Dynamic Covariate Balancing: Estimating Treatment Effects over Time with Potential Local Projections

Davide Viviano   Jelena Bradic
Harvard   UC San Diego

July 26, 2023

Two Examples

Ex1 "Effect of negative advertisement on election outcome?" (Blackwell, 2013)

- Dynamic choice of which advertisement to send;
- Politicians may make strategic choices based on past polls.

Ex2 "What is the effect of democracy on economic growth?" (Acemoglu et al., 2019)

- Hard question.
- Long and short term impacts may be different;
- Past growth may determine the selection into treatment;
- It is necessary to control for many potential confounders.

Inference on Treatment Histories

Suppose we collect data from an observational study with T periods and n individuals.

Goal: Inference on the effect of a treatment history.

Inference on Treatment Histories

Suppose we collect data from an observational study with T periods and n individuals.

Goal: Inference on the effect of a treatment history.

Challenges:

- Individuals select dynamically on arbitrary past information;
- Selection mechanism (propensity score) hard to estimate/unknown;
- Intermediate covariates and outcomes depend on past treatment assignments;
- Many potential confounders (high-dimensional covariates).

Content

Data and Notation

Ex-post evaluation:

- Collect data from T periods, then conduct inference.

Data and Notation

Ex-post evaluation:

- Collect data from T periods, then conduct inference.

Data:

$$\Big(X_{i,1}, X_{i,2}, \cdots, Y_{i,1}, \cdots, Y_{i,T}, D_{i,1}, \cdots, D_{i,T}\Big)_{i=1}^{n} \sim_{i.i.d.} \mathcal{P}$$

- $X_{i,t} \in \mathcal{X}_t$ - time-varying covariates;
- $Y_{i,t} \in \mathcal{Y}$ - intermediate outcomes;
- $D_{i,t} \in \{0,1\}$ - dynamic treatments;

$\Rightarrow$ Potential outcomes: $Y_{i,t}(d_1, \cdots, d_t)$.

Data and Notation

Ex-post evaluation:

- Collect data from T periods, then conduct inference.

Data:

$$\Big(X_{i,1}, X_{i,2}, \cdots, Y_{i,1}, \cdots, Y_{i,T}, D_{i,1}, \cdots, D_{i,T}\Big)_{i=1}^{n} \sim_{i.i.d.} \mathcal{P}$$

- $X_{i,t} \in \mathcal{X}_t$ - time-varying covariates;
- $Y_{i,t} \in \mathcal{Y}$ - intermediate outcomes;
- $D_{i,t} \in \{0, 1\}$ - dynamic treatments;

$\Rightarrow$ Potential outcomes: $Y_{i,t}(d_1, \cdots, d_t)$.

Goal

- Inference on treatment effects at time T.

Illustration: Two Periods

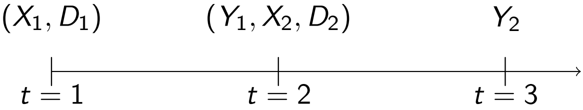
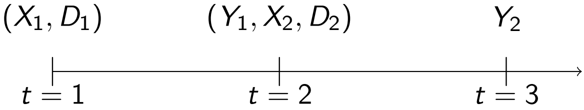


Illustration: Two Periods



Dynamic model:

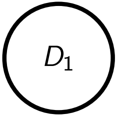
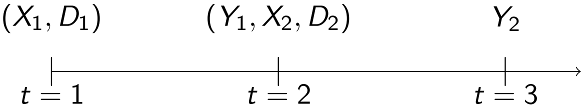


Illustration: Two Periods



Dynamic model:

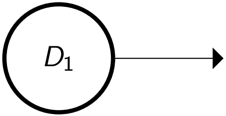
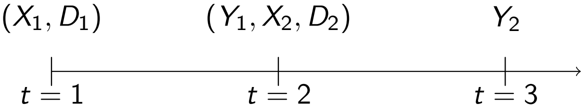


Illustration: Two Periods



Dynamic model:

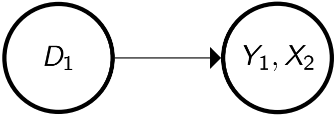
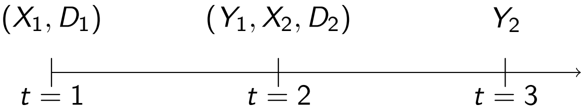


Illustration: Two Periods



Dynamic model:

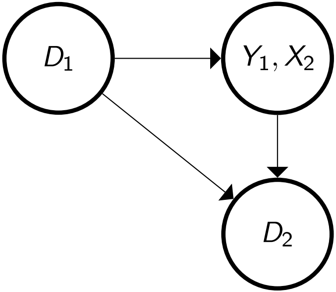
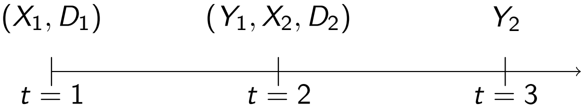


Illustration: Two Periods



Dynamic model:

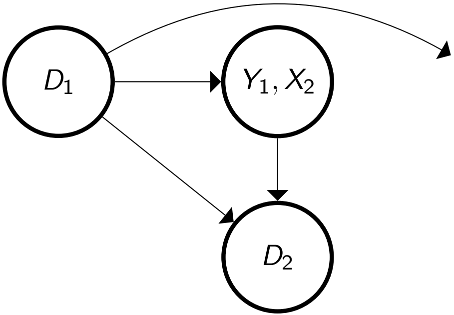
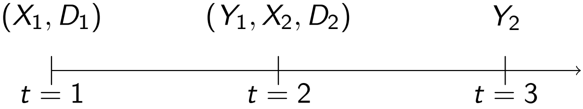


Illustration: Two Periods



Dynamic model:

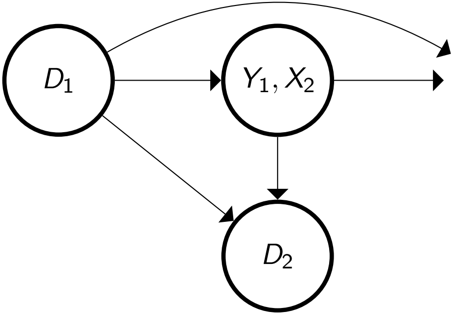
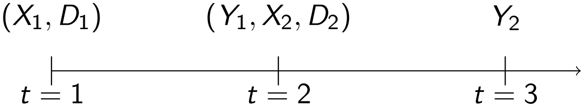


Illustration: Two Periods



Dynamic model:

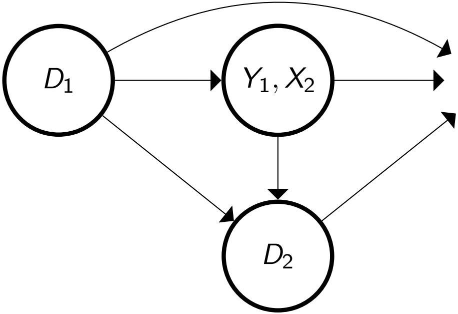
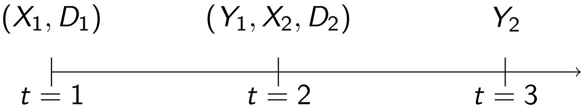


Illustration: Two Periods



Dynamic model:

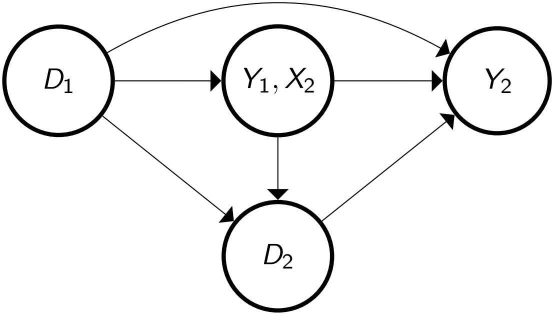
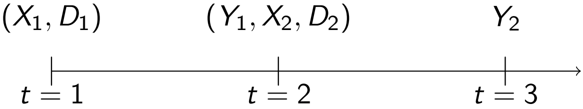
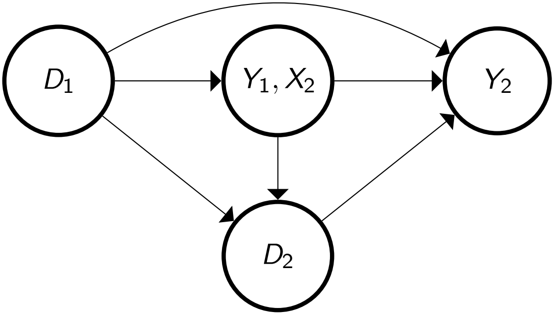


Illustration: Two Periods

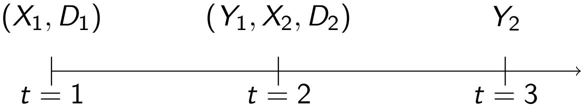


Dynamic model:

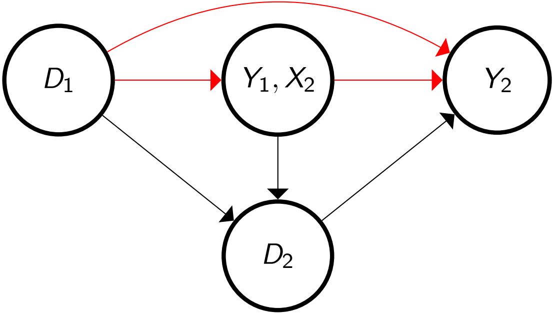


Goal Inference on $\mathbb{E}\Big[Y_2(1, 0) - Y_2(0, 0)\Big]$ (or conditional on baselines).

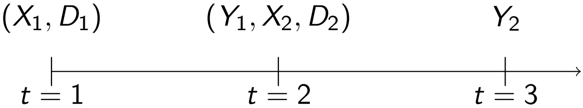
Illustration: Two Periods



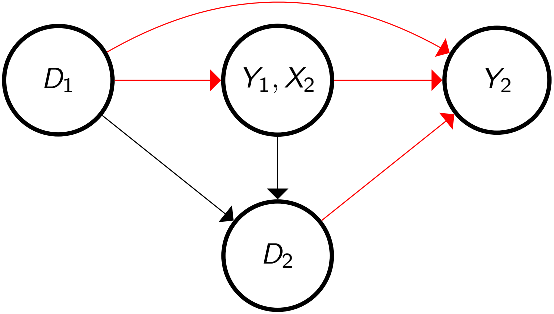
Dynamic model:



Goal Inference on $\mathbb{E}\Big[Y_2(1, 0) - Y_2(0, 0)\Big]$ (or conditional on baselines).



Dynamic model:

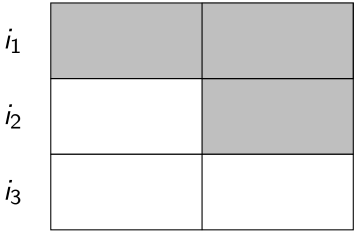


Goal Inference on $\mathbb{E}\Big[Y_2(1,1) - Y_2(0,0)\Big]$ (or conditional on baselines).

Data

Goal Inference on $\mathbb{E}\Big[Y_2(1,1) - Y_2(0,0)\Big]$.

Data:



i_1 $\Big(Y_1(0), X_2(0), Y_2(0,0)\Big)$?;

i_2 $\Big(Y_1(1), X_2(1), Y_2(0,0), Y_2(1,1)\Big)$?;

i_3 ...

Data

Goal Inference on $\mathbb{E}\Big[Y_2(1,1) - Y_2(0,0)\Big]$.

