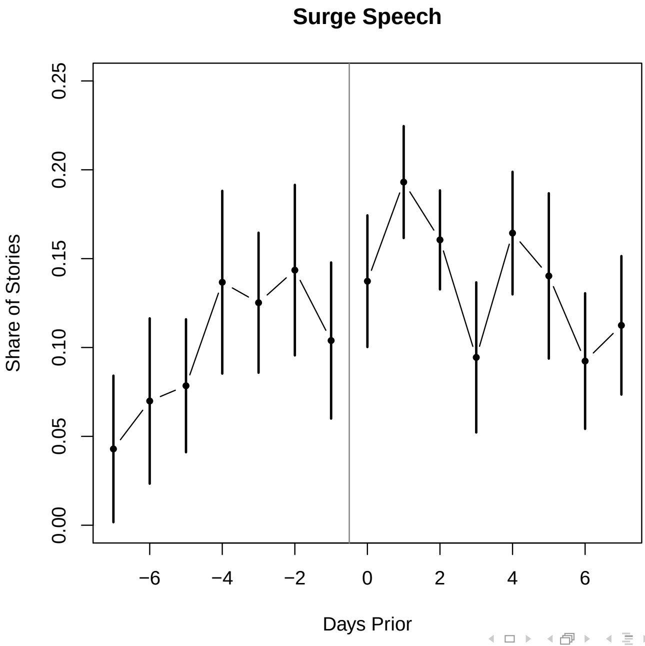
A President's effect on newspaper agenda

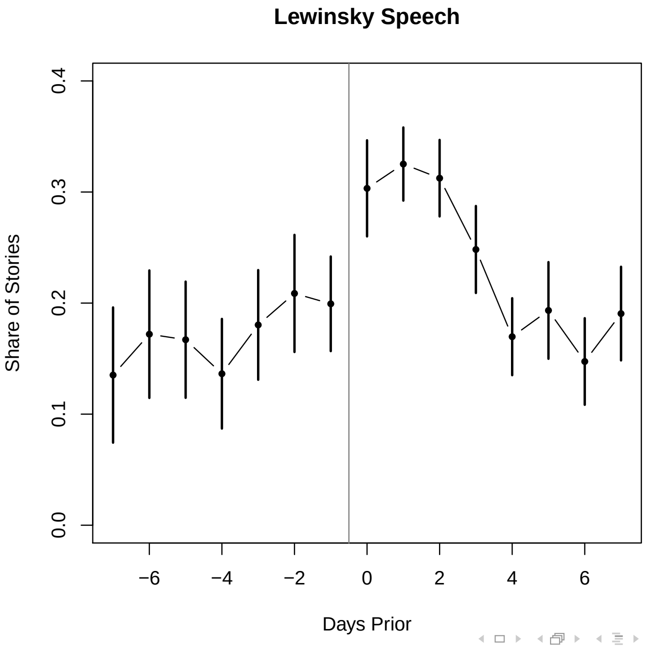
- Response: newspaper articles mentioning `president` in 10 highest circulation papers, two-week window around speech
- Treatment: Number of days before/after speech article was published
- 159,217 articles
- Train: 10%, Test 90%
- Effect estimate: interrupted time series design on topic prevalence (compare share immediately before to share immediately after)

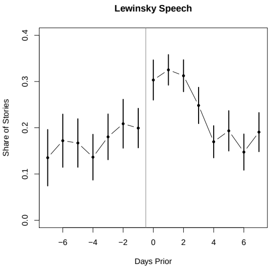

gettyimages
Pool
90534940


Health Care Speech
Share of Stories
0.00
0.05
0.10
0.15
0.20
−6
−4
−2
0
2
4
6
Days Prior




Surge Speech
Share of Stories
0.00
0.05
0.10
0.15
0.20
0.25
−6
−4
−2
0
2
4
6
Days Prior


Lewinsky Speech
Share of Stories
0.0
0.1
0.2
0.3
0.4
−6
−4
−2
0
2
4
6
Days Prior



Across speeches $\rightsquigarrow$ consistent effect on agenda

Immigration Application

Immigration Application

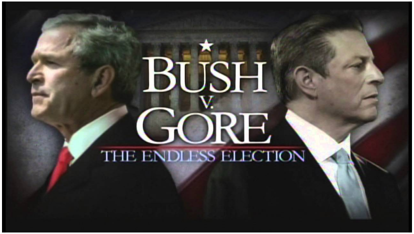
- Example application on a survey experiment about attitudes toward immigration.

Immigration Application

- Example application on a survey experiment about attitudes toward immigration.
- Uses data from a study by Cohen, Rust and Steen (2004), telephone random-digit dial of 1300 respondents (conducted in 2000). Train: 50%, Test 50%.

Immigration Application

- Example application on a survey experiment about attitudes toward immigration.
- Uses data from a study by Cohen, Rust and Steen (2004), telephone random-digit dial of 1300 respondents (conducted in 2000). Train: 50%, Test 50%.



