Matching Methods for High-Dimensional Data with Applications to Text

Molly Roberts, Brandon Stewart and Rich Nielsen

UCSD, Princeton and MIT

May 11, 2016
How do people react to online repression?

Lots of governments try to control online information.

Censoring the whole internet is hard (# of bloggers $\gg$ # of censors).

Limited external enforcement $\Rightarrow$ governments scare people into self-policing.

Governments might jail some bloggers to scare people.

Then encourage self-censorship by signaling off-limits topics.
How do people react to online repression?

- Lots of governments try to control online information
Introduction

How do people react to online repression?

- Lots of governments try to control online information
- Censoring the whole internet is hard (# of bloggers $\gg$ # of censors)
How do people react to online repression?

- Lots of governments try to control online information
- Censoring the whole internet is hard (\# of bloggers $\gg$ \# of censors)
- Limited external enforcement $\leadsto$ govt's scare people into self-policing
How do people react to online repression?

- Lots of governments try to control online information
- Censoring the whole internet is hard (# of bloggers $\gg$ # of censors)
- Limited external enforcement $\Rightarrow$ gov’ts scare people into self-policing
How do people react to online repression?

- Lots of governments try to control online information
- Censoring the whole internet is hard (number of bloggers $\gg$ number of censors)
- Limited external enforcement $\Rightarrow$ governments scare people into self-policing

- Governments might jail some bloggers to scare people
How do people react to online repression?

- Lots of governments try to control online information
- Censoring the whole internet is hard (number of bloggers $\gg$ number of censors)
- Limited external enforcement $\Rightarrow$ gov'ts scare people into self-policing

- Governments might jail some bloggers to scare people
- Then encourage self-censorship by signaling off-limits topics
How do people react to online repression?

- Lots of governments try to control online information
- Censoring the whole internet is hard (number of bloggers $\gg$ number of censors)
- Limited external enforcement $\leadsto$ governments scare people into self-policing

- Governments might jail some bloggers to scare people
- Then encourage self-censorship by signaling off-limits topics

Roberts (UCSD)
How do people react to online repression?

- Lots of governments try to control online information
- Censoring the whole internet is hard (# of bloggers $\gg$ # of censors)
- Limited external enforcement $\Rightarrow$ gov’ts scare people into self-policing

- Governments might jail some bloggers to scare people
- Then encourage self-censorship by signaling off-limits topics
- But this could turn out one of two ways:
How do people react to online repression?

- Lots of governments try to control online information
- Censoring the whole internet is hard (\# of bloggers $\gg$ \# of censors)
- Limited external enforcement $\Rightarrow$ gov’ts scare people into self-policing

- Governments might jail some bloggers to scare people
- Then encourage self-censorship by signaling off-limits topics
- But this could turn out one of two ways:
  - Bloggers might take cues to self-censor OR
How do people react to online repression?

- Lots of governments try to control online information
- Censoring the whole internet is hard (# of bloggers ≫ # of censors)
- Limited external enforcement ➝ gov’ts scare people into self-policing
  - Governments might jail some bloggers to scare people
  - Then encourage self-censorship by signaling off-limits topics
  - But this could turn out one of two ways:
    - Bloggers might take cues to self-censor OR
    - Bloggers might hate censorship and rebel
How do people react to online repression?

- Lots of governments try to control online information
- Censoring the whole internet is hard (# of bloggers ≫ # of censors)
- Limited external enforcement ⟷ gov’ts scare people into self-policing
  - Governments might jail some bloggers to scare people
  - Then encourage self-censorship by signaling off-limits topics
  - But this could turn out one of two ways:
    - Bloggers might take cues to self-censor OR
    - Bloggers might hate censorship and rebel
  - how censorship works: important to understand self-censorship
How do people react to online repression?

- Lots of governments try to control online information
- Censoring the whole internet is hard (# of bloggers ≫ # of censors)
- Limited external enforcement ⟷ gov’ts scare people into self-policing

- Governments might jail some bloggers to scare people
- Then encourage self-censorship by signaling off-limits topics
- But this could turn out one of two ways:
  - Bloggers might take cues to self-censor OR
  - Bloggers might hate censorship and rebel
- how censorship works: important to understand self-censorship
- BUT self-censorship very hard to measure
How do people react to online repression?

- Lots of governments try to control online information
- Censoring the whole internet is hard (\# of bloggers ≫ \# of censors)
- Limited external enforcement \(\leadsto\) gov’ts scare people into self-policing

- Governments might jail some bloggers to scare people
- Then encourage self-censorship by signaling off-limits topics
- But this could turn out one of two ways:
  - Bloggers might take cues to self-censor OR
  - Bloggers might hate censorship and rebel
- how censorship works: important to understand self-censorship
- BUT self-censorship very hard to measure
The perfect experiment

1. Be the Chinese government
2. Randomly assign censorship
3. See what bloggers write after censorship

Problem 1: unethical
Problem 2: we aren't the Chinese government
The perfect experiment

1. Be the Chinese government
The perfect experiment

1. Be the Chinese government
2. Randomly assign censorship
The perfect experiment

1. Be the Chinese government
2. Randomly assign censorship
3. See what bloggers write after censorship
The perfect experiment

1. Be the Chinese government
2. Randomly assign censorship
3. See what bloggers write after censorship

Problem 1: unethical
The perfect experiment

1. Be the Chinese government
2. Randomly assign censorship
3. See what bloggers write after censorship

Problem 1: unethical
Problem 2: we aren’t the Chinese government
How Can We Measure Deterrence?

The best approximation:

Find two bloggers

- similar users,
- similar censorship histories,
- similar numbers of posts
- similar previous post sensitivity
- with very similar posts
- written on the same day
- Only one censored

Censorship 'Mistake'

Does the censored blogger's behavior change?

Does the censored blogger stay away from the topic?

Does the censored blogger pursue the topic?
How Can We Measure Deterrence?

The best approximation:

Find two bloggers

✓ similar users,
How Can We Measure Deterrence?

The best approximation:

Find two bloggers

✓ similar users,
✓ similar censorship histories,
How Can We Measure Deterrence?

The best approximation:

Find two bloggers

✓ similar users,
✓ similar censorship histories,
✓ similar numbers of posts
How Can We Measure Deterrence?

The best approximation:

- similar users,
- similar censorship histories,
- similar numbers of posts

Find two bloggers with very similar posts

Roberts (UCSD)
How Can We Measure Deterrence?

The best approximation:

✓ similar users,
✓ similar censorship histories,
✓ similar numbers of posts

Find two bloggers with similar posts

written on the same day
How Can We Measure Deterrence?

The best approximation:

✓ similar users,
✓ similar censorship histories,
✓ similar numbers of posts

Find two bloggers

Only one censored

Does the censored blogger's behavior change?
Does the censored blogger stay away from the topic?
Does the censored blogger pursue the topic?
How Can We Measure Deterrence?

The best approximation:

✓ similar users,
✓ similar censorship histories,
✓ similar numbers of posts

Find two bloggers

Only one censored

Censorship ‘Mistake’
How Can We Measure Deterrence?

The best approximation:

- similar users,
- similar censorship histories,
- similar numbers of posts

Find two bloggers

Only one censored

Censorship ‘Mistake’

Does the censored blogger’s behavior change?
How Can We Measure Deterrence?

The best approximation:

- Find two bloggers with very similar posts written on the same day
- Only one censored

✓ similar users,
✓ similar censorship histories,
✓ similar numbers of posts

Does the censored blogger’s behavior change?
Does the censored blogger stay away from the topic?
How Can We Measure Deterrence?

The best approximation:

Find two bloggers

✓ similar users,
✓ similar censorship histories,
✓ similar numbers of posts

Only one censored

Censorship ‘Mistake’

Does the censored blogger’s behavior change?
Does the censored blogger stay away from the topic?
Does the censored blogger pursue the topic?
Introduction

Text Matching

Text as pre-treatment confounder ⇝ a surprisingly frequent problem

Applications

▶ Does censorship change a blogger's behavior?
▶ Do targeted killings of Islamic extremists create interest in their work?
▶ In International Relations, are women cited less frequently than men?
▶ Control for letters of recommendation, trade treaties, Congressional bills, etc

BUT existing matching methods impossible to apply to high-dimensional data

▶ You can't possibly match on every word! (and you wouldn't want to)

We care about controlling for covariates predictive of treatment

But with text, we don't know what predicts treatment

Very little work on this.

