DOUBLY WEIGHTED CAUSAL PANEL ESTIMATORS

Susan Athey – Stanford University

Guido Imbens – Stanford University

Zhaonan Qu – Stanford University

Davide Viviano– Harvard University

Barcelona, August 28th, 2023

SUMMARY

- Many Synthetic Control type estimators
  - Are limited to block assignment mechanisms
  - Don't exploit time ordering
  - Use balancing weights rather than distance weights.
- We propose a new estimator, Doubly Weighted Causal Panel (DWCP) estimator
  - More than competitive with SC, competitive with SDID
  - More generally applicable
  - Uses time ordering explicitly.

RMSE COMPARISONS WITH BLOCK ASSIGNMENT

	State Earnings	Country GDP	State Smoking
	N = 50, *T* = 40	*N* = 111, *T* = 48	*N* = 38, *T* = 31
SC	0.116	0.129	0.232
SDID	0.111	0.029	0.135
MC	0.110 (3)	0.042 (29)	0.264 (3)
TWFE	0.134	0.358	0.446
DWCP-Intercept	0.146 (12,0.7)	0.065 (24,0.4)	0.207 (18,2.1)
DWCP-TWFE	0.107 (2,0.21)	0.031 (0.8,1.6)	0.137 (0.8,1.4)
DWCP–1 Factor	0.104 (0,0.08)	0.031 (0.15,0.3)	0.161 (0.1,0.1)

OUTLINE

1. Set Up
2. Existing Estimators
3. Three Issues with Existing Estimators
   - General Assignment Mechanisms
   - Time Invariance
   - Balancing Weights versus Distance Weights
4. Proposed Method
   - Outcome Model
   - Unit and Time Weights
   - Choosing Tuning Parameters
5. Evidence

SET UP

Outcomes: $Y_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T$

Binary Treatment: $W_{it} \in \{0, 1\}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T.$

Interest in (average) treatment effect.

Primarily focus on case with single treated unit/period:

$$\mathbf{W} = \begin{pmatrix} 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

We need to predict Y_{NT} with $W_{NT} = 1$, as $\hat{Y}_{NT}$, and then

$$\hat{\tau} = Y_{NT} - \hat{Y}_{NT}.$$

WE ILLUSTRATE FINDINGS ON THREE DATA SETS

- State Earnings, $N = 50$, $T = 40$, data on average earnings from CPS, transformed to logarithms, from Arkhangelsky et al (2021).
- Country GDP, $N = 111$, $T = 48$, from Penn World Tables, for particular sample, see Arkhangelsky et al (2021).
- State Smoking, $N = 38$, $T = 31$, from Abadie-Diamond-Hainmueller (2010) (leaving out states with smoking cessation programs)

EXISTING METHODS

1. Two Way Fixed Effects / Difference In Differences
2. Synthetic Control
3. Augmented Synthetic Control
4. Synthetic Difference In Differences
5. Matrix Completion / Factor Models

See surveys by Abadie (2021), Arkhangelsky & Imbens (2023)

SYNTHETIC CONTROL

Abadie-Diamond-Hainmueller (JASA, 2010):

$$\hat{Y}_{NT} = \sum_{i=1}^{N-1} \omega_i Y_{iT}$$

The weights ω are choosen by restricted least squares regression (Douchenko-Imbens, 2016)

$$\min_{\omega:\sum_i \omega_i=1,\omega_i\geq 0} \sum_{t=1}^{T-1} \left(Y_{Nt} - \sum_{j=1}^{N-1} \omega_j Y_{jt} \right)^2$$

Given weights, weighted least squares estimator

$$\min_{\tau,\beta} \sum_{i=1}^{N} \sum_{t=1}^{T} \omega_i \left(Y_{it} - \beta_t - \tau W_{it}\right)^2$$

SYNTHETIC DIFFERENCE IN DIFFERENCES

Arkhangelsky et al (2021): add unit fixed effects and time weights

$$\min_{\tau,\alpha,\beta} \sum_{i=1}^{N} \sum_{t=1}^{T} \omega_i \lambda_t \left(Y_{it} - \alpha_i - \beta_t - \tau W_{it}\right)^2$$

ω_i: synthetic control weights, calculated as before.

