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

Kosuke Imai In Song Kim Erik Wang

Harvard MIT Princeton

Keynote Talk at
Latin American Political Methodology Meeting

November 9, 2018

Motivation and Overview

- 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
- 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 estimator
- Provide a model-based standard error
- Develop an open-source software package PanelMatch
- Empirical applications:
  - Democracy and economic growth (Acemoglu et al.)
  - Interstate war and inheritance tax (Scheve & Stasavage)

Democracy and Economic Growth

- Acemoglu et al. (2017): an up-to-date empirical study of the long-standing question in political economy
- TSCS data set: 184 countries from 1960 to 2010
- Dynamic linear regression model with fixed effects:

$$Y_{it} \quad = \quad \alpha_i + \gamma_t + \beta X_{it} + \sum_{\ell=1}^{4} \left\{ \rho_\ell Y_{i,t-\ell} + \zeta_\ell^\top \mathbf{Z}_{i,t-\ell} \right\} + \epsilon_{it}$$

  - X_{it}: binary democracy indicator
  - Y_{it}: log real GDP per capita
  - $\mathbf{Z}_{it}$: time-varying covariates (population, trade, social unrest, etc.)

- Sequential exogeneity assumption:

$$\mathbb{E}(\epsilon_{it} \mid \{Y_{it'}\}_{t'=1}^{t-1}, \{X_{it'}\}_{t'=1}^{t}, \{\mathbf{Z}_{it'}\}_{t'=1}^{t-1}, \alpha_i, \gamma_t) \ = \ 0$$

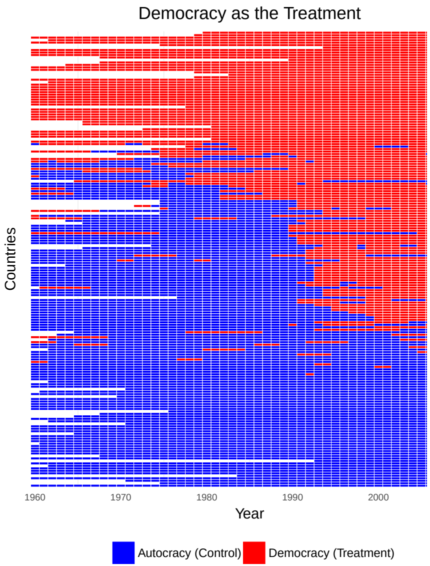
- Nickell bias $\rightsquigarrow$ GMM estimation with instruments (Arellano & Bond)

Regression Results

ATE ($\hat{\beta}$)	0.787	0.875	0.666	0.917
	(0.226)	(0.374)	(0.307)	(0.461)
$\hat{\rho}_1$	1.238	1.204	1.100	1.046
	(0.038)	(0.041)	(0.042)	(0.043)
$\hat{\rho}_2$	-0.207	-0.193	-0.133	-0.121
	(0.043)	(0.045)	(0.041)	(0.038)
$\hat{\rho}_3$	-0.026	-0.028	0.005	0.014
	(0.028)	(0.028)	(0.030)	(0.029)
$\hat{\rho}_4$	-0.043	-0.036	0.003	-0.018
	(0.017)	(0.020)	(0.024)	(0.023)
country FE	Yes	Yes	Yes	Yes
time FE	Yes	Yes	Yes	Yes
time trends	No	No	No	No
covariates	No	No	Yes	Yes
estimation	OLS	GMM	OLS	GMM
N	6,336	4,416	6,161	4,245

Treatment Variation Plot

- Regression models does not tell us where the variation comes from
- Estimation of counterfactual outcomes requires comparison between treated and control observations
- Identification strategy:
    - within-unit over-time variation
    - within-time across-units variation



War and Taxation

- Inheritance tax plays a central role in wealth accumulations and income inequality
- Scheve and Stasavage (2012): war increases inheritance taxation
- TSCS Data: 19 countries over 185 years from 1816 to 2000
- Static model:

$$Y_{it} \;=\; \alpha_i + \gamma_t + \beta X_{i,t-1} + \delta^\top \mathbf{Z}_{i,t-1} + \lambda_i t + \epsilon_{it}$$

  - $X_{i,t-1}$: interstate war for country i in year $t-1$
  - Y_{it}: top rate of inheritance tax
  - $\mathbf{Z}_{i,t-1}$: time-varying covariates (leftist executive, a binary variable for the universal male suffrage, and logged real GDP per capita)
  - Strict exogeneity:

$$\mathbb{E}(\epsilon_{it} \mid \mathbf{X}_i, \mathbf{Z}_i, \alpha_i, \gamma_t, \lambda_i) \;=\; 0$$

  where $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iT})$ and $\mathbf{Z}_i = (\mathbf{Z}_{i1}^\top, \mathbf{Z}_{i2}^\top, \ldots, \mathbf{Z}_{iT}^\top)^\top$

- Dynamic model without country fixed effects:

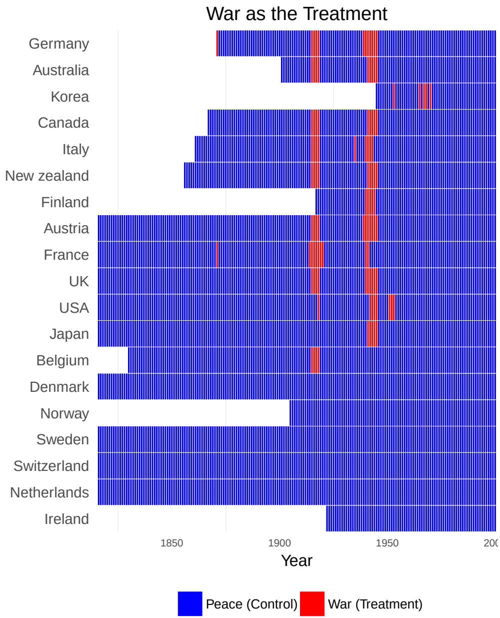
$$Y_{it} = \gamma_t + \beta X_{i,t-1} + \rho Y_{i,t-1} + \delta \mathbf{Z}_{i,t-1} + \lambda_i t + \epsilon_{it}$$

where the strict exogeneity assumption is now given by,

$$\mathbb{E}(\epsilon_{it} \mid \mathbf{X}_i, \mathbf{Z}_i, Y_{i,t-1}, \gamma_t, \lambda_i) = 0$$

- Regression results:

ATE ($\hat{\beta}$)	6.775 (2.392)	1.745 (0.729)	5.970 (2.081)	1.636 (0.757)
$\hat{\rho}_1$		0.908 (0.014)		0.904 (0.014)
country FE	Yes	No	Yes	No
time FE	Yes	Yes	Yes	Yes
time trends	Yes	Yes	Yes	Yes
covariates	No	No	Yes	Yes
N	2,780	2,537	2,779	2,536



- Treatment is concentrated in a few years
- How should we estimate counterfactual outcomes?

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):

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

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

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

Constructing Matched Sets

- Control units with identical treatment history from time $t - L$ to $t - 1$
- Construct a matched set for each treated observation
- 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

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

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 weighting:
  1. Estimate 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

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

The 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 DiD estimator:

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

- Equivalent to the weighted two-way fixed effects estimator:

$$\operatorname*{argmin}_{\beta} \sum_{i=1}^{N}\sum_{t=1}^{T} W_{it}\{(Y_{it}-\overline{Y}_i^{*}-\overline{Y}_t^{*}+\overline{Y}^{*})-\beta(X_{it}-\overline{X}_i^{*}-\overline{X}_t^{*}+\overline{X}^{*})\}^2$$

Checking Covariate Balance and Computing Standard Error

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

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

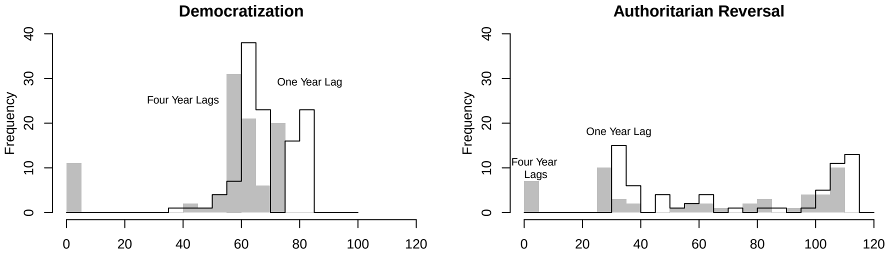
- Average this measure across all treated observations:

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

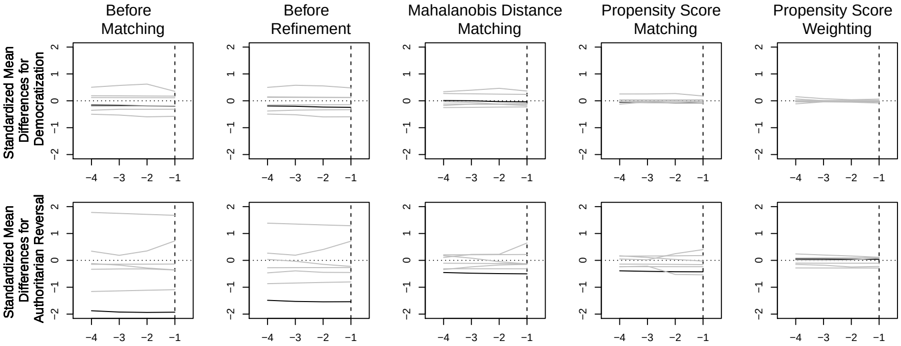
- Standard error calculation $\rightsquigarrow$ consider weight as a covariate
    1. Block bootstrap
    2. Model-based cluster robust standard error within the GMM framework

Empirical Application (1)

