Matching Methods for Causal Inference with Time-Series Cross-Sectional Data

Kosuke Imai In Song Kim Erik Wang

Harvard MIT ANU

Mannheim Centre for European Social Research (MZES) Social Science Data Lab, University of Mannheim

Oct 5, 2022

Our Research in a Nutshell

We propose a matching method for time-series cross-sectional data

Our Research in a Nutshell

We propose a matching method for time-series cross-sectional data

1. create a matched set for each treated observation

Our Research in a Nutshell

We propose a matching method for time-series cross-sectional data

1. create a matched set for each treated observation
2. refine the matched set via any matching or weighting method

Our Research in a Nutshell

We propose a matching method for time-series cross-sectional data

1. create a matched set for each treated observation
2. refine the matched set via any matching or weighting method
3. compute the difference-in-differences (DiD) estimator

Our Research in a Nutshell

We propose a matching method for time-series cross-sectional data

1. create a matched set for each treated observation
2. refine the matched set via any matching or weighting method
3. compute the difference-in-differences (DiD) estimator

Several advantages

Our Research in a Nutshell

We propose a matching method for time-series cross-sectional data

1. create a matched set for each treated observation
2. refine the matched set via any matching or weighting method
3. compute the difference-in-differences (DiD) estimator

Several advantages

- Contra two-way FE: matching is non-parametric

Our Research in a Nutshell

We propose a matching method for time-series cross-sectional data

1. create a matched set for each treated observation
2. refine the matched set via any matching or weighting method
3. compute the difference-in-differences (DiD) estimator

Several advantages

- Contra two-way FE: matching is non-parametric
- Reduces model dependency and mitigates confounding

Our Research in a Nutshell

We propose a matching method for time-series cross-sectional data

1. create a matched set for each treated observation
2. refine the matched set via any matching or weighting method
3. compute the difference-in-differences (DiD) estimator

Several advantages

- Contra two-way FE: matching is non-parametric
- Reduces model dependency and mitigates confounding
- Simple diagnostic tools

Our Research in a Nutshell

We propose a matching method for time-series cross-sectional data

1. create a matched set for each treated observation
2. refine the matched set via any matching or weighting method
3. compute the difference-in-differences (DiD) estimator

Several advantages

- Contra two-way FE: matching is non-parametric
- Reduces model dependency and mitigates confounding
- Simple diagnostic tools
- À la two-way FE: eliminates unit-specific (but time-invariant) and time-specific (but unit-invariant) confounding

Our Research in a Nutshell

We propose a matching method for time-series cross-sectional data

1. create a matched set for each treated observation
2. refine the matched set via any matching or weighting method
3. compute the difference-in-differences (DiD) estimator

Several advantages

- Contra two-way FE: matching is non-parametric
- Reduces model dependency and mitigates confounding
- Simple diagnostic tools
- À la two-way FE: eliminates unit-specific (but time-invariant) and time-specific (but unit-invariant) confounding
- Resolves and goes beyond the "Staggered DiD" in the econ literature

Our Research in a Nutshell

We propose a matching method for time-series cross-sectional data

1. create a matched set for each treated observation
2. refine the matched set via any matching or weighting method
3. compute the difference-in-differences (DiD) estimator

Several advantages

- Contra two-way FE: matching is non-parametric
- Reduces model dependency and mitigates confounding
- Simple diagnostic tools
- À la two-way FE: eliminates unit-specific (but time-invariant) and time-specific (but unit-invariant) confounding
- Resolves and goes beyond the "Staggered DiD" in the econ literature
    - *any* binary treatment pattern

Outline of talk

Outline of the Talk

1. Motivation
2. The Proposed Matching Methodology
3. Simulation Studies
4. Empirical Analysis
5. Conclusion

Motivation

- Matching methods have become part of toolkit for social scientists

Motivation

- Matching methods have become part of toolkit for social scientists
  1. reduces model dependence in observational studies

Motivation

- Matching methods have become part of toolkit for social scientists
  1. reduces model dependence in observational studies
  2. provides diagnostics through balance checks

Motivation

- Matching methods have become part of toolkit for social scientists
  1. reduces model dependence in observational studies
  2. provides diagnostics through balance checks
  3. clarifies comparison between treated and control units

Motivation

- Matching methods have become part of toolkit for social scientists
  1. reduces model dependence in observational studies
  2. provides diagnostics through balance checks
  3. clarifies comparison between treated and control units
- Yet, almost all existing matching methods deal with cross-sectional data

Motivation

- Matching methods have become part of toolkit for social scientists
  1. reduces model dependence in observational studies
  2. provides diagnostics through balance checks
  3. clarifies comparison between treated and control units
- Yet, almost all existing matching methods deal with cross-sectional data
- Causal inference with TSCS data is trickier for matching

Motivation

- Matching methods have become part of toolkit for social scientists
  1. reduces model dependence in observational studies
  2. provides diagnostics through balance checks
  3. clarifies comparison between treated and control units
- Yet, almost all existing matching methods deal with cross-sectional data
- Causal inference with TSCS data is trickier for matching
  - different units receive treatment at different times (staggered adoption)
  - treatment could also "reverse:" units could go in and out of treatment condition at different points in time

Motivation

- Matching methods have become part of toolkit for social scientists
  1. reduces model dependence in observational studies
  2. provides diagnostics through balance checks
  3. clarifies comparison between treated and control units
- Yet, almost all existing matching methods deal with cross-sectional data
- Causal inference with TSCS data is trickier for matching
  - different units receive treatment at different times (staggered adoption)
  - treatment could also "reverse:" units could go in and out of treatment condition at different points in time
- But matching representation is crucial for TSCS data

Motivation

- Matching methods have become part of toolkit for social scientists
  1. reduces model dependence in observational studies
  2. provides diagnostics through balance checks
  3. clarifies comparison between treated and control units
- Yet, almost all existing matching methods deal with cross-sectional data
- Causal inference with TSCS data is trickier for matching
  - different units receive treatment at different times (staggered adoption)
  - treatment could also "reverse:" units could go in and out of treatment condition at different points in time
- But matching representation is crucial for TSCS data
  - Find **all** two-period DiD cases in the data with a matching algorithm

Motivation

- Matching methods have become part of toolkit for social scientists
  1. reduces model dependence in observational studies
  2. provides diagnostics through balance checks
  3. clarifies comparison between treated and control units
- Yet, almost all existing matching methods deal with cross-sectional data
- Causal inference with TSCS data is trickier for matching
  - different units receive treatment at different times (staggered adoption)
  - treatment could also "reverse:" units could go in and out of treatment condition at different points in time
- But matching representation is crucial for TSCS data
  - Find **all** two-period DiD cases in the data with a matching algorithm
- Empirical applications:
  - Democracy and economic growth (Acemoglu et al.)
  - Interstate war and inheritance tax (Scheve & Stasavage)

Democracy and Economic Growth

- Long-standing question in political economy (e.g. Przeworski et al. 2000; Papaioannou and Siourounis, 2008)