λ_t: time weights calculated analogously:

$$\arg \min_{\lambda:\sum_t \lambda_t=1, \lambda_t \geq 0} \sum_{i=1}^{N-1} \left(Y_{iT} - \sum_{s=1}^{T-1} \lambda_s Y_{is} \right)^2$$

MATRIX COMPLETION

Athey et al (2021): Fit a factor model + regularization for rank:

$$\min_{\alpha,\beta,\mathbf{L}} \sum_{i=1}^{N} \sum_{t=1}^{T} (1 - W_{it}) (Y_{it} - \alpha_i - \beta_t - \mathbf{L})^2 + \lambda \|\mathbf{L}\|_{\text{nn}}$$

with nuclear norm penalty on **L**, leading to low-rank factor model:

$$\hat{\mathbf{L}}_{it} = \sum_{r=1}^{R} \delta_{ir} \gamma_{rt}$$

Tuning parameter λ chosen through cross-validation.

THREE ISSUES WITH SYNTHETIC CONTROL AND RELATED ESTIMATORS

- Does not directly generalize to settings with more general assignment patterns, *e.g.*, staggered adoption.
- Does not use time-series information: Early observations get as much weight as recent observations. Does not seem plausible: structure is likely to (possibly slowly) change over time.
- The synthetic control weights are not "local": distance to treated units does not matter, only whether treated unit is in convex hull.

THREE ISSUES WITH SYNTHETIC CONTROL AND RELATED ESTIMATORS

- Does not directly generalize to settings with more general assignment patterns, *e.g.*, staggered adoption.

- Does not use time-series information: Early observations get as much weight as recent observations. Does not seem plausible: structure is likely to (possibly slowly) change over time.

- The synthetic control weights are not “local”: distance to treated units does not matter, only whether treated unit is in convex hull.

THREE ISSUES WITH SYNTHETIC CONTROL AND RELATED ESTIMATORS

- Does not directly generalize to settings with more general assignment patterns, *e.g.*, staggered adoption.
- Does not use time-series information: Early observations get as much weight as recent observations. Does not seem plausible: structure is likely to (possibly slowly) change over time.
- The synthetic control weights are not “local”: distance to treated units does not matter, only whether treated unit is in convex hull.

BLOCK ASSIGNMENT PATTERNS

Synthetic control easily extends to block assignments

Block Assignment $\quad \mathbf{W} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$

Less clear, but still feasible, with staggered adoption

Staggered Adoption $\quad \mathbf{W} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{pmatrix}$

GENERAL ASSIGNMENT PATTERNS

How do we deal with general assignment patterns?

General Assignment $$\mathbf{W} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 1 \\ 1 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 1 & \dots & 1 \end{pmatrix}$$

- If we split **W** in blocks, the blocks will be very small.

IGNORING TIME SERIES INFORMATION, PART I

Should $\mathbf{Y}$ and $\mathbf{Y}'$ (with columns 1 and $T-1$ swapped) lead to the same predictions for $Y_{NT}(0)$?

$$\mathbf{Y} = \begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1,T-1} & Y_{1T} \\ Y_{21} & Y_{22} & \cdots & Y_{2,T-1} & Y_{2T} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Y_{N1} & Y_{N2} & \cdots & Y_{N,T-1} & Y_{NT} \end{pmatrix}$$

$$\mathbf{Y}' = \begin{pmatrix} Y_{1,T-1} & Y_{12} & \cdots & Y_{11} & Y_{1T} \\ Y_{2,T-1} & Y_{22} & \cdots & Y_{21} & Y_{2T} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Y_{N,T-1} & Y_{N2} & \cdots & Y_{N1} & Y_{NT} \end{pmatrix}$$

IGNORING TIME SERIES INFORMATION, PART II

Consider the SC estimator, with the weights estimated on periods $1, \ldots, T_0/2$, vs weights estimated on $T_0/2 + 1, \ldots, T_0$.

Under a factor model both should do equally well. Not true in practice. Second half does systematically better.