Roberts (UCSD) 28 April 2016 6 / 32
Text Matching

- Text as pre-treatment confounder
Text Matching

- Text as pre-treatment confounder \(\sim\) a surprisingly frequent problem
Text Matching

- Text as pre-treatment *confounder* ⇔ a surprisingly frequent problem
- Applications
Text Matching

- Text as pre-treatment confounder \(\sim\) a surprisingly frequent problem
- Applications
  - Does censorship change a bloggers behavior?
Text Matching

- Text as pre-treatment *confounder* \(\rightarrow\) a surprisingly frequent problem
- Applications
  - Does censorship change a blogger's behavior?
  - Do targeted killings of Islamic extremists create interest in their work?
Text Matching

- Text as pre-treatment confounder → a surprisingly frequent problem
- Applications
  - Does censorship change a blogger's behavior?
  - Do targeted killings of Islamic extremists create interest in their work?
  - In International Relations, are women cited less frequently than men?

- Control for letters of recommendation, trade treaties, Congressional bills, etc
- But existing matching methods impossible to apply to high-dimensional data
- You can't possibly match on every word! (and you wouldn't want to)
- We care about controlling for covariates predictive of treatment
- But with text, we don't know what predicts treatment
- Very little work on this.
Text Matching

- Text as pre-treatment confounder \(\sim\) a surprisingly frequent problem
- Applications
  - Does censorship change a bloggers behavior?
  - Do targeted killings of islamic extremists create interest in their work?
  - In International Relations, are women cited less frequently than men?
  - Control for letters of recommendation, trade treaties, Congressional bills, etc
Text Matching

- Text as pre-treatment confounder \( \leadsto \) a surprisingly frequent problem
- Applications
  - Does censorship change a blogger's behavior?
  - Do targeted killings of Islamic extremists create interest in their work?
  - In International Relations, are women cited less frequently than men?
  - Control for letters of recommendation, trade treaties, Congressional bills, etc
- BUT existing matching methods \textbf{impossible} to apply to high-dimensional data
Text Matching

- Text as pre-treatment confounder \(\rightsquigarrow\) a surprisingly frequent problem

- Applications
  - Does censorship change a bloggers behavior?
  - Do targeted killings of islamic extremists create interest in their work?
  - In International Relations, are women cited less frequently than men?
  - Control for letters of recommendation, trade treaties, Congressional bills, etc

- BUT existing matching methods impossible to apply to high-dimensional data
  - You can't possibly match on every word! (and you wouldn't want to)
Text Matching

- Text as pre-treatment *confounder* \(\rightsquigarrow\) a surprisingly frequent problem
- Applications
  - Does censorship change a blogger's behavior?
  - Do targeted killings of Islamic extremists create interest in their work?
  - In International Relations, are women cited less frequently than men?
  - Control for letters of recommendation, trade treaties, Congressional bills, etc
- BUT existing matching methods *impossible* to apply to high-dimensional data
  - You can't possibly match on every word! (and you wouldn't want to)
  - We care about controlling for covariates predictive of treatment
Text Matching

- Text as pre-treatment confounder $\leadsto$ a surprisingly frequent problem

Applications
  - Does censorship change a bloggers behavior?
  - Do targeted killings of islamic extremists create interest in their work?
  - In International Relations, are women cited less frequently than men?
  - Control for letters of recommendation, trade treaties, Congressional bills, etc

BUT existing matching methods impossible to apply to high-dimensional data
  - You can't possibly match on every word! (and you wouldn't want to)
  - We care about controlling for covariates predictive of treatment
  - But with text, we don’t know what predicts treatment
Text Matching

- Text as pre-treatment *confounder* \(\leadsto\) a surprisingly frequent problem

- Applications
  - Does censorship change a bloggers behavior?
  - Do targeted killings of islamic extremists create interest in their work?
  - In International Relations, are women cited less frequently than men?
  - Control for letters of recommendation, trade treaties, Congressional bills, etc

- BUT existing matching methods *impossible* to apply to high-dimensional data
  - You can’t possibly match on *every* word! (and you wouldn’t want to)
  - We care about controlling for covariates predictive of treatment
  - But with text, we don’t know what predicts treatment

Very little work on this.
Our Approach to Text Matching

1. Construct analogs to current methods
   - Propensity score matching
   - Multinomial Inverse Regression
   - Coarsened exact matching
   - Topically Coarsened Matching

2. Identify benefits and drawbacks of each

3. Create a new method
   - Topical Inverse Regression Matching (TIRM), by combining the two
Our Approach to Text Matching

1. Construct analogs to current methods
Our Approach to Text Matching

1. Construct analogs to current methods
   ▶ Propensity score matching $\leadsto$ Multinomial Inverse Regression
Our Approach to Text Matching

1. Construct analogs to current methods
   - Propensity score matching $\leadsto$ Multinomial Inverse Regression
   - Coarsened exact matching $\leadsto$ Topically Coarsened Matching
Our Approach to Text Matching

1. Construct **analog**s to current methods
   - Propensity score matching $\rightsquigarrow$ Multinomial Inverse Regression
   - Coarsened exact matching $\rightsquigarrow$ Topically Coarsened Matching

2. Identify **benefits** and **drawbacks** of each
Our Approach to Text Matching

1. Construct analogs to current methods
   - Propensity score matching $\leadsto$ Multinomial Inverse Regression
   - Coarsened exact matching $\leadsto$ Topically Coarsened Matching

2. Identify benefits and drawbacks of each

3. Create a new method
Introduction

Our Approach to Text Matching

1. Construct analogs to current methods
   - Propensity score matching $\leadsto$ Multinomial Inverse Regression
   - Coarsened exact matching $\leadsto$ Topically Coarsened Matching

2. Identify benefits and drawbacks of each

3. Create a new method Topical Inverse Regression Matching (TIRM)
Our Approach to Text Matching

1. Construct analogs to current methods
   - Propensity score matching $\leadsto$ Multinomial Inverse Regression
   - Coarsened exact matching $\leadsto$ Topically Coarsened Matching

2. Identify benefits and drawbacks of each

3. Create a new method Topical Inverse Regression Matching (TIRM), by combining the two
Outline of Talk
Outline of Talk

- A Quick Review of Matching
Outline of Talk

- A Quick Review of Matching
- Text Analogs to Current Matching Method
Introduction

Outline of Talk

- A Quick Review of Matching
- Text Analogs to Current Matching Method
- Topical Inverse Regression Matching
Introduction

Outline of Talk

- A Quick Review of Matching
- Text Analogs to Current Matching Method
- Topical Inverse Regression Matching
- Applications
Previous Approaches to Matching

Goal:

1. Model $p(t_i | \vec{x}_i) \rightarrow$ propensity scores
2. Match on all $\vec{x}_i$ \rightarrow coarsened exact matching

Both strategies scale poorly with high-dimensional covariates.
Previous Approaches to Matching

Goal: \( t_i \perp y_i(1), y_i(0) | \vec{x}_i \)
Previous Approaches to Matching

- Goal: $t_i \perp \! \! \! \perp y_i(1), y_i(0) | \vec{x}_i$
- Many approaches: propensity score matching
Previous Approaches to Matching

- Goal: $t_i \perp y_i(1), y_i(0) | \bar{x}_i$
- Many approaches: propensity score matching, coarsened exact matching

Both strategies scale poorly with high-dimensional covariates.
Previous Approaches to Matching

- **Goal:** \( t_i \perp y_i(1), y_i(0) \mid \bar{x}_i \)
- **Many approaches:** propensity score matching, coarsened exact matching, genetic matching
Previous Approaches to Matching

- Goal: \( t_i \perp \perp y_i(1), y_i(0) | \tilde{x}_i \)
- Many approaches: propensity score matching, coarsened exact matching, genetic matching, covariate-balanced propensity scores, entropy balancing, synthetic matching
Previous Approaches to Matching

- Goal: $t_i \perp y_i(1), y_i(0) | \vec{x}_i$

- Many approaches: propensity score matching, coarsened exact matching, genetic matching, covariate-balanced propensity scores, entropy balancing, synthetic matching, mahalanobis matching, exact matching, subclassification matching, nearest neighbor matching, full matching...
Previous Approaches to Matching

- **Goal:** \( t_i \perp y_i(1), y_i(0) | \vec{x}_i \)

- **Many approaches:** propensity score matching, coarsened exact matching, genetic matching, covariate-balanced propensity scores, entropy balancing, synthetic matching, mahalanobis matching, exact matching, subclassification matching, nearest neighbor matching, full matching ...