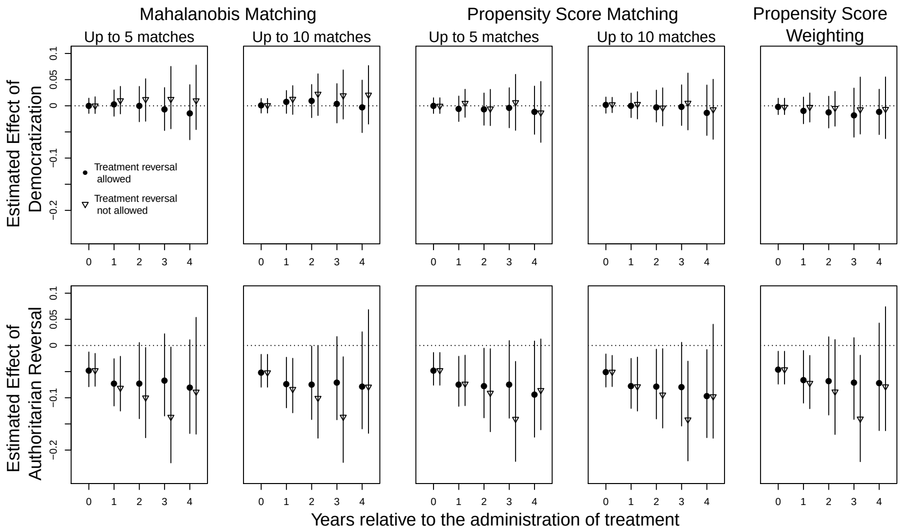
- ATT with $L = 4$ and $F = 1, 2, 3, 4$
- We consider democratization and authoritarian reversal
- Examine the number of matched control units
- 18 (13) treated observations have no matched control



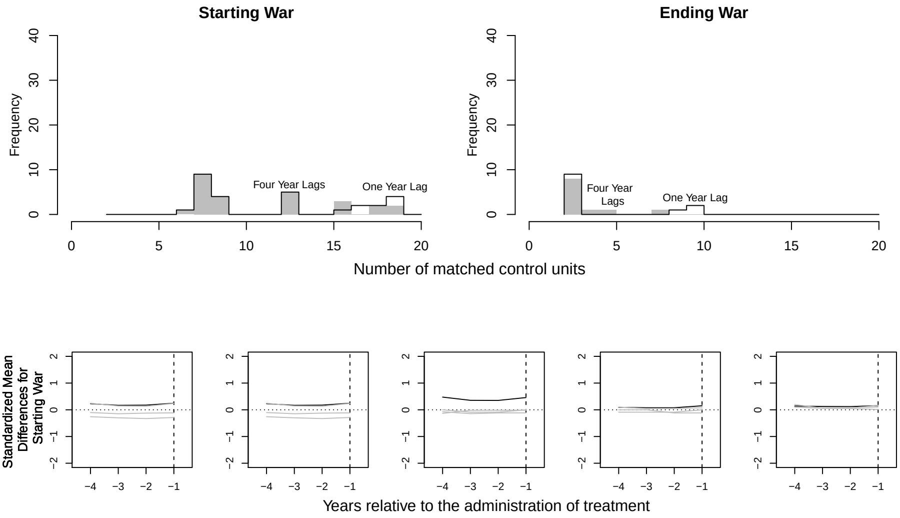
Improved Covariate Balance



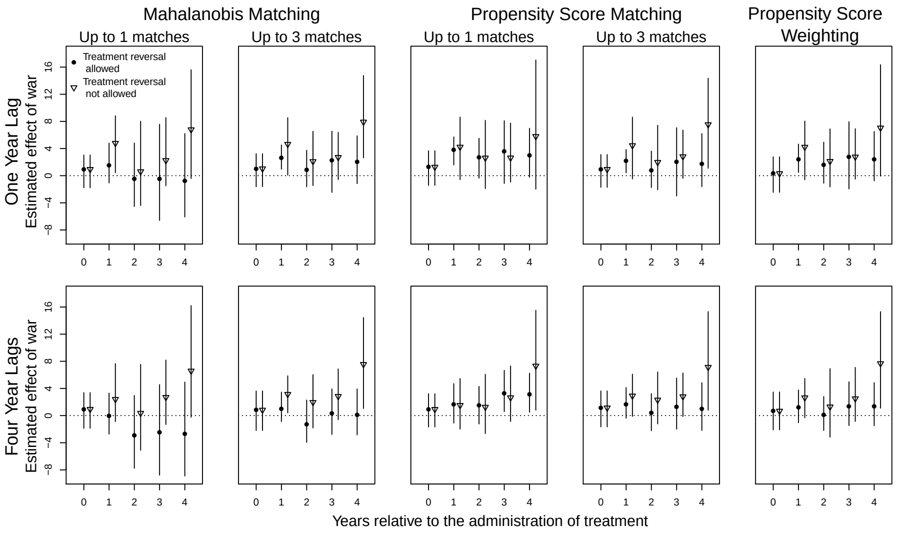
Estimated Causal Effects



Empirical Application (2)



Estimated Causal Effects



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

- Future research: addressing possible spillover effects