Training Period ↓	State Earnings $N = 50, T = 40$	Country GDP $N = 111, T = 48$	State Smoking $N = 38, T = 31$
First Half	0.066	0.492	14.3
Second Half	0.059	0.182	9.3
All	0.057	0.200	10.5
% Improvement First to Second Half	10%	63%	55%

IGNORING TIME SERIES INFORMATION, PART III

- Simple factor model can pick up main variation over time.
- Taking account of time-series patterns can aid in exploiting weak signals.
- Smoothing over time improves precision.

BALANCING VERSUS DISTANCE/KERNEL WEIGHTS, PART I

Suppose, for large c, small ε

$$\mathbf{Y} = \begin{pmatrix} \varepsilon & 0 & 0 & 0 & Y_{1T} \\ 0 & \varepsilon & 0 & 0 & Y_{2T} \\ 0 & 0 & 0 & \varepsilon & Y_{3T} \\ c & c & c & c & Y_{4T} \\ -c & -c & -c & -c & Y_{5T} \\ 0 & 0 & 0 & 0 & Y_{NT} \end{pmatrix}, \quad \boldsymbol{\omega} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \\ \omega_5 \\ \emptyset \end{pmatrix}$$

Distance/kernel weights lead to $\omega_1^K = \omega_2^K = \omega_3^K \approx 1/3$, $\omega_4^K = \omega_5^K \approx 0$.

SC/balancing weights lead to $\omega_1^{SC} = \omega_2^{SC} = \omega_3^{SC} \approx 0$, $\omega_4^{SC} = \omega_5^{SC} \approx 1/2$.

Which weights are more attractive?

BALANCING VERSUS DISTANCE/KERNEL WEIGHTS, PART I

Suppose, for large c, small ε

$$\mathbf{Y} = \begin{pmatrix} \varepsilon & 0 & 0 & 0 & Y_{1T} \\ 0 & \varepsilon & 0 & 0 & Y_{2T} \\ 0 & 0 & 0 & \varepsilon & Y_{3T} \\ c & c & c & c & Y_{4T} \\ -c & -c & -c & -c & Y_{5T} \\ 0 & 0 & 0 & 0 & Y_{NT} \end{pmatrix}, \quad \boldsymbol{\omega} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \\ \omega_5 \\ \emptyset \end{pmatrix}$$

Distance/kernel weights lead to $\omega_1^K = \omega_2^K = \omega_3^K \approx 1/3, \omega_4^K = \omega_5^K \approx 0.$

SC/balancing weights $\omega_1^{\mathrm{SC}} = \omega_2^{\mathrm{SC}} = \omega_3^{\mathrm{SC}} \approx 0, \omega_4^{\mathrm{SC}} = \omega_5^{\mathrm{SC}} \approx 1/2.$

Which weights are more attractive?

BALANCING VERSUS DISTANCE/KERNEL WEIGHTS, PART I

Suppose, for large c, small ε

$$\mathbf{Y} = \begin{pmatrix} \varepsilon & 0 & 0 & 0 & Y_{1T} \\ 0 & \varepsilon & 0 & 0 & Y_{2T} \\ 0 & 0 & 0 & \varepsilon & Y_{3T} \\ c & c & c & c & Y_{4T} \\ -c & -c & -c & -c & Y_{5T} \\ 0 & 0 & 0 & 0 & Y_{NT} \end{pmatrix}, \quad \boldsymbol{\omega} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \\ \omega_5 \\ \emptyset \end{pmatrix}$$

Distance/kernel weights lead to $\omega_1^K = \omega_2^K = \omega_3^K \approx 1/3$, $\omega_4^K = \omega_5^K \approx 0$.

SC/balancing weights $\omega_1^{\mathrm{SC}} = \omega_2^{\mathrm{SC}} = \omega_3^{\mathrm{SC}} \approx 0$, $\omega_4^{\mathrm{SC}} = \omega_5^{\mathrm{SC}} \approx 1/2$.

Which weights are more attractive?