Data:

i_1	$(Y_1, X_2)(1)$	$Y_2(1,1)$
i_2		
i_3		

i_1 $\Big(Y_1(0), X_2(0), Y_2(0,0)\Big)$?;

i_2 $\Big(Y_1(1), X_2(1), Y_2(0,0), Y_2(1,1)\Big)$?;

i_3 ...

Data

Goal Inference on $\mathbb{E}\Big[Y_2(1,1) - Y_2(0,0)\Big]$.

Data:

i_1	$(Y_1, X_2)(1)$	$Y_2(1,1)$
i_2	$(Y_1, X_2)(0)$	$Y_2(1,0)$
i_3	$(Y_1, X_2)(0)$	$Y_2(0,0)$

i_1 $\Big(Y_1(0), X_2(0), Y_2(0,0)\Big)$?;

i_2 $\Big(Y_1(1), X_2(1), Y_2(0,0), Y_2(1,1)\Big)$?;

i_3 ...

Data

Goal Inference on $\mathbb{E}\Big[Y_2(1,1) - Y_2(0,0)\Big]$.

Data:

i_1	$(Y_1, X_2)(1)$	$Y_2(1,1)$
i_2	$(Y_1, X_2)(0)$	$Y_2(1,0)$
i_3	$(Y_1, X_2)(0)$	$Y_2(0,0)$

- IPW?
- Estimate each counterfactual?

i_1 $\Big(Y_1(0), X_2(0), Y_2(0,0)\Big)$?;

i_2 $\Big(Y_1(1), X_2(1), Y_2(0,0), Y_2(1,1)\Big)$?;

i_3 ...

Data

Goal Inference on $\mathbb{E}\Big[Y_2(1,1) - Y_2(0,0)\Big]$.

Data:

i_1	$(Y_1, X_2)(1)$	$Y_2(1,1)$
i_2	$(Y_1, X_2)(0)$	$Y_2(1,0)$
i_3	$(Y_1, X_2)(0)$	$Y_2(0,0)$

- IPW?
- Estimate each counterfactual?

i_1 $\Big(Y_1(0), X_2(0), Y_2(0,0)\Big)$?;

i_2 $\Big(Y_1(1), X_2(1), Y_2(0,0), Y_2(1,1)\Big)$?;

i_3 ...

Data

Goal Inference on $\mathbb{E}\Big[Y_2(1,1) - Y_2(0,0)\Big]$.

Data:

i_1	$(Y_1, X_2)(1)$	$Y_2(1,1)$
i_2	$(Y_1, X_2)(0)$	$Y_2(1,0)$
i_3	$(Y_1, X_2)(0)$	$Y_2(0,0)$

- IPW?
- Estimate each counterfactual?

i_1 $\Big(Y_1(0), X_2(0), Y_2(0,0)\Big)$?;

i_2 $\Big(Y_1(1), X_2(1), Y_2(0,0), Y_2(1,1)\Big)$?;

i_3 ...

Review: some intuitive estimators

1. Regress end-line outcomes on covariates:

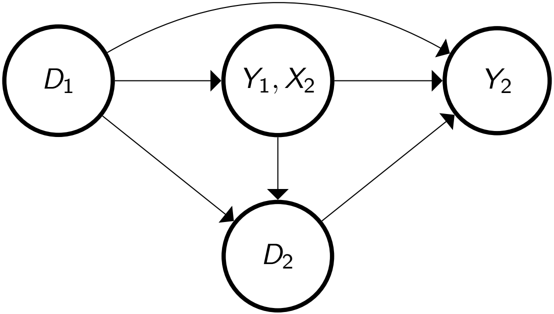
$$Y_{i,T} = \sum_{t=0}^{T-1} \alpha_t D_{i,t} + X_{i,t}\beta_t + \sum_{t=1}^{T-1} Y_{i,t}\phi_t + \varepsilon_{i,T}.$$

$D_{i,t-1}$ also affects $Y_{i,t}$ $\Rightarrow$ we main underestimate the overall effect.