Democracy and Economic Growth

- Long-standing question in political economy (e.g. Przeworski et al. 2000; Papaioannou and Siourounis, 2008)
- Acemoglu et al. (2019): an up-to-date empirical study

Democracy and Economic Growth

- Long-standing question in political economy (e.g. Przeworski et al. 2000; Papaioannou and Siourounis, 2008)
- Acemoglu et al. (2019): an up-to-date empirical study
- TSCS data set: 184 countries from 1960 to 2010

Democracy and Economic Growth

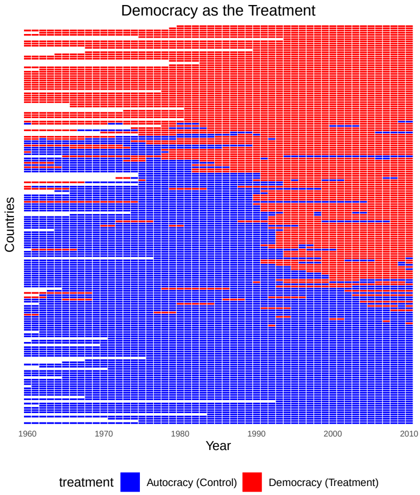
- Long-standing question in political economy (e.g. Przeworski et al. 2000; Papaioannou and Siourounis, 2008)
- Acemoglu et al. (2019): an up-to-date empirical study
- TSCS data set: 184 countries from 1960 to 2010
- Two-way FE: Country and year fixed effects

Democracy and Economic Growth

- Long-standing question in political economy (e.g. Przeworski et al. 2000; Papaioannou and Siourounis, 2008)
- Acemoglu et al. (2019): an up-to-date empirical study
- TSCS data set: 184 countries from 1960 to 2010
- Two-way FE: Country and year fixed effects
- Binary variable for democracy or not as the treatment

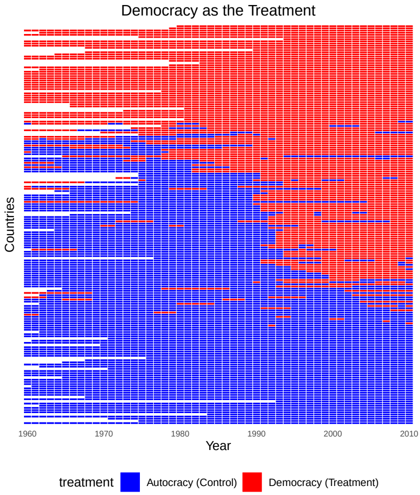
Treatment Variation Plot

- Treatment introduced at different times for different countries



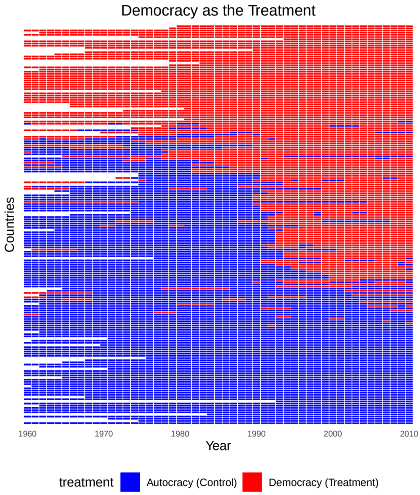
Treatment Variation Plot

- Treatment introduced at different times for different countries
- Beyond staggered adoption:



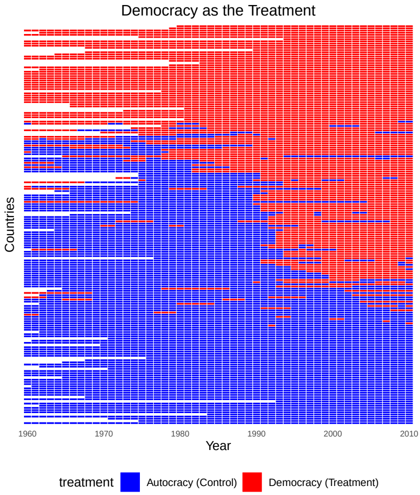
Treatment Variation Plot

- Treatment introduced at different times for different countries
- Beyond staggered adoption:
  - democratic transition and authoritarian reversal



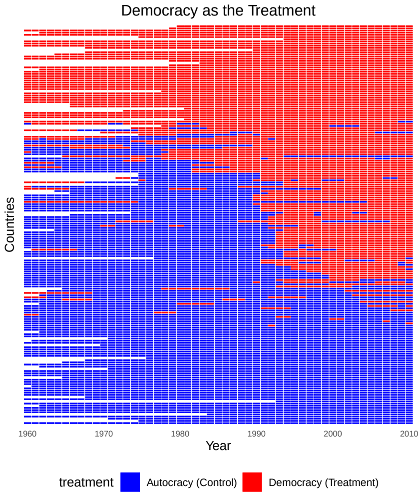
Treatment Variation Plot

- Treatment introduced at different times for different countries
- Beyond staggered adoption:
  - democratic transition and authoritarian reversal
- Identification strategy:



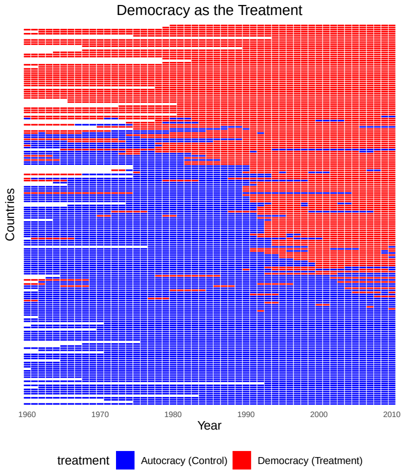
Treatment Variation Plot

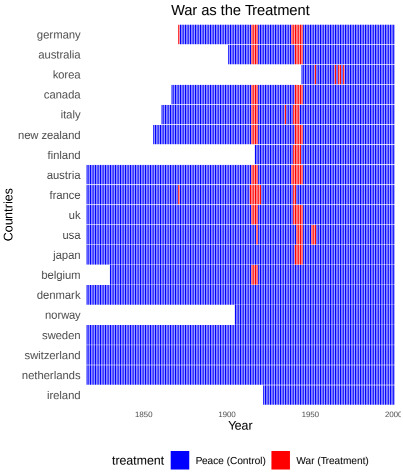
- Treatment introduced at different times for different countries
- Beyond staggered adoption:
  - democratic transition and authoritarian reversal
- Identification strategy:
  - within-unit over-time variation
  - within-time across-units variation
  - variation in the timing of treatment



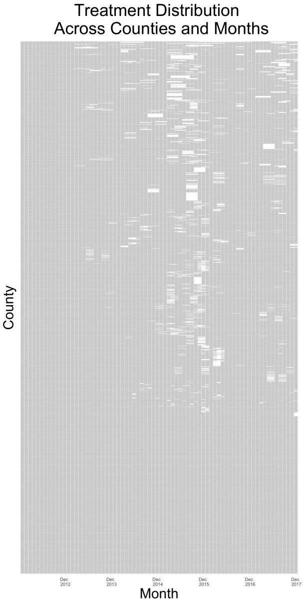
Treatment Variation Plot

Democracy and Growth by Acemoglu et al.(Left)
War and Inheritance Taxation by Scheve and Stasavage (Right)