BALANCING VERSUS DISTANCE/KERNEL WEIGHTS, PART II

- Balancing weights (depend on **all** units) ensure that average pretreatment outcomes and covariates are balanced.

$$\omega \text{ solves } \frac{1}{N_0}\sum_i \omega_i(1 - W_i)X_i \approx \frac{1}{N_1}\sum_i W_iX_i$$

- Distance weights (depend only on **pairs** of units) ensure that only units similar to treated unit receive much weight.

  For treated unit i, control unit j:

$$\omega^i_j \propto \text{distance}(X_i, X_j)$$

SYNTHETIC CONTROL, SYNTHETIC DIFFERENCE IN DIFFERENCES, AUGMENTED SYNTHETIC CONTROL, MATRIX COMPLETION

	General Assignment	Time Invariance	Weights (Time)	Weights (Units)
TWFE	Yes	Yes	None	None
Synthetic Control	No	Yes	None	Balancing
Augmented SC	No	Yes	None	Balancing
Synthetic DID	No	Yes	Balancing	Balancing
Matrix Completion	Yes	Yes	None	None

PROPOSAL

Estimate model on control outcomes, using distance-based weights ω_{it}, given an outcome model $g(i, t; \theta)$ with unknown parameter θ:

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{N} \sum_{t=1}^{T} \omega_{it}(1 - W_{it}) \left(Y_{it} - g(i, t; \theta)\right)^2$$

The missing value $Y_{NT}(0)$ is imputed as

$$\hat{Y}_{NT}(0) = g(N, T; \hat{\theta})$$

The estimated treatment effect is

$$\hat{\tau} = Y_{NT} - \hat{Y}_{NT}(0)$$

CANDIDATE OUTCOME MODELS

1. Constant (for illustrative purposes)

$$Y_{it}(0) = \mu + \varepsilon_{it}$$

2. Two Way Fixed Effects

$$Y_{it}(0) = \mu + \alpha_i + \beta_t + \varepsilon_{it}$$

3. Two Way Fixed Effects with Additional Factor

$$Y_{it}(0) = \mu + \alpha_i + \beta_t + \delta_i \gamma_t + \varepsilon_{it}$$

4. More general

PROPOSAL: GENERAL DISTANCE WEIGHTS

$$\omega_{it} = \exp\left(-\lambda_i^{\mathrm{un}} \underbrace{\mathrm{dist}^{\mathrm{un}}(i, N)}_{\text{unit distance}} - \lambda^{\mathrm{ti}} \underbrace{\mathrm{dist}^{\mathrm{ti}}(t, T)}_{\text{time distance}}\right)$$

- $\mathrm{dist}^{\mathrm{un}}(i, j)$ and $\mathrm{dist}^{\mathrm{ti}}(s, t)$ possibly vector-valued.
- Tuning par λ^{un} and λ^{ti}, chosen jointly through cross-validation
- Unit weights: $\mathrm{dist}^{\mathrm{un}}(i, j)$ depend only on outcomes/treatments for units i and j.
- Time weights: $\mathrm{dist}^{\mathrm{ti}}(s, t)$ depends on s and t, possibly on outcomes, *e.g.*, difference $|t - s|$, or incorporating seasonals, *e.g.*, days of the week, $t - s - 7 \times \lfloor (t - s)/7 \rfloor$.

PROPOSAL: DISTANCE WEIGHTS, A SIMPLE EXAMPLE

Weight contributions to objective function by unit and time:

$$\omega_{it} = \exp\left(-\lambda_i^{\text{un}} \underbrace{\text{dist}^{\text{un}}(i, N)}_{\text{unit distance}} - \lambda^{\text{ti}} \underbrace{\text{dist}^{\text{ti}}(t, T)}_{\text{time distance}}\right)$$

Simple example (but more general weights functions possible):

1. Unit weights: distance for control outcomes

$$\text{dist}^{\text{un}}(i, j) = \left(\frac{\sum_{t=1}^{T}(1 - W_{it})(1 - W_{jt})(Y_{it} - Y_{jt})^2}{\sum_{t=1}^{T}(1 - W_{it})(1 - W_{jt})}\right)^{1/2} \approx \|\mathbf{Y}_{i\cdot} - \mathbf{Y}_{j\cdot}\|_2$$