- **Today two of these strategies:**
Previous Approaches to Matching

- **Goal:** \( t_i \perp y_i(1), y_i(0) | \vec{x}_i \)
- **Many approaches:** propensity score matching, coarsened exact matching, genetic matching, covariate-balanced propensity scores, entropy balancing, synthetic matching, mahalanobis matching, exact matching, subclassification matching, nearest neighbor matching, full matching ...
- **Today two of these strategies:**
  1. model \( p(t_i | \vec{x}_i) \rightarrow \) propensity scores
Goal: $t_i \perp y_i(1), y_i(0) | \vec{x}_i$

Many approaches: propensity score matching, coarsened exact matching, genetic matching, covariate-balanced propensity scores, entropy balancing, synthetic matching, mahalanobis matching, exact matching, subclassification matching, nearest neighbor matching, full matching ...

Today two of these strategies:

1. model $p(t_i | \vec{x}_i) \leadsto$ propensity scores
2. match on all $\vec{x}_i \leadsto$ coarsened exact matching
Previous Approaches to Matching

- **Goal:** \( t_i \perp \perp y_i(1), y_i(0) | \vec{x}_i \)
- **Many approaches:** propensity score matching, coarsened exact matching, genetic matching, covariate-balanced propensity scores, entropy balancing, synthetic matching, mahalanobis matching, exact matching, subclassification matching, nearest neighbor matching, full matching . . .
- **Today two of these strategies:**
  1. model \( p(t_i|\vec{x}_i) \leadsto \) propensity scores
  2. match on all \( \vec{x}_i \leadsto \) coarsened exact matching
- **Both strategies scale poorly with high-dimensional covariates.**
Matching Methods for Text

Propensity Scores: An Analog for Text

Classical approach

- fit logistic regression \( \hat{\pi}_i = p(t_i | \vec{x}_i) \)
- match units with similar probability of treatment
- pros: units matched by scalar \( \hat{\pi}_i \) instead of long vector \( \vec{x}_i \)
- cons: only produces balance in expectation

Problem: high-dimensional confounders

\( X \) is \( N \times V \) (number of documents by number of words in vocab)

- can only estimate \( \hat{\pi}_i \) well when \( N \gg V \), which isn't the case!

Roberts (UCSD)
Propensity Scores: An Analog for Text

- Classical approach
Matching Methods for Text

Propensity Scores: An Analog for Text

- Classical approach
  - fit logistic regression $\hat{\pi}_i = p(t_i | \tilde{x}_i)$
Propensity Scores: An Analog for Text

- Classical approach
  - fit logistic regression $\hat{\pi}_i = p(t_i | \vec{x}_i)$
  - match units with similar probability of treatment
Propensity Scores: An Analog for Text

- Classical approach
  - fit logistic regression \( \hat{\pi}_i = p(t_i|\vec{x}_i) \)
  - match units with similar probability of treatment
  - pros: units matched by scalar (\( \hat{\pi}_i \)) instead of long vector (\( \vec{x}_i \))
Propensity Scores: An Analog for Text

- Classical approach
  - fit logistic regression \( \hat{\pi}_i = p(t_i|x_i) \)
  - match units with similar probability of treatment
  - pros: units matched by scalar \( \hat{\pi}_i \) instead of long vector \( x_i \)
  - cons: only produces balance in expectation

\( X \) is \( N \times V \) (number of documents by number of words in vocab)

\( N \gg V \), which isn't the case!
Matching Methods for Text

Propensity Scores: An Analog for Text

- Classical approach
  - fit logistic regression $\hat{\pi}_i = p(t_i|x_i)$
  - match units with similar probability of treatment
  - pros: units matched by scalar ($\hat{\pi}_i$) instead of long vector ($x_i$)
  - cons: only produces balance in expectation

- Problem: high-dimensional confounders

Roberts (UCSD)
Propensity Scores: An Analog for Text

- Classical approach
  - fit logistic regression $\hat{\pi}_i = p(t_i|x_i)$
  - match units with similar probability of treatment
  - pros: units matched by scalar ($\hat{\pi}_i$) instead of long vector ($x_i$)
  - cons: only produces balance in expectation

- Problem: high-dimensional confounders
  - $X$ is $N \times V$ (# of documents by # of words in vocab)
Propensity Scores: An Analog for Text

- Classical approach
  - fit logistic regression \( \hat{\pi}_i = p(t_i|x_i) \)
  - match units with similar probability of treatment
  - pros: units matched by scalar (\( \hat{\pi}_i \)) instead of long vector (\( x_i \))
  - cons: only produces balance in expectation

- Problem: high-dimensional confounders
  - \( X \) is \( N \times V \) (\# of documents by \# of words in vocab)
  - can only estimate \( \hat{\pi}_i \) well when \( N \gg V \), which isn’t the case!
Propensity Scores: An Analog for Text

- Classical approach
  - fit logistic regression $\hat{\pi}_i = p(t_i|\vec{x}_i)$
  - match units with similar probability of treatment
  - pros: units matched by scalar ($\hat{\pi}_i$) instead of long vector ($\vec{x}_i$)
  - cons: only produces balance in expectation

- Problem: high-dimensional confounders
  - $X$ is $N \times V$ (# of documents by # of words in vocab)
  - can only estimate $\hat{\pi}_i$ well when $N \gg V$, which isn’t the case!
Propensity Scores: An Analog for Text

Matching Methods for Text

Solution: Multinomial Inverse Regression

$\begin{align*}
\text{x}_i & \sim \text{Multinomial}(\vec{q}_i, \text{m}_i = \sum v \text{x}_i, v) \\
\phi_v & \text{measures relationship between treatment and word projection} \\
\text{z}_i = \Phi'(\text{x}_i / \text{m}_i) & \text{is a sufficient reduction} \\
\text{X} \perp \perp \text{T} | \text{Z} & \rightarrow \text{estimate } \hat{\pi}_i \\
\text{Match on } \text{z}_i \text{ or } \hat{\pi}_i
\end{align*}$
Propensity Scores: An Analog for Text

- Solution: Multinomial Inverse Regression (Cook 2007, Taddy 2013)
Propensity Scores: An Analog for Text

Solution: Multinomial Inverse Regression (Cook 2007, Taddy 2013)
Propensity Scores: An Analog for Text

- Solution: Multinomial Inverse Regression (Cook 2007, Taddy 2013)
- assume $x_i \sim \text{Multinomial}(\vec{q}_i, m_i = \sum_v x_{i,v})$

$\phi_v$ measures relationship between treatment and word projection $z_i = \Phi'(\vec{x}_i / m_i)$ is a sufficient reduction $X \perp \perp T | Z \Rightarrow$ estimate $\hat{\pi}_i$ with projection Match on $z_i$ or $\hat{\pi}_i$
Solution: Multinomial Inverse Regression (Cook 2007, Taddy 2013)

- assume $x_i \sim \text{Multinomial}(\vec{q}_i, m_i = \sum_v x_{i,v})$
- where $q_{i,v} \propto \exp(\alpha_v + t_i \phi_v)$
Solution: Multinomial Inverse Regression (Cook 2007, Taddy 2013)

- assume $x_i \sim \text{Multinomial}(\bar{q}_i, m_i = \sum_v x_{i,v})$
- where $q_{i,v} \propto \exp(\alpha_v + t_i \phi_v)$
Propensity Scores: An Analog for Text

Solution: Multinomial Inverse Regression (Cook 2007, Taddy 2013)

- assume $x_i \sim \text{Multinomial}(\vec{q}_i, m_i = \sum_v x_{i,v})$
- where $q_{i,v} \propto \exp(\alpha_v + t_i\phi_v)$
- $\phi_v$ measures relationship between treatment and word
Propensity Scores: An Analog for Text