Immigration Application

- Example application on a survey experiment about attitudes toward immigration.
- Uses data from a study by Cohen, Rust and Steen (2004), telephone random-digit dial of 1300 respondents (conducted in 2000). Train: 50%, Test 50%.

Immigration Application

- Example application on a survey experiment about attitudes toward immigration.
- Uses data from a study by Cohen, Rust and Steen (2004), telephone random-digit dial of 1300 respondents (conducted in 2000). Train: 50%, Test 50%.
- Respondents given a prompt about an immigrant, asked if she should be imprisoned and why.

Immigration Application

- Example application on a survey experiment about attitudes toward immigration.
- Uses data from a study by Cohen, Rust and Steen (2004), telephone random-digit dial of 1300 respondents (conducted in 2000). Train: 50%, Test 50%.
- Respondents given a prompt about an immigrant, asked if she should be imprisoned and why.

"A 28-year-old single man, a citizen of another country, was convicted of illegally entering the United States. Prior to this offense, he had served two previous prison sentences each more than a year. One of these previous sentences was for a violent crime and he had been deported back to his home country."

Immigration Application

- Example application on a survey experiment about attitudes toward immigration.
- Uses data from a study by Cohen, Rust and Steen (2004), telephone random-digit dial of 1300 respondents (conducted in 2000). Train: 50%, Test 50%.
- Respondents given a prompt about an immigrant, asked if she should be imprisoned and why.

"A 28-year-old single man, a citizen of another country, was convicted of illegally entering the United States. Prior to this offense, he had never been imprisoned before."

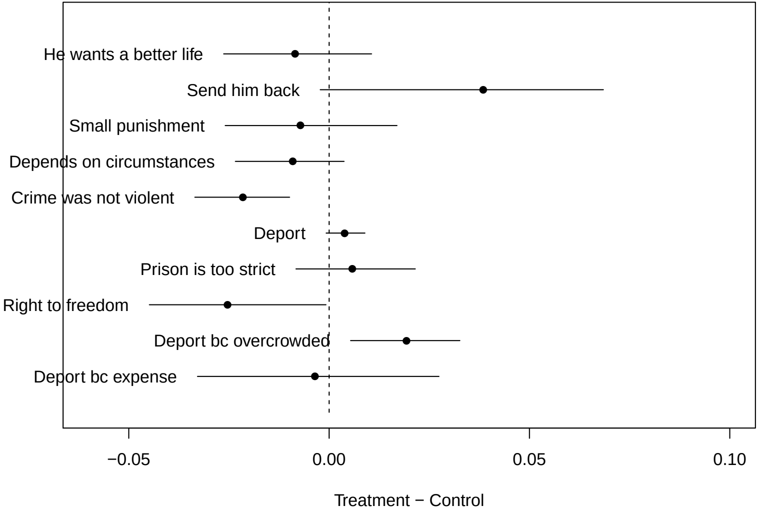
Immigration Experiment Results

Label	Highest Probability Words
He wants a better life	didnt, want, pay, better, life, probabl, isnt
Send him back	back, countri, send, home, well, charg
Small punishment	offens, reason, like, chanc, first, can, citizen
Depends on circum.	come, depend, doesnt, free, feel, law
Crime was not violent	crime, commit, violent, immigr, wasnt, look
Deport	deport, that, give, counti, peopl, look, guilti
Prison is too strict	enter, anyth, right, live, realli, illeg, anybodi
Right to freedom	just, tri, get, hes, came, freedom, put
Deport bc overcrowded	sent, prison, think, alreadi, anoth, done
Deport bc expense	dont, think, know, time, need, serv, crimin

Immigration Experiment Results

Label	Representative Document
He wants a better life	we're the land of opportunity everybody wants a better life
Send him back	send him back to his country
Small punishment	"it was his first offense, didn't hurt anybody, maybe a fine though, probation or something. that's nice serious like murder or robbery"
Depends on circumstances	it depends on reaason why he is coming into state if he was coming to beter himself its ok if he has a record he should be disbarred or deported
Crime was not violent	because he didnt commit a crime that was effecting someone else's individual liberties
Deport	he should be deported
Prison is too strict	because he didnt do anything except illegally enter
Right to freedom	Because he's just trying to get his freedom. Maybe he's trying to away from a tough situation/that country- maybe it's not good for him.
Deport bc overcrowded	he should be sent to prison in another country our prisons are over crowded already
Deport bc expense	because i think he shold be deported-p-i don't think he should be supported in our prison system and i don't think he should be allowed to immigrate here

Immigration Experiment Results



Conclusions and Future Directions

- Sequential (inductive) approach to social science: build theory with successive experiments
- Testing assumptions and new causal quantities of interest
- General Framework: Application to non-text settings (images, voting records)
- Text as Treatment , Text as Outcome , Text as Outcome and Treatment