2. Time weights: difference in time

$$\text{dist}^{\text{ti}}(s, t) = |t - s|$$

WHY LOCAL WEIGHTS? SIMPLE EXAMPLE

- Unit weights can improve prediction over time-varying unobservables
- Time weights can improve prediction over unit-varying unobservables

- Simple example: take $\gamma_t = \mu t$, $\delta_i \in \{0, 1\}$, $\delta_N = 1$

$$Y_{i,t}(0) = \delta_i \gamma_t + \varepsilon_{i,t}, \varepsilon_{i,t} \sim_{i.i.d.} \mathcal{N}(0, \sigma^2)$$

$\Rightarrow$ $d^{unit}(i, j)$: weights more units predictive of confounder γ_T

$\Rightarrow$ d^{time}: weight more units with similar γ_t/predictive of δ_N

TUNING PARAMETERS (WITH TWFE MODEL)

Choose $\lambda^{\mathrm{ti}}, \lambda^{\mathrm{un}}$ jointly by leave-one-out cross-validation

$$\min_{\lambda^{\mathrm{ti}},\lambda^{\mathrm{un}}} Q(\lambda^{\mathrm{ti}}, \lambda^{\mathrm{un}}) = \min_{\lambda^{\mathrm{ti}},\lambda^{\mathrm{un}}} \sum_{i=1}^{N} \sum_{t=1}^{T} (1 - W_{it}) \left(Y_{it} - \hat{Y}_{it}(\lambda^{\mathrm{ti}}, \lambda^{\mathrm{un}}) \right)^2$$

$$\hat{Y}_{it}(\lambda^{\mathrm{ti}}, \lambda^{\mathrm{un}}) = \underbrace{\hat{\alpha}_i(\lambda^{\mathrm{ti}}, \lambda^{\mathrm{un}}, (i,t)) + \hat{\beta}_t(\lambda^{\mathrm{ti}}, \lambda^{\mathrm{un}}, (i,t))}_{\text{TWFE Model}}$$

$$\hat{\alpha}_i(\lambda^{\mathrm{ti}}, \lambda^{\mathrm{un}}, (i,t)) + \hat{\beta}_t(\lambda^{\mathrm{ti}}, \lambda^{\mathrm{un}}, (i,t)) =$$

$$\arg\min_{\alpha,\beta} \sum_{j=1}^{N} \sum_{s=1}^{T} \underbrace{\mathbf{1}_{(j,s)\neq(i,t)} \omega_{js}^{(i,t)}(\lambda^{\mathrm{ti}}, \lambda^{\mathrm{un}})}_{\text{Cross Validation}} (1 - W_{js}) \left(Y_{js} - \alpha_j - \beta_s \right)^2$$

TUNING PARAMETERS (WITH TWFE MODEL)

Choose $\lambda^{\text{ti}}, \lambda^{\text{un}}$ jointly by leave-one-out cross-validation

$$\min_{\lambda^{\text{ti}},\lambda^{\text{un}}} Q(\lambda^{\text{ti}}, \lambda^{\text{un}}) = \min_{\lambda^{\text{ti}},\lambda^{\text{un}}} \sum_{i=1}^{N} \sum_{t=1}^{T} (1 - W_{it}) \left(Y_{it} - \hat{Y}_{it}(\lambda^{\text{ti}}, \lambda^{\text{un}}) \right)^2$$

$$\hat{Y}_{it}(\lambda^{\text{ti}}, \lambda^{\text{un}}) = \underbrace{\hat{\alpha}_i(\lambda^{\text{ti}}, \lambda^{\text{un}}, (i,t)) + \hat{\beta}_t(\lambda^{\text{ti}}, \lambda^{\text{un}}, (i,t))}_{\text{TWFE Model}}$$

$$\hat{\alpha}_i(\lambda^{\text{ti}}, \lambda^{\text{un}}, (i,t)) + \hat{\beta}_t(\lambda^{\text{ti}}, \lambda^{\text{un}}, (i,t)) =$$

$$\arg\min_{\alpha,\beta} \sum_{j=1}^{N} \sum_{s=1}^{T} \underbrace{\mathbf{1}_{(j,s)\neq(i,t)} \omega_{js}^{(i,t)}(\lambda^{\text{ti}}, \lambda^{\text{un}})}_{\text{Cross Validation}} (1 - W_{js}) \left(Y_{js} - \alpha_j - \beta_s \right)^2$$