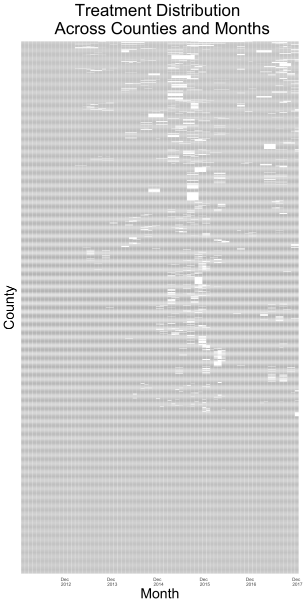




Treatment Variation Plot



Treatment Variation Plot



A micro data example: anti-corruption inspection and bureaucrats' productivity (Wang 2022)

72 months and $2,698$ counties

Each cell a "county-month"

Treated (under inspection) cells in white

Two-way FE and Difference-in-Differences

Two-way FE and Difference-in-Differences

- Two-way FE inconsistent for causal effect unless (e.g. Imai & Kim 2011):
  - Treatment effect constant across units OR
  - Proportion of treated observations constant across units

Two-way FE and Difference-in-Differences

- Two-way FE inconsistent for causal effect unless (e.g. Imai & Kim 2011):
  - Treatment effect constant across units OR
  - Proportion of treated observations constant across units
- Strong parametric assumptions in linear regression

  e.g. Aronow & Samii 2015

Two-way FE and Difference-in-Differences

- Two-way FE inconsistent for causal effect unless (e.g. Imai & Kim 2011):
  - Treatment effect constant across units OR
  - Proportion of treated observations constant across units
- Strong parametric assumptions in linear regression

  e.g. Aronow & Samii 2015
- Two-way FE $\neq$ DiD beyond the simplest 2-period setting

Two-way FE and Difference-in-Differences

- Two-way FE inconsistent for causal effect unless (e.g. Imai & Kim 2011):
  - Treatment effect constant across units OR
  - Proportion of treated observations constant across units
- Strong parametric assumptions in linear regression

  e.g. Aronow & Samii 2015
- Two-way FE $\neq$ DiD beyond the simplest 2-period setting
- Recent works on TSCS data focus on staggered DiD

  e.g. Callaway & Sant'Anna 2021; Athey & Imbens 2021

Two-way FE and Difference-in-Differences

- Two-way FE inconsistent for causal effect unless (e.g. Imai & Kim 2011):
  - Treatment effect constant across units OR
  - Proportion of treated observations constant across units
- Strong parametric assumptions in linear regression

  e.g. Aronow & Samii 2015
- Two-way FE $\neq$ DiD beyond the simplest 2-period setting
- Recent works on TSCS data focus on staggered DiD

  e.g. Callaway & Sant'Anna 2021; Athey & Imbens 2021
  - Treatment timing differs across units
  - Once treated, will remain treated in the following periods

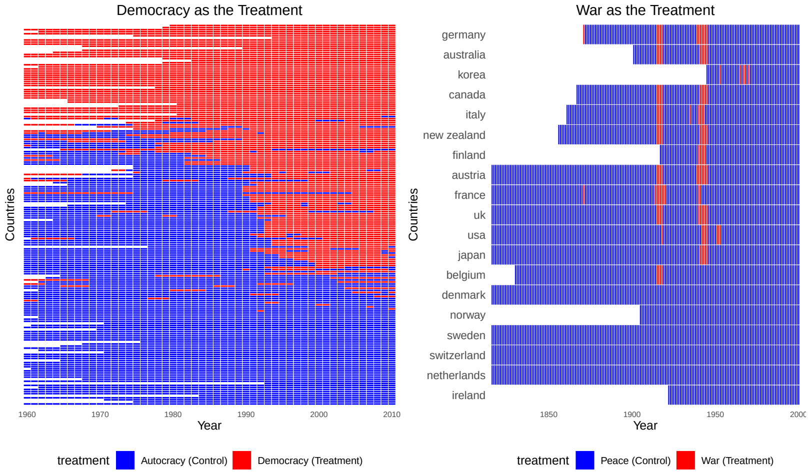
DiD with Staggered Adoption

The $N \times T$ matrix $\boldsymbol{W}$ with typical element W_{it} has the form:

$$\boldsymbol{W}_{N\times T} = \left(\begin{array}{cccccccr}
1 & 2 & 3 & 4 & \dots & T & & \text{(time period)} \\
 & & & & & & & \\
0 & 0 & 0 & 0 & \dots & 0 & & \text{(never adopter)} \\
0 & 0 & 0 & 0 & \dots & 1 & & \text{(late adopter)} \\
0 & 0 & 0 & 0 & \dots & 1 & & \\
0 & 0 & 1 & 1 & \dots & 1 & & \\
0 & 0 & 1 & 1 & \dots & 1 & & \text{(medium adopter)} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & & \\
0 & 1 & 1 & 1 & \dots & 1 & & \text{(early adopter)}
\end{array} \right).$$

Source: Athey & Imbens (2021)

Democracy and Growth by Acemoglu et al.(Left)
War and Inheritance Taxation by Scheve and Stasavage (Right)



Outline of the Talk

Quantity of Interest and Assumptions

- Choose number of lags $L = 2, \ldots$, for confounder adjustment

Quantity of Interest and Assumptions

- Choose number of lags $L = 2, \ldots,$ for confounder adjustment
- Choose number of leads, $F = 0, 1, \ldots,$ for short or long term effects

Quantity of Interest and Assumptions

- Choose number of lags $L = 2, \ldots,$ for confounder adjustment
- Choose number of leads, $F = 0, 1, \ldots,$ for short or long term effects
- Average Treatment Effect of Policy Change for the Treated (ATT):

$$\delta(F, L) = \mathbb{E}\left\{ Y_{i,t+F}\left(X_{it} = 1, X_{i,t-1} = 0, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) - Y_{i,t+F}\left(X_{it} = 0, X_{i,t-1} = 0, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) \mid X_{it} = 1, X_{i,t-1} = 0 \right\}$$

Quantity of Interest and Assumptions

- Choose number of lags $L = 2, \ldots,$ for confounder adjustment
- Choose number of leads, $F = 0, 1, \ldots,$ for short or long term effects
- Average Treatment Effect of Policy Change for the Treated (ATT):

$$\delta(F, L) = \mathbb{E}\left\{ Y_{i,t+F}\left(X_{it} = 1, X_{i,t-1} = 0, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) - Y_{i,t+F}\left(X_{it} = 0, X_{i,t-1} = 0, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) \mid X_{it} = 1, X_{i,t-1} = 0 \right\}$$

- Assumptions:

Quantity of Interest and Assumptions

- Choose number of lags $L = 2, \ldots,$ for confounder adjustment
- Choose number of leads, $F = 0, 1, \ldots,$ for short or long term effects
- Average Treatment Effect of Policy Change for the Treated (ATT):

$$\delta(F, L) = \mathbb{E}\left\{ Y_{i,t+F}\left(X_{it} = 1, X_{i,t-1} = 0, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) - Y_{i,t+F}\left(X_{it} = 0, X_{i,t-1} = 0, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) \mid X_{it} = 1, X_{i,t-1} = 0 \right\}$$

- Assumptions:
  1. No spillover effect

Quantity of Interest and Assumptions

- Choose number of lags $L = 2, \ldots,$ for confounder adjustment
- Choose number of leads, $F = 0, 1, \ldots,$ for short or long term effects
- Average Treatment Effect of Policy Change for the Treated (ATT):

$$\delta(F, L) = \mathbb{E}\left\{ Y_{i,t+F}\left(X_{it} = 1, X_{i,t-1} = 0, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) - Y_{i,t+F}\left(X_{it} = 0, X_{i,t-1} = 0, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) \mid X_{it} = 1, X_{i,t-1} = 0 \right\}$$

- Assumptions:
  1. No spillover effect
  2. Limited carryover effect (up to L time periods)

Quantity of Interest and Assumptions

- Choose number of lags $L = 2, \ldots,$ for confounder adjustment
- Choose number of leads, $F = 0, 1, \ldots,$ for short or long term effects
- Average Treatment Effect of Policy Change for the Treated (ATT):

$$\delta(F, L) = \mathbb{E}\left\{ Y_{i,t+F}\left(X_{it} = 1, X_{i,t-1} = 0, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) - Y_{i,t+F}\left(X_{it} = 0, X_{i,t-1} = 0, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) \mid X_{it} = 1, X_{i,t-1} = 0 \right\}$$

- Assumptions:
  1. No spillover effect
  2. Limited carryover effect (up to L time periods)
  3. Parallel trend after conditioning:

$$\begin{aligned} &\mathbb{E}[Y_{i,t+F}\left(X_{it} = X_{i,t-1} = 0, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) - Y_{i,t-1} \\ &\qquad \mid X_{it} = 1, X_{i,t-1} = 0, \{X_{i,t-\ell}, Y_{i,t-\ell}\}_{\ell=2}^{L}, \{\mathbf{Z}_{i,t-\ell}\}_{\ell=0}^{L}] \\ = \; &\mathbb{E}[Y_{i,t+F}\left(X_{it} = X_{i,t-1} = 0, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) - Y_{i,t-1} \\ &\qquad \mid X_{it} = 0, X_{i,t-1} = 0, \{X_{i,t-\ell}, Y_{i,t-\ell}\}_{\ell=2}^{L}, \{\mathbf{Z}_{i,t-\ell}\}_{\ell=0}^{L}] \end{aligned}$$

Quantity of Interest and Assumptions

Average Treatment Effect of Policy Reversal for the Treated (ART):

$$\xi(F,L) = \mathbb{E}\left\{ Y_{i,t+F}\left(X_{it}=0, X_{i,t-1}=1, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) - \right.$$
$$\left. Y_{i,t+F}\left(X_{it}=1, X_{i,t-1}=1, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) \mid X_{it}=0, X_{i,t-1}=1 \right\}$$

Assumption:

$$\begin{aligned}
&\mathbb{E}[Y_{i,t+F}\left(X_{it}=X_{i,t-1}=1, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) - Y_{i,t-1} \\
&\qquad \mid X_{it}=0, X_{i,t-1}=1, \{X_{i,t-\ell}, Y_{i,t-\ell}\}_{\ell=2}^{L}, \{\mathbf{Z}_{i,t-\ell}\}_{\ell=0}^{L}] \\
= \ &\mathbb{E}[Y_{i,t+F}\left(X_{it}=X_{i,t-1}=1, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) - Y_{i,t-1} \\
&\qquad \mid X_{it}=1, X_{i,t-1}=1, \{X_{i,t-\ell}, Y_{i,t-\ell}\}_{\ell=2}^{L}, \{\mathbf{Z}_{i,t-\ell}\}_{\ell=0}^{L}]
\end{aligned}$$

Constructing Matched Sets

- Construct a matched set for each treated observation ($X_{it} = 1, X_{i,t-1} = 0$)

Constructing Matched Sets

- Construct a matched set for each treated observation ($X_{it} = 1, X_{i,t-1} = 0$)
- Control units with identical treatment history from time $t - L$ to $t - 1$

Constructing Matched Sets

- Construct a matched set for each treated observation $(X_{it} = 1, X_{i,t-1} = 0)$
- Control units with identical treatment history from time $t - L$ to $t - 1$
- Formal definition:

$$\mathcal{M}_{it} \;=\; \{i' : i' \neq i, X_{i't} = 0, X_{i't'} = X_{it'} \text{ for all } t' = t-1, \ldots, t-L\}$$

Constructing Matched Sets

- Construct a matched set for each treated observation $(X_{it} = 1, X_{i,t-1} = 0)$
- Control units with identical treatment history from time $t - L$ to $t - 1$
- Formal definition:

$$\mathcal{M}_{it} \quad = \quad \{i' : i' \neq i, X_{i't} = 0, X_{i't'} = X_{it'} \text{ for all } t' = t-1, \dots, t-L\}$$

- Some treated observations have no matched control
  $\rightsquigarrow$ change the quantity of interest by dropping them

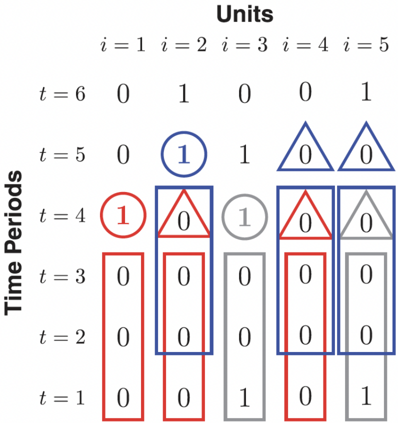
Constructing Matched Sets

- Construct a matched set for each treated observation $(X_{it} = 1, X_{i,t-1} = 0)$
- Control units with identical treatment history from time $t - L$ to $t - 1$
- Formal definition:

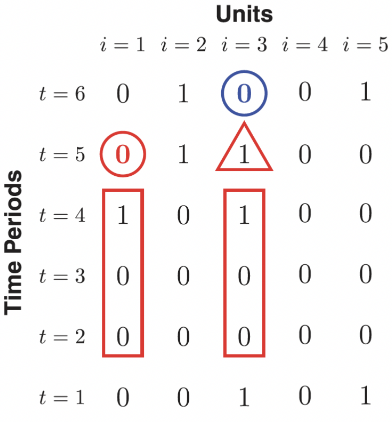
$$\mathcal{M}_{it} = \{i' : i' \neq i, X_{i't} = 0, X_{i't'} = X_{it'} \text{ for all } t' = t-1, \ldots, t-L\}$$

- Some treated observations have no matched control
  $\rightsquigarrow$ change the quantity of interest by dropping them
- Similar to the risk set of Li et al. (2001) but we do not exclude those who already receive the treatment