- Solution: Multinomial Inverse Regression (Cook 2007, Taddy 2013)
  - assume $x_i \sim \text{Multinomial}(\bar{q}_i, m_i = \sum_v x_{i,v})$
  - where $q_{i,v} \propto \exp(\alpha_v + t_i \phi_v)$
  - $\phi_v$ measures relationship between treatment and word
Solution: Multinomial Inverse Regression (Cook 2007, Taddy 2013)

- assume $x_i \sim \text{Multinomial}(q_i, m_i = \sum_v x_{i,v})$
- where $q_{i,v} \propto \exp(\alpha_v + t_i \phi_v)$
- $\phi_v$ measures relationship between treatment and word
- projection $z_i = \Phi'(\tilde{x}_i/m_i)$ is a sufficient reduction $\mathbf{X} \perp \perp T|Z$
  $\leadsto$ estimate $\hat{\pi}_i$ with projection
Solution: Multinomial Inverse Regression (Cook 2007, Taddy 2013)

- assume $x_i \sim \text{Multinomial}(q_i, m_i = \sum_v x_{i,v})$
- where $q_{i,v} \propto \exp(\alpha_v + t_i \phi_v)$
- $\phi_v$ measures relationship between treatment and word
- projection $z_i = \Phi'(\bar{x}_i/m_i)$ is a sufficient reduction $X \perp \perp T|Z$
- $\leadsto$ estimate $\hat{\pi}_i$ with projection
Solution: Multinomial Inverse Regression (Cook 2007, Taddy 2013)

- Assume $x_i \sim \text{Multinomial}(\vec{q}_i, m_i = \sum_v x_{i,v})$
- Where $q_{i,v} \propto \exp(\alpha_v + t_i \phi_v)$
- $\phi_v$ measures relationship between treatment and word
- Projection $z_i = \Phi'(\vec{x}_i / m_i)$ is a sufficient reduction $\mathbf{X} \perp \perp T|Z$
- $\Rightarrow$ estimate $\hat{\pi}_i$ with projection
- Match on $z_i$ or $\hat{\pi}_i$
Problems with MNIR Matching

Matching Methods for Text
Problems with MNIR Matching

Posts equally likely to be treated are not always semantically similar:
Problems with MNIR Matching

Posts equally likely to be treated are not always semantically similar:

![Parental Advisory Explicit Content](image)
Problems with MNIR Matching

Posts equally likely to be treated are not always semantically similar:
Problems with MNIR Matching

Posts equally likely to be treated are not always semantically similar:
Problems with MNIR Matching

Posts equally likely to be treated are not always semantically similar:

- wouldn’t be a problem in expectation BUT
Problems with MNIR Matching

Posts equally likely to be treated are not always semantically similar:

- wouldn’t be a problem in expectation BUT
- hard to assess balance in the text case
Problems with MNIR Matching

Posts equally likely to be treated are not always semantically similar:

- wouldn’t be a problem in expectation BUT
- hard to assess balance in the text case
- could be more efficient if matches were more similar
Coarsened Exact Matching: An Analog for Text

- Classical approach
Coarsened Exact Matching: An Analog for Text

- Classical approach
  - coarsen each variable into natural categories
    i.e. years of education $\sim\{\text{high school, elementary school, college}\}$
Coarsened Exact Matching: An Analog for Text

- Classical approach
  - coarsen each variable into natural categories
    i.e. years of education \(\sim\) \{high school, elementary school, college\}
  - exactly match on coarsened variable
Coarsened Exact Matching: An Analog for Text

- Classical approach
  - coarsen each variable into natural categories
    i.e. years of education $\sim \{\text{high school, elementary school, college}\}$
  - exactly match on coarsened variable
  - pros: bounds imbalance on each variable
Coarsened Exact Matching: An Analog for Text

- Classical approach
  - coarsen each variable into natural categories
    i.e. years of education $\rightsquigarrow \{\text{high school, elementary school, college}\}$
  - exactly match on coarsened variable
  - pros: bounds imbalance on each variable
- Problem: high-dimensional confounder
Coarsened Exact Matching: An Analog for Text

- Classical approach
  - coarsen each variable into natural categories
    i.e. years of education $\sim \{\text{high school, elementary school, college}\}$
  - exactly match on coarsened variable
  - pros: bounds imbalance on each variable

- Problem: high-dimensional confounder
  - thousands of variables, even if we coarsen, no exact match
Coarsened Exact Matching: An Analog for Text

- **Classical approach**
  - coarsen each variable into natural categories
    - i.e. years of education $\rightsquigarrow \{\text{high school, elementary school, college}\}$
  - exactly match on **coarsened** variable
  - pros: **bounds** imbalance on each variable

- **Problem: high-dimensional confounder**
  - thousands of variables, even if we coarsen, no exact match
Coarsened Exact Matching: An Analog for Text

- Solution: topically coarsened matching
Coarsened Exact Matching: An Analog for Text

- Solution: topically coarsened matching
  - innovation: coarsen across variables
    - simple example: “tax”, “income”, “tariff” $\leadsto$ “economics”
Coarsened Exact Matching: An Analog for Text

Solution: topically coarsened matching

- innovation: coarsen across variables
  - simple example: “tax”, “income”, “tariff” \(\rightsquigarrow\) “economics”
- topics must be equivalent across documents instead of words
Coarsened Exact Matching: An Analog for Text

- Solution: topically coarsened matching
  - innovation: coarsen across variables
    simple example: “tax”, “income”, “tariff” ↦ “economics”
  - topics must be equivalent across documents instead of words
  - bounds imbalance across groups of stochastically equivalent words
Coarsened Exact Matching: An Analog for Text

Solution: topically coarsened matching
- innovation: coarsen across variables
  - simple example: “tax”, “income”, “tariff” \(\sim\) “economics”
- topics must be equivalent across documents instead of words
- bounds imbalance across groups of stochastically equivalent words

Estimate a topic model
Coarsened Exact Matching: An Analog for Text

- Solution: topically coarsened matching
  - innovation: coarsen across variables
    - simple example: “tax”, “income”, “tariff” \(\sim\) “economics”
  - topics must be equivalent across documents instead of words
  - bounds imbalance across groups of stochastically equivalent words

- Estimate a topic model

- Match on the topic density rather than raw word counts
Problems with Topical CEM

Topics aren’t always the most important predictor of treatment:
Problems with Topical CEM

Topics aren’t always the most important predictor of treatment:
Problems with Topical CEM

Topics aren’t always the most important predictor of treatment:
Problems with Topical CEM

Topics aren’t always the most important predictor of treatment:
Topics aren’t always the most important predictor of treatment:
Problems with Topical CEM

Topics aren’t always the most important predictor of treatment:
Topical Inverse Regression Matching (TIRM)

We need something that:

1. Bounds imbalance between documents
2. Doesn’t leave out important words
Topical Inverse Regression Matching (TIRM)

- We need something that:
  1. **Bounds imbalances** between documents

   "Structural Topic Model" (Roberts, Stewart, Tingley et al 2014) with treatment as content covariate
Topical Inverse Regression Matching (TIRM)

we need something that:

1. Bounds imbalance between documents
2. Doesn’t leave out important words
Topical Inverse Regression Matching (TIRM)

- We need something that:
  1. Bounds imbalance between documents
  2. Doesn’t leave out important words
- TIRM: Jointly estimate probability of treatment and topic density
Topical Inverse Regression Matching (TIRM)

- We need something that:
  1. Bounds imbalance between documents
  2. Doesn’t leave out important words
- TIRM: Jointly estimate probability of treatment and topic density
- Match on topic proportions & topic-specific probability of treatment
Topical Inverse Regression Matching (TIRM)

- We need something that:
  1. **Bounds imbalance** between documents
  2. Doesn’t leave out **important words**

- **TIRM**: Jointly estimate **probability of treatment** and **topic density**
- Match on topic proportions & topic-specific probability of treatment
  - topical bounding properties
Topical Inverse Regression Matching (TIRM)

- We need something that:
  1. Bounds imbalance between documents
  2. Doesn’t leave out important words

- TIRM: Jointly estimate probability of treatment and topic density

- Match on topic proportions & topic-specific probability of treatment
  - topical bounding properties
  - estimates which words associated with treatment

Ingredients:
- Structural Topic Model (Roberts, Stewart, Tingley et al. 2014) with treatment as content covariate
Topical Inverse Regression Matching (TIRM)