WEIGHTS IN CROSSVALIDATION

We do not use Y_{it} in construction of unit distance measure:

$$\omega_{js}^{(i,t)}(\lambda^{\mathrm{ti}}, \lambda^{\mathrm{un}}) = \exp\left(-\lambda^{\mathrm{un}}\widetilde{\mathrm{dist}^{\mathrm{un}}}(i,j)^{t} - \lambda^{\mathrm{ti}}|t-s|\right)$$

where

$$\widetilde{\mathrm{dist}^{\mathrm{un}}}(i,j)^{t} = \left(\frac{\sum_{s=1}^{T} \mathbf{1}_{s\neq t}(1-W_{is})(1-W_{js})(Y_{is}-Y_{js})^2}{\sum_{t=1}^{T}(1-W_{it})(1-W_{jt})}\right)^{1/2}$$

WEIGHTS IN CROSSVALIDATION

We do not use Y_{it} in construction of unit distance measure:

$$\omega_{js}^{(i,t)}(\lambda^{\mathrm{ti}}, \lambda^{\mathrm{un}}) = \exp\left(-\lambda^{\mathrm{un}}\widetilde{\mathrm{dist}}^{\mathrm{un}}(i,j)^t - \lambda^{\mathrm{ti}}|t-s|\right)$$

where

$$\widetilde{\mathrm{dist}}^{\mathrm{un}}(i,j)^t = \left(\frac{\sum_{s=1}^{T} \mathbf{1}_{s\neq t}(1-W_{is})(1-W_{js})(Y_{is}-Y_{js})^2}{\sum_{t=1}^{T}(1-W_{it})(1-W_{jt})}\right)^{1/2}$$

SIMPLE NUMERICAL ILLUSTRATION

- The choice of λ^{ti} may depend on serial correlation. We provide a simple numerical example with two DGPs below.

- **Independence over time**:

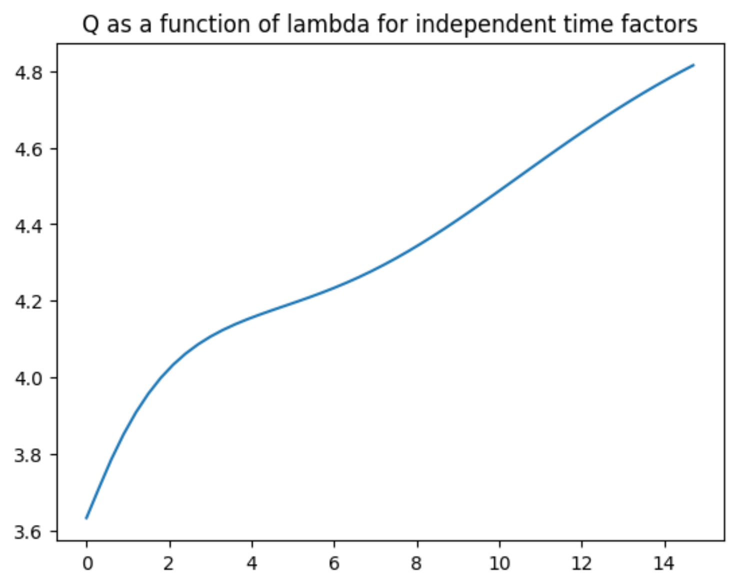
$$Y_{i,t} = \alpha_i + \beta_t + \varepsilon_{i,t}, \quad \alpha_i, \beta_t \sim_{i.i.d.} \mathcal{N}(0,1), \varepsilon_{i,t} \sim \mathcal{N}(0, 0.1)$$

- **Serial correlation**:

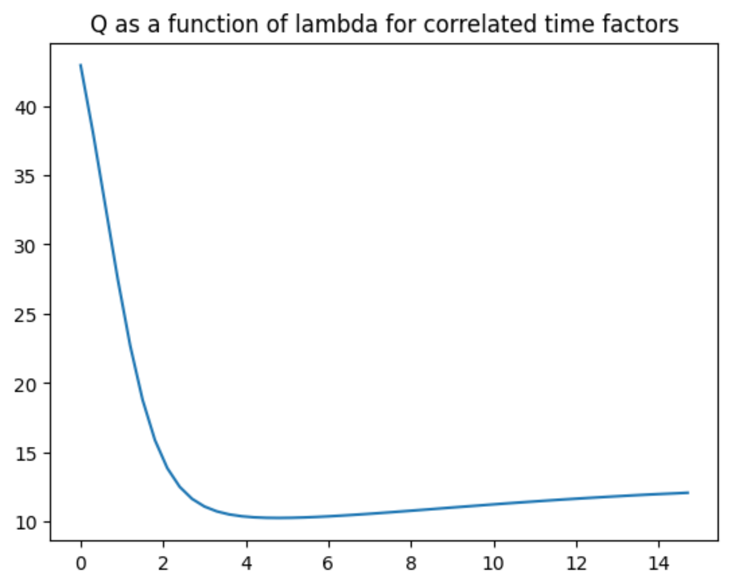
$$Y_{i,t} = \alpha_{i,t} + \beta_t + \varepsilon_{i,t}, \quad \alpha_{i,t}|\alpha_{i,t-1} \sim \mathcal{N}(\rho\alpha_{i,t-1}, \sigma^2)$$

where $\alpha_{i,1} \sim \mathcal{N}(0, \sigma^2/(1-\rho))$, $\rho = 0.9$, $\sigma^2 = 0.2$.

OBJECTIVE FOR λ^{TI} WITH INDEPENDENCE



OBJECTIVE FOR λ^{TI} WITH SERIAL CORRELATION



RMSE FOR SC, SDID, MC, TWFE AND WEIGHTED-TWFE

	State Earnings	Country GDP	State Smoking
	N = 50, T = 40	N = 111, T = 48	N = 38, T = 31
SC	0.116	0.129	0.232
SDID	0.111	0.029	0.135
MC	0.110 (3)	0.042 (29)	0.264 (3)
TWFE	0.134	0.358	0.446
DWCP-Intercept	0.146 (12,0.7)	0.065 (24,0.4)	0.207 (18,2.1)
DWCP-TWFE	0.107 (2,0.21)	0.031 (0.8,1.6)	0.137 (0.8,1.4)
DWCP-1 Factor	0.104 (0,0.08)	0.031 (0.15,0.3)	0.161 (0.1,0.1)

NORMALIZED WEIGHTS FOR TWFE, STATE EARNINGS, T=40, N=50.

$\lambda^{\mathrm{un}} = 2, \lambda^{\mathrm{ti}} = 0.21$

	T(=40)	T-1	T-2	...	20	...	10	...	1
CA	–	13.0	10.6	...	2.4	...	0.3	...	0.0
NY	12.0	9.8	7.9	...	1.8	...	0.2	...	0.0
HI	11.6	9.4	7.6	...	1.8	...	0.2	...	0.0
⋮	⋮	⋮	⋮	⋱	⋮	⋱	⋮	⋱	⋮
OH	7.6	6.2	5.0	...	1.2	...	0.1	...	0.0
⋮	⋮	⋮	⋮	⋱	⋮	⋱	⋮	⋱	⋮
ID	3.9	3.2	2.6	...	0.6	...	0.1	...	0.0
MT	3.6	3.0	2.4	...	0.6	...	0.1	...	0.0

NORMALIZED WEIGHTS FOR INTERCEPT MODEL, STATE SMOKING DATA, T=31, N=38.