Matched Sets for ATT and ART



(a) Matched Sets for ATT



(b) Matched Sets for ART

An Example of Matched Set

	Country	Year	Democracy	logGDP	Population	Trade
1	Argentina	1974	**1**	888.20	29.11	14.45
2	Argentina	1975	**1**	886.53	29.11	12.61
3	Argentina	1976	**0**	882.91	29.15	12.11
4	Argentina	1977	**0**	888.09	29.32	15.15
5	**Argentina**	**1978**	**0**	**881.99**	**29.57**	**15.54**
6	Argentina	1979	0	890.24	29.85	15.93
7	Argentina	1980	0	892.81	30.12	12.23
8	Argentina	1981	0	885.43	30.33	11.39
9	Argentina	1982	0	878.82	30.62	13.40
10	Thailand	1974	**1**	637.24	43.32	37.76
11	Thailand	1975	**1**	639.51	42.90	41.63
12	Thailand	1976	**0**	645.97	42.44	42.33
13	Thailand	1977	**0**	653.02	41.92	43.21
14	**Thailand**	**1978**	**1**	**660.57**	**41.39**	**42.66**
15	Thailand	1979	1	663.64	40.82	45.27
16	Thailand	1980	1	666.57	40.18	46.69
17	Thailand	1981	1	670.27	39.44	53.40
18	Thailand	1982	1	673.52	38.59	54.22

Another Example of Matched Set (Wang 2022)

	County	Time	Inspection
1	Youxi	Sep 2014	**0**
2	Youxi	Oct 2014	**1**
3	Youxi	Nov 2014	**1**
4	Youxi	Dec 2014	**0**
5	Youxi	Jan 2015	**1**
6	Guangze	Sep 2014	**0**
7	Guangze	Oct 2014	**1**
8	Guangze	Nov 2014	**1**
9	Guangze	Dec 2014	**0**
10	Guangze	Jan 2015	**0**
11	Mengla	Sep 2014	**0**
12	Mengla	Oct 2014	**1**
13	Mengla	Nov 2014	**1**
14	Mengla	Dec 2014	**0**
15	Mengla	Jan 2015	**0**

Refining Matched Sets

- Make additional adjustments for past outcomes and confounders

Refining Matched Sets

- Make additional adjustments for past outcomes and confounders
- Use any matching or weighting method

Refining Matched Sets

- Make additional adjustments for past outcomes and confounders
- Use any matching or weighting method
- Mahalanobis distance matching:

Refining Matched Sets

- Make additional adjustments for past outcomes and confounders
- Use any matching or weighting method
- Mahalanobis distance matching:
  1. Compute the distance between treated and matched control obs.

$$S_{it}(i') \;=\; \frac{1}{L}\sum_{\ell=1}^{L}\sqrt{(\mathbf{V}_{i,t-\ell}-\mathbf{V}_{i',t-\ell})^{\top}\mathbf{\Sigma}_{i,t-\ell}^{-1}(\mathbf{V}_{i,t-\ell}-\mathbf{V}_{i',t-\ell})}$$

where $\mathbf{V}_{it'} = (Y_{it'}, \mathbf{Z}_{i,t'+1}^{\top})^{\top}$ and $\mathbf{\Sigma}_{it'} = \mathrm{Cov}(\mathbf{V}_{it'})$

Refining Matched Sets

- Make additional adjustments for past outcomes and confounders
- Use any matching or weighting method
- Mahalanobis distance matching:
  1. Compute the distance between treated and matched control obs.

$$S_{it}(i') \; = \; \frac{1}{L}\sum_{\ell=1}^{L}\sqrt{(\mathbf{V}_{i,t-\ell}-\mathbf{V}_{i',t-\ell})^\top \mathbf{\Sigma}_{i,t-\ell}^{-1}(\mathbf{V}_{i,t-\ell}-\mathbf{V}_{i',t-\ell})}$$

     where $\mathbf{V}_{it'} = (Y_{it'}, \mathbf{Z}_{i,t'+1}^\top)^\top$ and $\mathbf{\Sigma}_{it'} = \mathrm{Cov}(\mathbf{V}_{it'})$
  2. Match the most similar J matched control observations

Refining Matched Sets

- Make additional adjustments for past outcomes and confounders
- Use any matching or weighting method
- Mahalanobis distance matching:
  1. Compute the distance between treated and matched control obs.

$$S_{it}(i') = \frac{1}{L}\sum_{\ell=1}^{L}\sqrt{(\mathbf{V}_{i,t-\ell}-\mathbf{V}_{i',t-\ell})^\top \mathbf{\Sigma}_{i,t-\ell}^{-1}(\mathbf{V}_{i,t-\ell}-\mathbf{V}_{i',t-\ell})}$$

     where $\mathbf{V}_{it'} = (Y_{it'}, \mathbf{Z}_{i,t'+1}^\top)^\top$ and $\mathbf{\Sigma}_{it'} = \mathrm{Cov}(\mathbf{V}_{it'})$
  2. Match the most similar J matched control observations
- Propensity score matching or weighting:

Refining Matched Sets

- Make additional adjustments for past outcomes and confounders
- Use any matching or weighting method
- Mahalanobis distance matching:
  1. Compute the distance between treated and matched control obs.

$$S_{it}(i') = \frac{1}{L}\sum_{\ell=1}^{L}\sqrt{(\mathbf{V}_{i,t-\ell} - \mathbf{V}_{i',t-\ell})^\top \boldsymbol{\Sigma}_{i,t-\ell}^{-1}(\mathbf{V}_{i,t-\ell} - \mathbf{V}_{i',t-\ell})}$$

     where $\mathbf{V}_{it'} = (Y_{it'}, \mathbf{Z}_{i,t'+1}^\top)^\top$ and $\boldsymbol{\Sigma}_{it'} = \mathrm{Cov}(\mathbf{V}_{it'})$
  2. Match the most similar J matched control observations
- Propensity score matching or weighting:
  1. Estimate the propensity score

$$e_{it}(\{\mathbf{V}_{i,t-\ell}\}_{\ell=1}^{L}) = \Pr(X_{it} = 1 \mid \{\mathbf{V}_{i,t-\ell}\}_{\ell=1}^{L})$$

Refining Matched Sets

- Make additional adjustments for past outcomes and confounders
- Use any matching or weighting method
- Mahalanobis distance matching:
  1. Compute the distance between treated and matched control obs.