- We need something that:
  1. Bounds imbalance between documents
  2. Doesn’t leave out important words

- TIRM: Jointly estimate probability of treatment and topic density
- Match on topic proportions & topic-specific probability of treatment
  - topical bounding properties
  - estimates which words associated with treatment

- Ingredients:
Topical Inverse Regression Matching (TIRM)

- We need something that:
  1. Bounds imbalance between documents
  2. Doesn’t leave out important words

- TIRM: Jointly estimate probability of treatment and topic density
- Match on topic proportions & topic-specific probability of treatment
  - topical bounding properties
  - estimates which words associated with treatment

- Ingredients:
  - Structural Topic Model (Roberts, Stewart, Tingley et al 2014)
Topical Inverse Regression Matching (TIRM)

We need something that:
1. Bounds imbalance between documents
2. Doesn’t leave out important words

TIRM: Jointly estimate probability of treatment and topic density
Match on topic proportions & topic-specific probability of treatment
- topical bounding properties
- estimates which words associated with treatment

Ingredients:
- Structural Topic Model (Roberts, Stewart, Tingley et al 2014)
- with treatment as content covariate
Matching Methods for Text

Structural Topic Model

- STM adds a “structure” to the Latent Dirichlet Allocation (Blei, Ng and Jordan 2003) via a prior

\[ P(\text{word} | \text{topic}, \text{doc}) \propto \exp(\kappa(m) + \text{topic} \ast \kappa(k) + \text{covariate doc} \ast \kappa(c) + \text{topic*covariate doc} \ast \kappa(int)) \]

\( \kappa(c) \) and \( \kappa(int) \) \( \Rightarrow \) how words are related to treatment.
Matching Methods for Text

Structural Topic Model

- STM adds a “structure” to the Latent Dirichlet Allocation (Blei, Ng and Jordan 2003) via a prior
  - Replace topic prevalence prior → (heuristically) glm with arbitrary covariates
    (Blei and Lafferty 2006, Mimno and McCallum 2008)
Structural Topic Model

- STM adds a “structure” to the Latent Dirichlet Allocation (Blei, Ng and Jordan 2003) via a prior
  - Replace topic prevalence prior → (heuristically) glm with arbitrary covariates
    (Blei and Lafferty 2006, Mimno and McCallum 2008)
  - Replace the distribution over words → multinomial logit (Eisenstein, Ahmed and Xing 2011)
Structural Topic Model

- STM adds a “structure” to the Latent Dirichlet Allocation (Blei, Ng and Jordan 2003) via a prior
  - Replace topic prevalence prior $\rightarrow$ (heuristically) glm with arbitrary covariates
    (Blei and Lafferty 2006, Mimno and McCallum 2008)
  - Replace the distribution over words $\rightarrow$ multinomial logit (Eisenstein, Ahmed and Xing 2011)
- Documents have different expected topic proportions based on observed covariates.
Structural Topic Model

- STM adds a “structure” to the Latent Dirichlet Allocation (Blei, Ng and Jordan 2003) via a prior
  - Replace topic prevalence prior $\rightarrow$ (heuristically) glm with arbitrary covariates
    (Blei and Lafferty 2006, Mimno and McCallum 2008)
  - Replace the distribution over words $\rightarrow$ multinomial logit (Eisenstein, Ahmed and Xing 2011)

- Documents have different expected topic proportions based on observed covariates.

- Topics are now deviations from a baseline distribution.
Structural Topic Model

- STM adds a “structure” to the Latent Dirichlet Allocation (Blei, Ng and Jordan 2003) via a prior
  - Replace topic prevalence prior $\rightarrow$ (heuristically) glm with arbitrary covariates (Blei and Lafferty 2006, Mimno and McCallum 2008)
  - Replace the distribution over words $\rightarrow$ multinomial logit (Eisenstein, Ahmed and Xing 2011)
- Documents have different expected topic proportions based on observed covariates.
- Topics are now deviations from a baseline distribution.

\[ P(\text{word}|\text{topic}, \text{doc}) \propto \exp(\kappa^{(m)} + \text{topic} \times \kappa^{(k)} + \text{covariate}_{\text{doc}} \times \kappa^{(c)} + \text{topic} \times \text{covariate}_{\text{doc}} \times \kappa^{(int)}) \]
Structural Topic Model

- STM adds a “structure” to the Latent Dirichlet Allocation (Blei, Ng and Jordan 2003) via a prior
  - Replace topic prevalence prior $\rightarrow$ (heuristically) glm with arbitrary covariates
    (Blei and Lafferty 2006, Mimno and McCallum 2008)
  - Replace the distribution over words $\rightarrow$ multinomial logit (Eisenstein, Ahmed and Xing 2011)

- Documents have different expected topic proportions based on observed covariates.

- Topics are now deviations from a baseline distribution.

$$P(word|topic, doc) \propto \exp(\kappa^{(m)} + \text{topic}\kappa^{(k)} + \text{covariate}_{doc}\kappa^{(c)} + \text{topic}\text{*covariate}_{doc}\kappa^{(int)})$$

$\kappa^{(c)}$ and $\kappa^{(int)} \leadsto$ how words are related to treatment.
TIRM

Match on:

1. $\theta$: Estimated topic proportion ($K$ covariates)

2. $\text{proj}$: $\text{let } \left(\frac{x_i}{m_i}\right)$ percentage of document $i$ that is word $x_i$ ($\kappa(c)$)

$\kappa(c)$

3. Any other covariates you think are important

We generally use CEM to match but other methods could be used.

Limitations of TIRM

$\text{New:}$ relies on a parametric method to reduce dimensions

$\text{Old:}$ requires SUTVA, relevant covariates

Roberts (UCSD)
TIRM

Match on:

1. $\theta$: Estimated topic proportion ($K$ covariates)
2. $proj$:
Match on:

1. $\theta$: Estimated topic proportion ($K$ covariates)
2. $proj$:
   - let $(x_i/m_i)$ % of document $i$ that is word $x$

Limitations of TIRM

- New: relies on a parametric method to reduce dimensions
- Old: requires SUTVA, relevant covariates
TIRM

Match on:

1. $\theta$: Estimated topic proportion ($K$ covariates)
2. $proj$:
   - let $(x_i / m_i)$ % of document $i$ that is word $x$
   - $((\kappa^{(c)})'(x_i / m_i))$

Limitations of TIRM
- New: relies on a parametric method to reduce dimensions
- Old: requires SUTVA, relevant covariates


**TIRM**

Match on:

1. $\theta$: Estimated topic proportion ($K$ covariates)
2. $proj$:
   - let $(x_i/m_i)$ % of document $i$ that is word $x$
   - $(\kappa^{(c)})'(x_i/m_i)$ covariate-only projection

Limitations of TIRM:

- New: relies on a parametric method to reduce dimensions
- Old: requires SUTVA, relevant covariates
TIRM

Match on:

1. \( \theta \): Estimated topic proportion (\( K \) covariates)

2. \( \text{proj} \):
   - let \( (x_i/m_i) \% \) of document \( i \) that is word \( x \)
   - \((\kappa^{(c)})'(x_i/m_i)\) covariate-only projection
   - \((\kappa^{(c)})'(x_i/m_i) + \frac{1}{m_i} \sum_v x_{i,v} \left((\kappa^{(int)}_v)'ight) \theta_i\)
TIRM

Match on:

1. $\theta$: Estimated topic proportion ($K$ covariates)
2. $proj$:
   - let $(x_i/m_i)$ % of document $i$ that is word $x$
   - $(\kappa^{(c)})'(x_i/m_i)$ covariate-only projection
   - $(\kappa^{(c)})'(x_i/m_i) + \frac{1}{m_i} \sum_v x_{i,v} \left((\kappa^{(int)}_v)'	heta_i\right)$ topic-covariate projection
Match on:

1. \( \theta \): Estimated topic proportion \((K \text{ covariates})\)
2. \( \text{proj} \):
   - let \((x_i/m_i)\) % of document \(i\) that is word \(x\)
   - \((\kappa^{(c)}(c))(x_i/m_i)\) covariate-only projection
   - \((\kappa^{(c)}(c))(x_i/m_i) + \frac{1}{m_i} \sum_{v} x_{i,v} \left( (\kappa^{(int)}_v)^\prime \theta_i \right)\) topic-covariate projection
3. Any other covariates you think are important
TIRM