1. Regress end-line outcomes on covariates:

$$Y_{i,T} = \sum_{t=0}^{T-1} \alpha_t D_{i,t} + X_{i,t}\beta_t + \sum_{t=1}^{T-1} Y_{i,t}\phi_t + \varepsilon_{i,T}.$$

$D_{i,t-1}$ also affects $Y_{i,t}$ $\Rightarrow$ we main underestimate the overall effect.

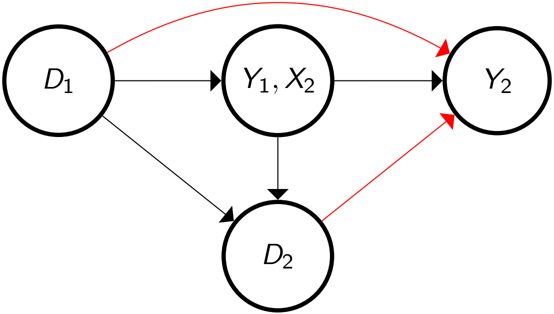


Review: some intuitive estimators

1. Regress end-line outcomes on covariates:

$$Y_{i,T} = \sum_{t=0}^{T-1} \alpha_t D_{i,t} + X_{i,t}\beta_t + \sum_{t=1}^{T-1} Y_{i,t}\phi_t + \varepsilon_{i,T}.$$

$D_{i,t-1}$ also affects $Y_{i,t}$ $\Rightarrow$ we main underestimate the overall effect.



Review: some intuitive estimators

2. Regress end-line outcomes and exclude the outcomes:

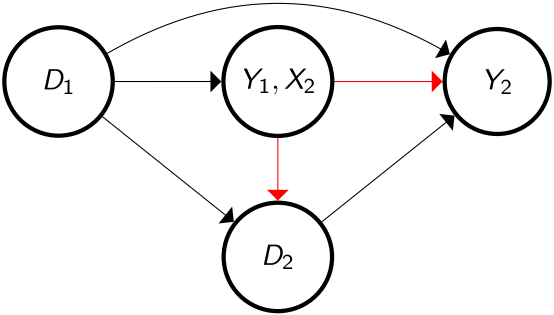
$$Y_{i,T} = \sum_{t=0}^{T-1} \alpha_t D_{i,t} + X_{i,1}\beta + \varepsilon_{i,T}.$$

Review: some intuitive estimators

2. Regress end-line outcomes and exclude the outcomes:

$$Y_{i,T} = \sum_{t=0}^{T-1} \alpha_t D_{i,t} + X_{i,1}\beta + \varepsilon_{i,T}.$$

$\Rightarrow$ omitted variable bias.



Review: some intuitive estimators

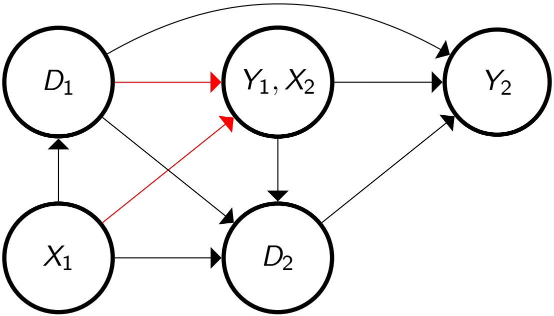
3. Run separate regressions for

$$Y_{i,t} = X_{i,t-1}\beta_t + D_{i,t-1}\alpha_t + \varepsilon_{i,t}, \quad X_{i,t} = X_{i,t-1}\tilde{\beta}_{t-1} + D_{i,t-1}\tilde{\alpha}_{t-1} + \varepsilon_{x,i,t}.$$

3. Run separate regressions for

$$Y_{i,t} = X_{i,t-1}\beta_t + D_{i,t-1}\alpha_t + \varepsilon_{i,t}, \quad X_{i,t} = X_{i,t-1}\tilde{\beta}_{t-1} + D_{i,t-1}\tilde{\alpha}_{t-1} + \varepsilon_{x,i,t}.$$

Prone to large estimation error in high dimensions.