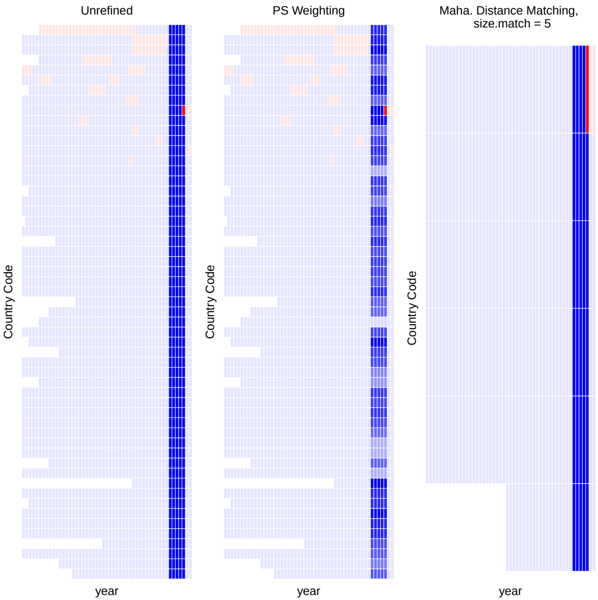
$$S_{it}(i') = \frac{1}{L}\sum_{\ell=1}^{L}\sqrt{(\mathbf{V}_{i,t-\ell} - \mathbf{V}_{i',t-\ell})^\top \mathbf{\Sigma}_{i,t-\ell}^{-1}(\mathbf{V}_{i,t-\ell} - \mathbf{V}_{i',t-\ell})}$$

     where $\mathbf{V}_{it'} = (Y_{it'}, \mathbf{Z}_{i,t'+1}^\top)^\top$ and $\mathbf{\Sigma}_{it'} = \mathrm{Cov}(\mathbf{V}_{it'})$
  2. Match the most similar J matched control observations
- Propensity score matching or weighting:
  1. Estimate the propensity score

$$e_{it}(\{\mathbf{V}_{i,t-\ell}\}_{\ell=1}^{L}) = \Pr(X_{it} = 1 \mid \{\mathbf{V}_{i,t-\ell}\}_{\ell=1}^{L})$$

  2. Weight each matched control observation by its inverse
     - could then use scores as weights or pick the closest J, each of which receives equal weight

An Example of Refinement: Bhutan in 2008



The Multi-period Difference-in-Differences Estimator

- Compute the weighted average of difference-in-differences among matched control observations

The Multi-period Difference-in-Differences Estimator

- Compute the weighted average of difference-in-differences among matched control observations
- Weights are based on refinement

The Multi-period Difference-in-Differences Estimator

- Compute the weighted average of difference-in-differences among matched control observations
- Weights are based on refinement
- A "synthetic control" for each treated observation

The Multi-period Difference-in-Differences Estimator

- Compute the weighted average of difference-in-differences among matched control observations
- Weights are based on refinement
- A "synthetic control" for each treated observation

- The Multi-period DiD estimator: $\widehat{\delta}_{att}(F, L) =$

$$\frac{1}{\sum_{i=1}^{N} \sum_{t=L+1}^{T-F} D_{it}} \sum_{i=1}^{N} \sum_{t=L+1}^{T-F} D_{it} \left\{ (Y_{i,t+F} - Y_{i,t-1}) - \sum_{i' \in \mathcal{M}_{it}} w_{it}^{i'} (Y_{i',t+F} - Y_{i',t-1}) \right\}$$

The Multi-period Difference-in-Differences Estimator

- Compute the weighted average of difference-in-differences among matched control observations
- Weights are based on refinement
- A "synthetic control" for each treated observation

- The Multi-period DiD estimator: $\widehat{\delta}_{att}(F, L) =$

$$\frac{1}{\sum_{i=1}^{N}\sum_{t=L+1}^{T-F} D_{it}} \sum_{i=1}^{N}\sum_{t=L+1}^{T-F} D_{it}\left\{(Y_{i,t+F} - Y_{i,t-1}) - \sum_{i' \in \mathcal{M}_{it}} w_{it}^{i'}\left(Y_{i',t+F} - Y_{i',t-1}\right)\right\}$$

- Same with ART

The Multi-period Difference-in-Differences Estimator

- Compute the weighted average of difference-in-differences among matched control observations
- Weights are based on refinement
- A "synthetic control" for each treated observation

- The Multi-period DiD estimator: $\widehat{\delta}_{att}(F, L) =$

$$\frac{1}{\sum_{i=1}^{N}\sum_{t=L+1}^{T-F} D_{it}} \sum_{i=1}^{N}\sum_{t=L+1}^{T-F} D_{it}\left\{(Y_{i,t+F} - Y_{i,t-1}) - \sum_{i' \in \mathcal{M}_{it}} w_{it}^{i'}\left(Y_{i',t+F} - Y_{i',t-1}\right)\right\}$$

- Same with ART
- Block bootstrap for standard error calculation

An Example of Refinement

	Country	Year	Democracy	logGDP	Population	Trade	**Weight**
1	Argentina	1979	**0**	890.24	29.85	15.93	1.00
2	Argentina	1980	**0**	892.81	30.12	12.23	1.00
3	Argentina	1981	**0**	885.43	30.33	11.39	1.00
4	Argentina	1982	**0**	878.82	30.62	13.40	1.00
5	**Argentina**	**1983**	**1**	**881.09**	**30.75**	**16.46**	**1.00**
6	Argentina	1984	1	881.76	30.77	15.67	1.00
7	Mali	1979	**0**	542.02	43.80	31.18	0.26
8	Mali	1980	**0**	535.65	43.96	41.82	0.26
9	Mali	1981	**0**	529.10	44.07	41.92	0.26
10	Mali	1982	**0**	522.25	44.45	42.53	0.26
11	**Mali**	**1983**	**0**	**524.84**	**44.74**	**43.65**	**0.26**
12	Mali	1984	0	527.13	44.95	45.92	0.26
13	Chad	1979	**0**	506.71	44.61	44.80	0.27
14	Chad	1980	**0**	498.36	44.84	45.75	0.27
15	Chad	1981	**0**	497.18	45.07	51.58	0.27
16	Chad	1982	**0**	500.07	45.44	43.97	0.27
17	**Chad**	**1983**	**0**	**512.20**	**45.76**	**29.22**	**0.27**
18	Chad	1984	0	511.63	46.04	29.91	0.27
19	Uruguay	1979	**0**	858.39	27.23	41.51	0.47
20	Uruguay	1980	**0**	863.39	27.04	37.99	0.47
21	Uruguay	1981	**0**	864.28	26.93	36.20	0.47
22	Uruguay	1982	**0**	853.36	26.86	35.84	0.47
23	**Uruguay**	**1983**	**0**	**841.87**	**26.83**	**33.36**	**0.47**
24	Uruguay	1984	0	840.08	26.82	42.98	0.47

Checking Covariate Balance

- Balance for covariate j at time $t - \ell$ in each matched set:

$$B_{it}(j, \ell) \; = \; \frac{V_{i,t-\ell,j} - \sum_{i' \in \mathcal{M}_{it}} w_{it}^{i'} \; V_{i',t-\ell,j}}{\sqrt{\frac{1}{N_1 - 1} \sum_{i'=1}^{N} \sum_{t'=L+1}^{T-F} D_{i't'}(V_{i',t'-\ell,j} - \overline{V}_{t'-\ell,j})^2}}$$

Checking Covariate Balance

- Balance for covariate j at time $t-\ell$ in each matched set:

$$B_{it}(j,\ell) \;=\; \frac{V_{i,t-\ell,j} - \sum_{i' \in \mathcal{M}_{it}} w_{it}^{i'} \; V_{i',t-\ell,j}}{\sqrt{\frac{1}{N_1-1}\sum_{i'=1}^{N}\sum_{t'=L+1}^{T-F} D_{i't'}(V_{i',t'-\ell,j} - \overline{V}_{t'-\ell,j})^2}}$$