Match on:

1. $\theta$: Estimated topic proportion ($K$ covariates)
2. $\text{proj}$:
   - let $(x_i/m_i)$ % of document $i$ that is word $x$
   - $(\kappa^{(c)})'(x_i/m_i)$ covariate-only projection
   - $(\kappa^{(c)})'(x_i/m_i) + \frac{1}{m_i} \sum_v x_{i,v} \left( (\kappa_v^{(\text{int})})' \theta_i \right)$ topic-covariate projection
3. Any other covariates you think are important

We generally use CEM to match but other methods could be used.
Matching Methods for Text

**TIRM**

Match on:

1. \( \theta \): Estimated topic proportion (\( K \) covariates)
2. \( \text{proj} \):
   - let \( (x_i / m_i) \% \) of document \( i \) that is word \( x \)
   - \( (\kappa^{(c)})(x_i / m_i) \) covariate-only projection
   - \( (\kappa^{(c)})(x_i / m_i) + \frac{1}{m_i} \sum_{v} x_{i,v} \left( (\kappa^{(\text{int})})' \theta_i \right) \) topic-covariate projection
3. Any other covariates you think are important

We generally use CEM to match but other methods could be used.

Limitations of TIRM
Match on:

1. $\theta$: Estimated topic proportion ($K$ covariates)

2. $\text{proj}$:
   - let ($x_i/m_i$) % of document $i$ that is word $x$
   - $(\kappa^{(c)})'(x_i/m_i)$ covariate-only projection
   - $(\kappa^{(c)})'(x_i/m_i) + \frac{1}{m_i} \sum_v x_{i,v} \left( (\kappa^{(\text{int})}_v)' \theta_i \right)$ topic-covariate projection

3. Any other covariates you think are important

We generally use CEM to match but other methods could be used.

Limitations of TIRM

- New: relies on a parametric method to reduce dimensions
TIRM

Match on:

1. $\theta$: Estimated topic proportion ($K$ covariates)
2. $proj$:
   - let $(x_i/m_i) \%$ of document $i$ that is word $x$
   - $(k^{(c)})'(x_i/m_i)$ covariate-only projection
   - \[ (k^{(c)})'(x_i/m_i) + \frac{1}{m_i} \sum_v x_{i,v} \left( (k_v^{(int)})' \theta_i \right) \] topic-covariate projection
3. Any other covariates you think are important

We generally use CEM to match but other methods could be used.

Limitations of TIRM

- New: relies on a parametric method to reduce dimensions
- Old: requires SUTVA, relevant covariates
Simulations

Set up:

1. Simulate 200 outcome and treatment with confounding topics and words
2. Estimate STM
3. Condition on topics and projection

Roberts (UCSD)
Simulations

Set up:

1. Simulate 200 outcome and treatment with confounding topics and words
Simulations

Set up:
1. Simulate 200 outcome and treatment with confounding topics and words
2. Estimate STM
Simulations

Set up:

1. Simulate 200 outcome and treatment with confounding topics and words
2. Estimate STM
3. Condition on topics and projection
Simulations

Set up:
1. Simulate 200 outcome and treatment with confounding topics and words
2. Estimate STM
3. Condition on topics and projection
Example 1: How do bloggers react to censorship?

Data: 593 bloggers over 6 months spanning 2011, 2012

150,000 posts

Return to blogs to measure censorship

Find censors’ mistakes: two similar blogs, different censorship

Also match on date, previous censorship, previous sensitivity.

How do ‘treated’ bloggers react to censorship?

Outcome: Bloggers’ writings after censorship:

- Censorship rate after
- Sensitivity of blog text after (estimated by TIRM)
- Topical content of blogs after
Example 1: How do bloggers react to censorship?

- **Data**: 593 bloggers over 6 months spanning 2011, 2012
Example 1: How do bloggers react to censorship?

- **Data**: 593 bloggers over 6 months spanning 2011, 2012
- **150,000 posts**
Example 1: How do bloggers react to censorship?

- **Data:** 593 bloggers over 6 months spanning 2011, 2012
- 150,000 posts
- Return to blogs to measure censorship
Example 1: How do bloggers react to censorship?

- **Data**: 593 bloggers over 6 months spanning 2011, 2012
- 150,000 posts
- Return to blogs to measure censorship
- Find censors’ mistakes: two similar blogs, different censorship
- Outcome: Bloggers’ writings after censorship:
  - censorship rate after
  - sensitivity of blog text after (estimated by TIRM)
  - topical content of blogs after
Example 1: How do bloggers react to censorship?

- **Data:** 593 bloggers over 6 months spanning 2011, 2012
- 150,000 posts
- Return to blogs to measure censorship
- Find censors’ mistakes: two similar blogs, different censorship
- Also match on date, previous censorship, previous sensitivity.
Example 1: How do bloggers react to censorship?

- **Data:** 593 bloggers over 6 months spanning 2011, 2012
- 150,000 posts
- Return to blogs to measure censorship
- Find censors’ mistakes: two similar blogs, different censorship
- Also match on date, previous censorship, previous sensitivity.
- How do ‘treated’ bloggers react to censorship?
Example 1: How do bloggers react to censorship?

- **Data:** 593 bloggers over 6 months spanning 2011, 2012
- 150,000 posts
- Return to blogs to measure censorship
- Find censors’ mistakes: two similar blogs, different censorship
- Also match on date, previous censorship, previous sensitivity.
- How do ’treated’ bloggers react to censorship?
- **Outcome:** Bloggers’ writings after censorship:
Example 1: How do bloggers react to censorship?

- **Data:** 593 bloggers over 6 months spanning 2011, 2012
- 150,000 posts
- Return to blogs to measure **censorship**
- Find censors’ **mistakes:** two similar blogs, different censorship
- Also match on date, previous censorship, previous sensitivity.
- How do ’treated’ bloggers react to censorship?
- **Outcome:** Bloggers’ writings after censorship:
  - censorship rate after
Example 1: How do bloggers react to censorship?

- **Data:** 593 bloggers over 6 months spanning 2011, 2012
- 150,000 posts
- Return to blogs to measure censorship
- Find censors’ mistakes: two similar blogs, different censorship
- Also match on date, previous censorship, previous sensitivity.
- How do ’treated’ bloggers react to censorship?
- **Outcome:** Bloggers’ writings after censorship:
  - censorship rate after
  - sensitivity of blog text after (estimated by TIRM)
Example 1: How do bloggers react to censorship?

- **Data**: 593 bloggers over 6 months spanning 2011, 2012
- 150,000 posts
- Return to blogs to measure censorship
- Find censors’ mistakes: two similar blogs, different censorship
- Also match on date, previous censorship, previous sensitivity.
- How do ’treated’ bloggers react to censorship?
- **Outcome**: Bloggers’ writings after censorship:
  - censorship rate after
  - sensitivity of blog text after (estimated by TIRM)
  - topical content of blogs after
TIRM Finds Almost Identical Posts
TIRM Finds Almost Identical Posts

![Graph showing TIRM String Kernel Similarity](image-url)

- **TIRM**
- **String Kernel Similarity**
- **Frequency**
- **0.0 0.2 0.4 0.6 0.8 1.0**
- **0 5 10 15 20**

Roberts (UCSD)
TIRM Finds Almost Identical Posts

![Graph showing topic match against string kernel similarity.

The x-axis represents string kernel similarity ranging from 0.0 to 1.0.

The y-axis represents frequency, with labels at 0, 10, 20, and 30.

The graph displays a concentration of matches at higher similarity values, indicating a high degree of topic similarity in the posts detected by TIRM.]
TIRM Finds Almost Identical Posts
Results

We find 46 matched blogs (censors' mistakes).

Nearly perfect matches.

Most matched posts are about Bo Xilai incident, Maoist protests.

5 posts before treatment:
No statistical difference between actual censorship.
No statistical difference between TIRM-predicted censorship.
(Not surprising, we are matching on these!)

5 posts after treatment:
- Treated group: 20% censorship
- Control group: 7% censorship

TIRM estimates treated text significantly more sensitive than control.

Treated group talks significantly more about Bo Xilai incident after censorship than control.