$\lambda^{\mathrm{un}} = 18, \lambda^{\mathrm{ti}} = 2.1$

	T(=31)	T-1	T-2	...	26	...	20	...	1
CO	–	712.7	87.3	...	0.2	...	0.0	...	0.0
IL	173.2	21.2	2.6	...	0.0	...	0.0	...	0.0
MT	44.0	5.4	0.7	...	0.0	...	0.0	...	0.0
⋮	⋮	⋮	⋮	⋱	⋮	⋱	⋮	⋱	⋮
ME	0.6	0.1	0.0	...	0.0	...	0.0	...	0.0
⋮	⋮	⋮	⋮	⋱	⋮	⋱	⋮	⋱	⋮
KT	0.0	0.0	0.0	...	0.0	...	0.0	...	0.0
NH	0.0	0.0	0.0	...	0.0	...	0.0	...	0.0

NORMALIZED WEIGHTS FOR TWFE, STATE SMOKING DATA, T=31, N=38.

$\lambda^{\mathrm{un}} = 0.8, \lambda^{\mathrm{ti}} = 1.4$

	T(=31)	T-1	T-2	…	26	…	20	…	1
CO	–	9.8	2.4	…	0.0	…	0.0	…	0.0
IL	33.9	8.4	2.1	…	0.0	…	0.0	…	0.0
MT	31.9	7.9	1.9	…	0.0	…	0.0	…	0.0
⋮	⋮	⋮	⋮	⋱	⋮	⋱	⋮	⋱	⋮
ME	26.4	6.5	1.6	…	0.0	…	0.0	…	0.0
⋮	⋮	⋮	⋮	⋱	⋮	⋱	⋮	⋱	⋮
KT	5.5	1.4	0.3	…	0.0	…	0.0	…	0.0
NH	2.7	0.7	0.2	…	0.0	…	0.0	…	0.0

GENERAL ASSIGNMENT PATTERNS

1. Given $(\lambda^{\text{un}}, \lambda^{\text{ti}})$, impute missing Y_{it} by first minimizing (say, for the TWFE outcome model) seperately for each missing pair (i, t)

$$\min_{\alpha,\beta} \sum_{j=1}^{N} \sum_{s=1}^{T} (1 - W_{js}) \exp\left(-\lambda^{\text{un}} \text{dist}^{\text{un}}(i, j) - \lambda^{\text{ti}} \text{dist}^{\text{ti}}(t, s)\right) \left(Y_{js} - \alpha_j - \beta_s\right)^2$$

2. Impute Y_{it} as

$$\hat{Y}_{it} = \hat{\alpha}_i + \hat{\beta}_t$$

3. Repeat for all missing values and average.

FUTURE WORK

- The choice of the time weights must depend on the behavior of unobserved time-varying factors. Other distance measures are possible (seasonality, distance analogous to d^{un}, etc.)

- Different weights' choice may correspond to robustness to different models of unobserved confounders. Future work may formalize robustness properties based on the weights' choice

- Choice of number of factors can use out-of-sample validation or other model selection criteria

REFERENCES

- Abadie, Alberto. "Using synthetic controls: Feasibility, data requirements, and methodological aspects." Journal of Economic Literature 59, no. 2 (2021): 391-425.
- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program." Journal of the American statistical Association 105, no. 490 (2010): 493-505.
- Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. "Synthetic difference-in-differences." American Economic Review 111, no. 12 (2021): 4088-4118.

REFERENCES (CTD)

- Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. "Matrix completion methods for causal panel data models." Journal of the American Statistical Association 116, no. 536 (2021): 1716-1730.
- Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. "The augmented synthetic control method." Journal of the American Statistical Association 116, no. 536 (2021): 1789-1803.

  Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. "Synthetic controls with staggered adoption." Journal of the Royal Statistical Society Series B: Statistical Methodology 84, no. 2 (2022): 351-381.

REFERENCES (CTD)

- Brodersen, Kay H., Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott. "Inferring causal impact using Bayesian structural time-series models." (2015): 247-274.
- Doudchenko, Nikolay, and Guido W. Imbens. Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. No. w22791. National Bureau of Economic Research, 2016.
- Imbens, Guido, and Davide Viviano. Identification and Inference for Synthetic Controls with Confounding, 2023.