- Average this measure across all treated observations:

$$\overline{B}(j,\ell) \;=\; \frac{1}{N_1}\sum_{i=1}^{N}\sum_{t=L+1}^{T-F} D_{it}B_{it}(j,\ell)$$

Outline of the Talk

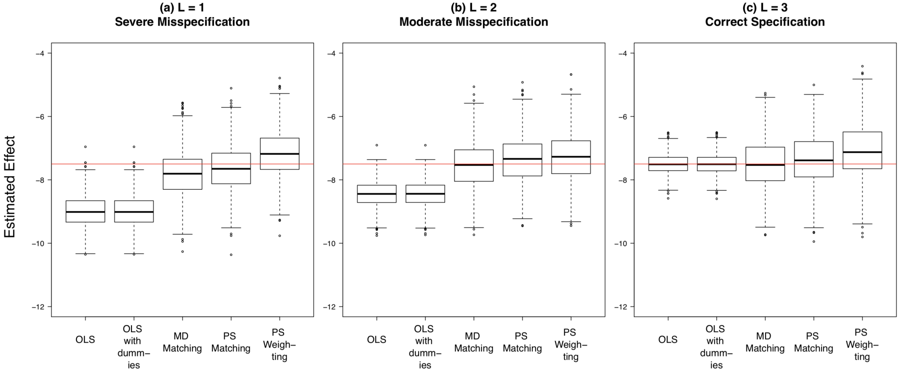
Simulation Studies

- Make simulations realistic by using the Acemoglu et al. data
- Balanced TSCS data with $N = 162$ and $T = 51$
- Impute missing values in the original data set
- Treatment and outcome variables:

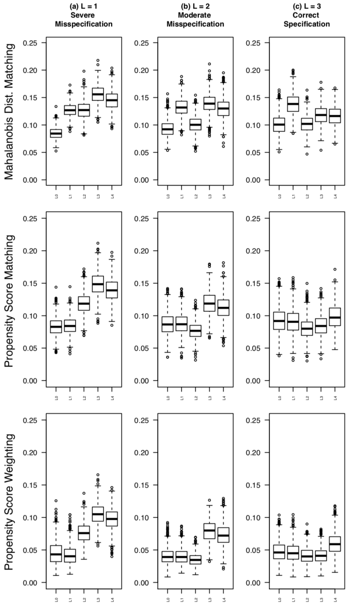
$$
\begin{aligned}
X_{it} &\sim \text{Benoulli}\left(\text{logit}^{-1}\left\{\tilde{\alpha}_i + \tilde{\gamma}_t + \sum_{\ell=1}^{L} \tilde{\beta}_\ell^\top X_{i,t-\ell} + \right.\right. \\
&\qquad\qquad \left.\left. \sum_{\ell=0}^{L}\left(\tilde{\zeta}_\ell^\top \mathbf{Z}_{i,t-\ell} + \tilde{\phi}_\ell^\top [\mathbf{Z}_{i,t-\ell}^{(1)} : \mathbf{Z}_{i,t-\ell}^{(3)}]\right)\right\}\right) \\
Y_{it} &= \alpha_i + \gamma_t + \sum_{\ell=0}^{L} \beta_\ell^\top X_{i,t-\ell} + \sum_{\ell=0}^{L}\left(\zeta_\ell^\top \mathbf{Z}_{i,t-\ell} + \phi_\ell^\top [\mathbf{Z}_{i,t-\ell}^{(1)} : \mathbf{Z}_{i,t-\ell}^{(3)}]\right) + \epsilon_{it}
\end{aligned}
$$

where $\epsilon \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$, and $\mathbf{Z}_{i,t-\ell} = (\mathbf{Z}_{i,t-\ell}^{\top(1)}, \mathbf{Z}_{i,t-\ell}^{\top(2)}, \mathbf{Z}_{i,t-\ell}^{\top(3)})^\top$

Robustness to Model Misspecification



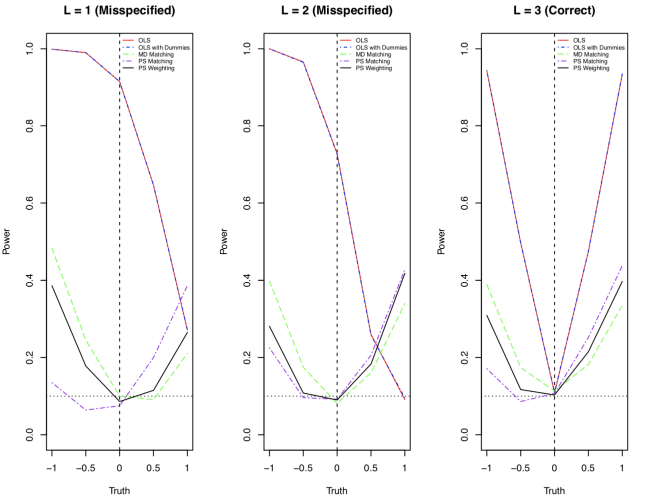
Balance Metrics in the Simulations



Average absolute standardized mean difference across covariates for each L

Desirable balance for moderate misspecification and correct specification

Bias-variance Tradeoff



Outline of the Talk

Empirical Application

- $L = 4$ or 1 and $F = 0, 1, 2, 3, 4$

Empirical Application

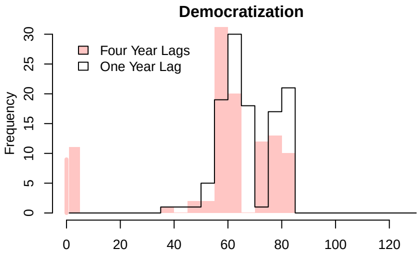
- $L = 4$ or 1 and $F = 0, 1, 2, 3, 4$
- We consider democratization and authoritarian reversal

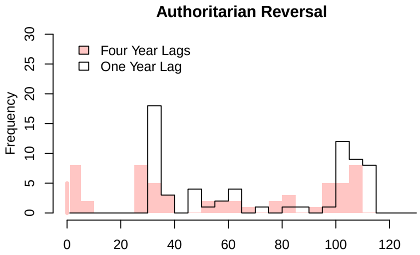
Empirical Application

- $L = 4$ or 1 and $F = 0, 1, 2, 3, 4$
- We consider democratization and authoritarian reversal
- Examine the number of matched control units