Treated group talks significantly more about CCP History/Mao after censorship than control.
Results

- We find 46 matched blogs (censors’ mistakes)
  - Nearly perfect matches
  - Most matched posts are about Bo Xilai incident, Maoist protests
  - 5 posts before treatment:
    - No statistical difference between actual censorship
    - No statistical difference between TIRM-predicted censorship
    - (Not surprising, we are matching on these!)
  - 5 posts after treatment:
    - Treated group: 20% censorship
    - Control group: 7% censorship
    - TIRM estimates treated text significantly more sensitive than control
    - Treated group talks significantly more about Bo Xilai incident after censorship than control
    - Treated group talks significantly more about CCP History/Mao after censorship than control
Results

- We find 46 matched blogs (censors’ mistakes)
- Nearly perfect matches
Results

- We find 46 matched blogs (censors’ mistakes)
- Nearly perfect matches
- Most matched posts are about Bo Xilai incident, Maoist protests
Results

- We find 46 matched blogs (censors’ mistakes)
- Nearly perfect matches
- Most matched posts are about Bo Xilai incident, Maoist protests
- 5 posts before treatment:

  ▶ No statistical difference between actual censorship
  ▶ No statistical difference between TIRM-predicted censorship
  ▶ (Not surprising, we are matching on these!)

  ▶ 5 posts after treatment:
    ▶ Treated group: 20% censorship
    ▶ Control group: 7% censorship
    ▶ TIRM estimates treated text significantly more sensitive than control
    ▶ Treated group talks significantly more about Bo Xilai incident after censorship than control
    ▶ Treated group talks significantly more about CCP History/Mao after censorship than control
Results

- We find 46 matched blogs (censors’ mistakes)
- Nearly perfect matches
- Most matched posts are about Bo Xilai incident, Maoist protests
- 5 posts before treatment:
  - No statistical difference between actual censorship
Results

- We find 46 matched blogs (censors’ mistakes)
- Nearly perfect matches
- Most matched posts are about Bo Xilai incident, Maoist protests
- 5 posts before treatment:
  - No statistical difference between actual censorship
  - No statistical difference between TIRM-predicted censorship
- Treated group: 20% censorship
- Control group: 7% censorship
- Treated group talks significantly more about Bo Xilai incident after censorship than control
- Treated group talks significantly more about CCP History/Mao after censorship than control
Results

- We find 46 matched blogs (censors’ mistakes)
- Nearly perfect matches
- Most matched posts are about Bo Xilai incident, Maoist protests

5 posts before treatment:
  - No statistical difference between actual censorship
  - No statistical difference between TIRM-predicted censorship
  - (Not surprising, we are matching on these!)
Results

- We find 46 matched blogs (censors’ mistakes)
- Nearly perfect matches
- Most matched posts are about Bo Xilai incident, Maoist protests

- 5 posts before treatment:
  - No statistical difference between actual censorship
  - No statistical difference between TIRM-predicted censorship
  - (Not surprising, we are matching on these!)

- 5 posts after treatment:
Results

- We find 46 matched blogs (censors’ mistakes)
- Nearly perfect matches
- Most matched posts are about Bo Xilai incident, Maoist protests

- 5 posts before treatment:
  - No statistical difference between actual censorship
  - No statistical difference between TIRM-predicted censorship
  - (Not surprising, we are matching on these!)

- 5 posts after treatment:
  - Treated group: 20% censorship
Results

- We find 46 matched blogs (censors’ mistakes)
- Nearly perfect matches
- Most matched posts are about Bo Xilai incident, Maoist protests

- 5 posts before treatment:
  - No statistical difference between actual censorship
  - No statistical difference between TIRM-predicted censorship
  - (Not surprising, we are matching on these!)

- 5 posts after treatment:
  - Treated group: 20% censorship Control group: 7% censorship
Results

- We find 46 matched blogs (censors’ mistakes)
- Nearly perfect matches
- Most matched posts are about Bo Xilai incident, Maoist protests

- 5 posts before treatment:
  - No statistical difference between actual censorship
  - No statistical difference between TIRM-predicted censorship
  - (Not surprising, we are matching on these!)

- 5 posts after treatment:
  - Treated group: 20% censorship Control group: 7% censorship
  - TIRM estimates treated text significantly more sensitive than control
Results

- We find 46 matched blogs (censors’ mistakes)
- Nearly perfect matches
- Most matched posts are about Bo Xilai incident, Maoist protests

5 posts before treatment:
- No statistical difference between actual censorship
- No statistical difference between TIRM-predicted censorship
- (Not surprising, we are matching on these!)

5 posts after treatment:
- Treated group: 20% censorship Control group: 7% censorship
- TIRM estimates treated text significantly more sensitive than control
- Treated group talks significantly more about Bo Xilai incident after censorship than control
Results

- We find 46 matched blogs (censors’ mistakes)
- Nearly perfect matches
- Most matched posts are about Bo Xilai incident, Maoist protests

- 5 posts before treatment:
  - No statistical difference between actual censorship
  - No statistical difference between TIRM-predicted censorship
  - (Not surprising, we are matching on these!)

- 5 posts after treatment:
  - Treated group: 20% censorship Control group: 7% censorship
  - TIRM estimates treated text significantly more sensitive than control
  - Treated group talks significantly more about Bo Xilai incident after censorship than control
  - Treated group talks significantly more about CCP History/Mao after censorship than control
Example 2: Does gender affect citations in Political Science?

Maliniak, Powers, Walter (2013): women get cited less than men in IR

Problem: women write about different topics than men

Maliniak et al. solution: Code articles into (many) categories

Our solution: Text matching!

Data: 3,201 journal articles from top 12 IR journals, 1980-2006.

Code lots of variables, including gender, article age, tenure, etc.

Treatment: all-female Control: co-ed/all-male

Our motive: Find similar articles, see how they are cited differently.
Example 2: Does gender affect citations in Political Science?

- Maliniak, Powers, Walter (2013): women get cited less than men in IR
- Problem: women write about different topics than men
Example 2: Does gender affect citations in Political Science?

- Maliniak, Powers, Walter (2013): women get cited less than men in IR
- Problem: women write about different topics than men
- Maliniak et al solution: Code articles into (many) categories
Example 2: Does gender affect citations in Political Science?

- Maliniak, Powers, Walter (2013): women get cited less than men in IR
- Problem: women write about different topics than men
- Maliniak et al solution: Code articles into (many) categories
- Our solution: Text matching!
Example 2: Does gender affect citations in Political Science?

- Maliniak, Powers, Walter (2013): women get cited less than men in IR
- Problem: women write about different topics than men
- Maliniak et al solution: Code articles into (many) categories
- Our solution: Text matching!
- Data: 3,201 journal articles from top 12 IR journals, 1980-2006.
Example 2: Does gender affect citations in Political Science?

- Maliniak, Powers, Walter (2013): women get cited less than men in IR
- Problem: women write about different topics than men
- Maliniak et al solution: Code articles into (many) categories
- Our solution: Text matching!
- Data: 3,201 journal articles from top 12 IR journals, 1980-2006.
- Code lots of variables, including gender, article age, tenure, etc.
Example 2: Does gender affect citations in Political Science?

- Maliniak, Powers, Walter (2013): women get cited less than men in IR
- Problem: women write about different topics than men
- Maliniak et al solution: Code articles into (many) categories
- Our solution: Text matching!
- Data: 3,201 journal articles from top 12 IR journals, 1980-2006.
- Code lots of variables, including gender, article age, tenure, etc.
- Treatment: all-female Control: co-ed/all-male
Example 2: Does gender affect citations in Political Science?