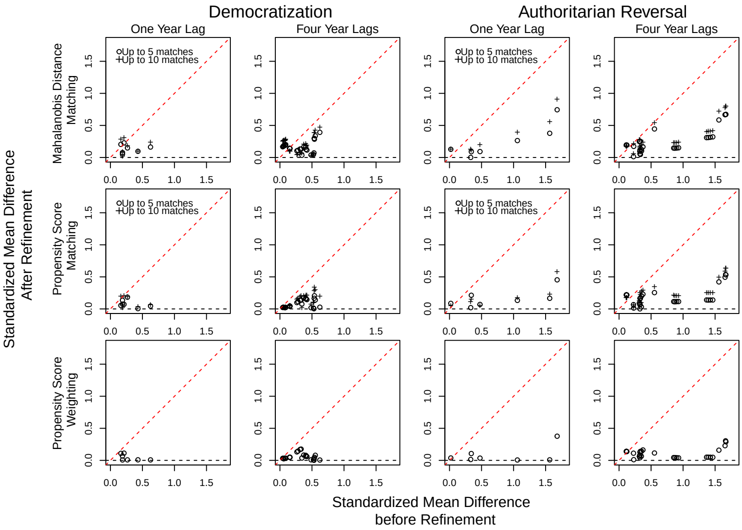
Empirical Application

- $L = 4$ or 1 and $F = 0, 1, 2, 3, 4$
- We consider democratization and authoritarian reversal
- Examine the number of matched control units
- 9 (5) treated observations have no matched control

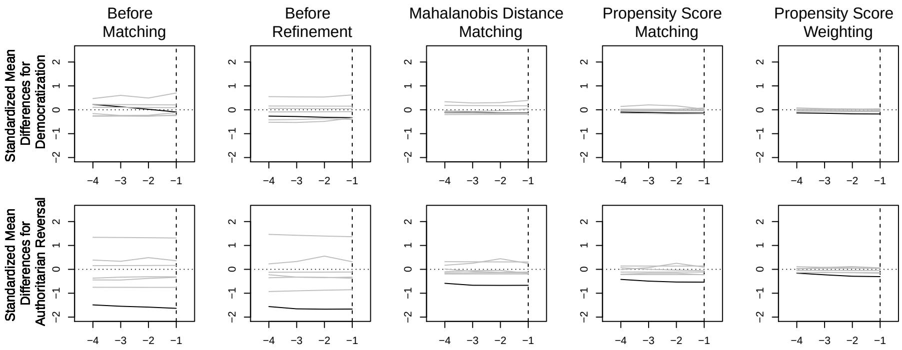




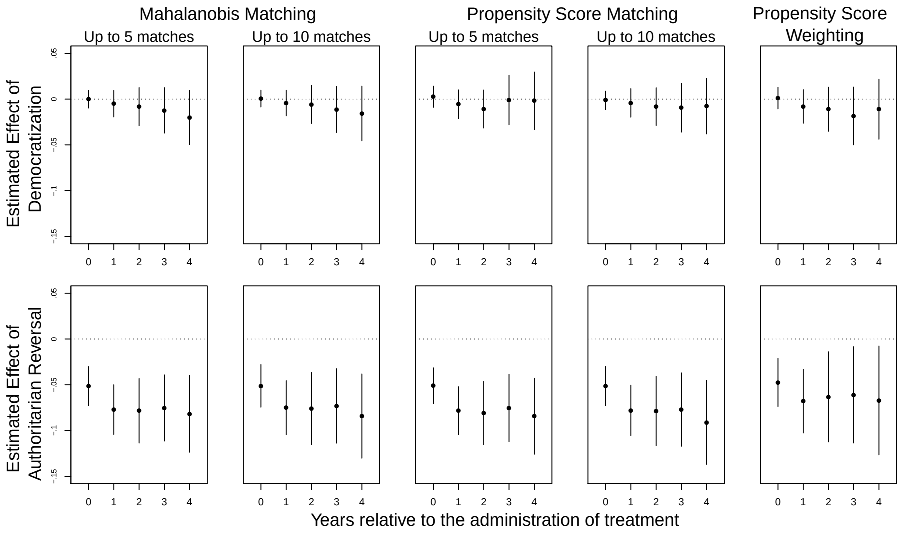
Improved Covariate Balance



Improved Covariate Balance



Estimated Causal Effects



Outline of the Talk

Concluding Remarks

- Matching as transparent and simple methods for causal inference

Concluding Remarks

- Matching as transparent and simple methods for causal inference
- Yet, matching has not been applied to time-series cross-sectional data

Concluding Remarks

- Matching as transparent and simple methods for causal inference
- Yet, matching has not been applied to time-series cross-sectional data

- We propose a matching framework for TSCS data

Concluding Remarks

- Matching as transparent and simple methods for causal inference
- Yet, matching has not been applied to time-series cross-sectional data

- We propose a matching framework for TSCS data
    1. construct matched sets

Concluding Remarks

- Matching as transparent and simple methods for causal inference
- Yet, matching has not been applied to time-series cross-sectional data

- We propose a matching framework for TSCS data
  1. construct matched sets
  2. refine matched sets

Concluding Remarks

- Matching as transparent and simple methods for causal inference
- Yet, matching has not been applied to time-series cross-sectional data

- We propose a matching framework for TSCS data
  1. construct matched sets
  2. refine matched sets
  3. compute difference-in-differences estimator

Concluding Remarks

- Matching as transparent and simple methods for causal inference
- Yet, matching has not been applied to time-series cross-sectional data

- We propose a matching framework for TSCS data
  1. construct matched sets
  2. refine matched sets
  3. compute difference-in-differences estimator
- Checking covariates and computing standard errors

Concluding Remarks

- Matching as transparent and simple methods for causal inference
- Yet, matching has not been applied to time-series cross-sectional data

- We propose a matching framework for TSCS data
  1. construct matched sets
  2. refine matched sets
  3. compute difference-in-differences estimator
- Checking covariates and computing standard errors
- R package PanelMatch implements all of these methods

Concluding Remarks

- Matching as transparent and simple methods for causal inference
- Yet, matching has not been applied to time-series cross-sectional data

- We propose a matching framework for TSCS data
  1. construct matched sets
  2. refine matched sets
  3. compute difference-in-differences estimator
- Checking covariates and computing standard errors
- R package PanelMatch implements all of these methods

- Ongoing research: possible spillover effects and continuous treatment

References

- Imai, Kim, and Wang. "Matching Methods for Causal Inference with Time-Series Cross-Sectional Data." *American Journal of Political Science*, Forthcoming

References

- Imai, Kim, and Wang. "Matching Methods for Causal Inference with Time-Series Cross-Sectional Data." *American Journal of Political Science*, Forthcoming

- Related papers:
  1. Imai and Kim. (2019) "When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?." *American Journal of Political Science*, Vol. 63, No. 2, pp. 467-490.
  2. Imai and Kim. (2021) "On the Use of Two-way Fixed Effects Regression Models for Causal Inference with Panel Data." *Political Analysis*, Vol. 29, No. 3, pp. 405-415.

References

- Imai, Kim, and Wang. "Matching Methods for Causal Inference with Time-Series Cross-Sectional Data." *American Journal of Political Science*, Forthcoming

- Related papers:
  1. Imai and Kim. (2019) "When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?." *American Journal of Political Science*, Vol. 63, No. 2, pp. 467-490.
  2. Imai and Kim. (2021) "On the Use of Two-way Fixed Effects Regression Models for Causal Inference with Panel Data." *Political Analysis*, Vol. 29, No. 3, pp. 405-415.

- Send comments and suggestions to:

  erik.wang@anu.edu.au