- Maliniak, Powers, Walter (2013): women get cited less than men in IR
- Problem: women write about different topics than men
- Maliniak et al solution: Code articles into (many) categories
- Our solution: Text matching!
- Data: 3,201 journal articles from top 12 IR journals, 1980-2006.
- Code lots of variables, including gender, article age, tenure, etc.
- Treatment: all-female Control: co-ed/all-male
- Our motive: Find similar articles, see how they are cited differently.
Words men and women use differently in IR
Words men and women use differently in IR

Topic Matching:
Words men and women use differently in IR

TIRM:

Mutual Information in Unmatched Dataset

All Female Articles – Male/Coed Articles
## TIRM Reduces Topical Differences

| Topic 1 | State, power, intern, system, polit |
| Topic 2 | Model, variable, data, effect, measure |
| Topic 3 | Polit, conflict, group, ethnic, state |
| Topic 4 | Econom, development, industry, country, world |
| Topic 5 | Polit, social, one, theoria, world |
| Topic 6 | Game, will, cooperation, can, strategy |
| Topic 7 | Policy, foreign, public, political, decisions |
| Topic 8 | Polit, party, policy, government, vote |
| Topic 9 | Nuclear, weapon, arm, force, defense |
| Topic 10 | State, China, unit, foreign, policy |
| Topic 11 | International, state, organization, institute, law |
| Topic 12 | Soviet, military, war, force, defense |
| Topic 13 | Trade, economic, policy, bank, international |
| Topic 14 | War, conflict, state, dispute, democracy |
| Topic 15 | War, Israel, peace, conflict, Arab |

### Mean topic difference (Women−Men)

-0.10 -0.05 0.00 0.05 0.10
TIRM Reduces Topical Differences

Mean topic difference (Women−Men)

-0.10  -0.05  0.00  0.05  0.10

Full Data Set (Unmatched)
• TIRM
○ MNIR
△ Topic Matching
□ Human Coding Matched

Roberts (UCSD)
Results

Maliniak et al: Women receive 80% the citations of men

In our data: women receive fewer citations robust across matches

Final match: Women receive 40-60% the citations of men

Still looking into why we are getting more extreme results

Could be the difference is in very high citation counts
Results

- Maliniak et al: Women receive 80% the citations of men
Results

- Maliniak et al: Women receive 80% the citations of men
- In our data: women receive fewer citations robust across matches
Results

- Maliniak et al: Women receive 80% the citations of men
- In our data: women receive fewer citations robust across matches
- Final match: Women receive 40-60% the citations of men
Results

- Maliniak et al: Women receive 80% the citations of men
- In our data: women receive fewer citations robust across matches
- Final match: Women receive 40-60% the citations of men
- Still looking into why we are getting more extreme results
Results

- Maliniak et al: Women receive 80% the citations of men
- In our data: women receive fewer citations robust across matches
- Final match: Women receive 40-60% the citations of men
- Still looking into why we are getting more extreme results
- Could be the difference is in very high citation counts
Ex. 3: Did killing Bin Laden make his ideas less popular?

“His death will serve as a global clarion call for another generation of jihadists.”
– Ed Husain (CFR)

“al-Qaida may emerge even more radical, and more closely united under the banner of an iconic martyr.”
– Abdel Bari Atwan (The Guardian)
Ex. 3: Did killing Bin Laden make his ideas less popular?

“His death will serve as a global clarion call for another generation of jihadists.”
– Ed Husain (CFR)

“al-Qaida may emerge even more radical, and more closely united under the banner of an iconic martyr.”
– Abdel Bari Atwan (The Guardian)
Ex. 3: Did killing Bin Laden make his ideas less popular?

“The idea that Obama made a strategic misstep by killing a man responsible for the death of thousands of U.S. citizens and committed to killing thousands more is absurd. Rather than making him a martyr, Bin Laden’s killing demonstrated that he was, like the rest of us, mortal.” – Robert Simcox (LA Times)
Ex. 3: Did killing Bin Laden make his ideas less popular?

We don’t really know.

Usama Bin Laden
5/2/2011

Anwar al-Awlaki
9/30/2011

Abu Yahya al-Libi
6/5/2012
Ex. 3: Did killing Bin Laden make his ideas less popular?

We don’t really know.

Usama Bin Laden  
5/2/2011

Anwar al-Awlaki  
9/30/2011

Abu Yahya al-Libi  
6/5/2012
Empirical Strategy

View-count data from a Jihadist website, scraped over time

Does targeted killing of Bin Laden increase views of his work?

TIRM matching + match on pre-treatment page views.

QOI is ATT: nearest neighbor matching instead of CEM

Validation: Matches accord with sub-pages on website
Empirical Strategy

- View-count data from a Jihadist website, scraped over time
Empirical Strategy

- View-count data from a Jihadist website, scraped over time
- Does targeted killing of Bin Laden increase views of his work?
Empirical Strategy

- View-count data from a Jihadist website, scraped over time
- Does targeted killing of Bin Laden increase views of his work?
- TIRM matching + match on pre-treatment page views.
Empirical Strategy

- View-count data from a Jihadist website, scraped over time
- Does targeted killing of Bin Laden increase views of his work?
- TIRM matching + match on pre-treatment page views.
- QOI is ATT: nearest neighbor matching instead of CEM
Empirical Strategy

- View-count data from a Jihadist website, scraped over time
- Does targeted killing of Bin Laden increase views of his work?
- TIRM matching + match on pre-treatment page views.
- QOI is ATT: nearest neighbor matching instead of CEM
- Validation: Matches accord with sub-pages on website
Martyr Effect: Clear short-term increase in page views
**Martyr Effect**: Clear short-term increase in page views

**Figure**: Estimated effects of Usama Bin Laden’s death (on May 2, 2011) on subsequent page views of his documents on a large jihadist web-library.
Conclusion

Lots of applications measure pre-treatment confounders with text

No methods developed yet to do this

We develop a new method, Topical Inverse Regression Matching

Matching on topical density estimate

\( \rightarrow \) bounds differences between topics

Match on probability of treatment

\( \rightarrow \) balances on words related to treatment

Future work:

- Develop theoretical properties of TIRM
- Extend to high-dimensional cases other than text
- Create an R package
Conclusion

- Lots of applications measure pre-treatment confounders with text
Conclusion

- Lots of applications measure pre-treatment confounders with text
- No methods developed yet to do this
Conclusion

- Lots of applications measure pre-treatment confounders with text
- No methods developed yet to do this
- We develop a new method, Topical Inverse Regression Matching
Conclusion

- Lots of applications measure pre-treatment confounders with text
- No methods developed yet to do this
- We develop a new method, Topical Inverse Regression Matching
  - Matching on topical density estimate
Conclusion

- Lots of applications measure pre-treatment confounders with text
- No methods developed yet to do this
- We develop a new method, Topical Inverse Regression Matching
  - Matching on topical density estimate \( \sim \) bounds differences between topics
Conclusion

- Lots of applications measure pre-treatment confounders with text
- No methods developed yet to do this
- We develop a new method, Topical Inverse Regression Matching
  - Matching on topical density estimate $\sim$ bounds differences between topics
  - Match on probability of treatment
Conclusion

- Lots of applications measure pre-treatment confounders with text
- No methods developed yet to do this
- We develop a new method, Topical Inverse Regression Matching
  - Matching on topical density estimate $\sim$ bounds differences between topics
  - Match on probability of treatment $\sim$ balances on words related to treatment

Future work:
- Develop theoretical properties of TIRM
- Extend to high-dimensional cases other than text
- Create an R package
Conclusion

- Lots of applications measure pre-treatment confounders with text
- No methods developed yet to do this
- We develop a new method, Topical Inverse Regression Matching
  - Matching on topical density estimate $\rightarrow$ bounds differences between topics
  - Match on probability of treatment $\rightarrow$ balances on words related to treatment
- Future work:
Conclusion

- Lots of applications measure pre-treatment confounders with text
- No methods developed yet to do this
- We develop a new method, Topical Inverse Regression Matching
  - Matching on topical density estimate $\sim$ bounds differences between topics
  - Match on probability of treatment $\sim$ balances on words related to treatment
- Future work:
  - Develop theoretical properties of TIRM
Lots of applications measure pre-treatment confounders with text
No methods developed yet to do this
We develop a new method, Topical Inverse Regression Matching
  Matching on topical density estimate \(\leadsto\) bounds differences between topics
  Match on probability of treatment \(\leadsto\) balances on words related to treatment
Future work:
  Develop theoretical properties of TIRM
  Extend to high-dimensional cases other than text
Conclusion

- Lots of applications measure pre-treatment confounders with text
- No methods developed yet to do this
- We develop a new method, Topical Inverse Regression Matching
  - Matching on topical density estimate $\sim$ bounds differences between topics
  - Match on probability of treatment $\sim$ balances on words related to treatment
- Future work:
  - Develop theoretical properties of TIRM
  - Extend to high-dimensional cases other than text
  - Create an R package