Review: some intuitive estimators

4. Local projections (Jorda, 2005)

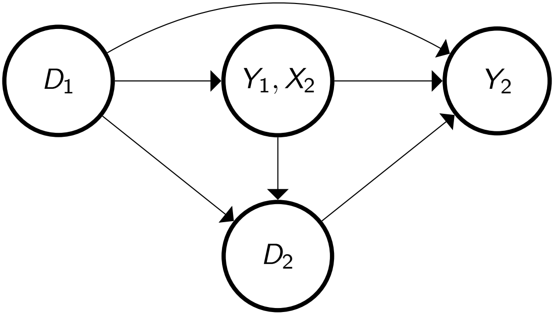
$$Y_{i,T} = \sum_{t=0}^{T-h} \alpha_{t,h} D_{i,t} + X_{i,t}\beta_{t,h} + Y_{i,t}\phi_{t,h} + \varepsilon_{i,T}$$

Review: some intuitive estimators

4. Local projections (Jorda, 2005)

$$Y_{i,T} = \sum_{t=0}^{T-h} \alpha_{t,h} D_{i,t} + X_{i,t}\beta_{t,h} + Y_{i,t}\phi_{t,h} + \varepsilon_{i,T}$$

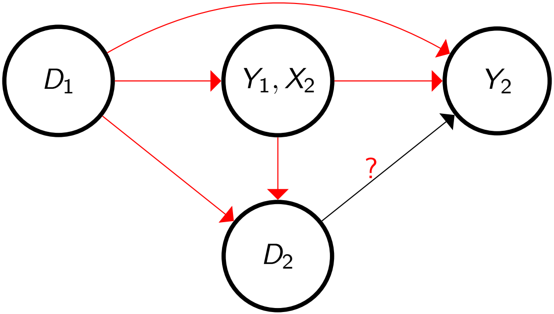
$\alpha_{t,h}$ targets a different estimand!



4. Local projections (Jorda, 2005)

$$Y_{i,T} = \sum_{t=0}^{T-h} \alpha_{t,h} D_{i,t} + X_{i,t}\beta_{t,h} + Y_{i,t}\phi_{t,h} + \varepsilon_{i,T}$$

$\alpha_{t,h}$ targets a different estimand!



Dynamic Covariate Balancing

Method

1. Model on potential outcomes with "potential" local projections:
   - ⇒ It does not require to estimate each counterfactual ("adapt" local projections of Jorda (2005) to dynamic treatments);

Dynamic Covariate Balancing

Method

1. Model on potential outcomes with "potential" local projections:
   - ⇒ It does not require to estimate each counterfactual ("adapt" local projections of Jorda (2005) to dynamic treatments);
2. Estimation balances covariates directly (dynamically)
   - ⇒ It does not require to specify and estimate the propensity score;
   - ⇒ It guarantees an asymptotic vanishing bias in high dimensions;
   - ⇒ More stable than (A)IPW estimators in the presence of poor overlap;
   - ⇒ Generalizes Zubizarreta (2015) and Athey et al. (2018) to dynamics.

Dynamic Covariate Balancing

Method

1. Model on potential outcomes with "potential" local projections:
   - ⇒ It does not require to estimate each counterfactual ("adapt" local projections of Jorda (2005) to dynamic treatments);
2. Estimation balances covariates directly (dynamically)
   - ⇒ It does not require to specify and estimate the propensity score;
   - ⇒ It guarantees an asymptotic vanishing bias in high dimensions;
   - ⇒ More stable than (A)IPW estimators in the presence of poor overlap;
   - ⇒ Generalizes Zubizarreta (2015) and Athey et al. (2018) to dynamics.

Conditions:

1. Sequential ignorability:
   - ⇒ Treatments are assigned sequentially based on past observations and exogenous unobservables.
2. Approximate linearity:
   - ⇒ Linearity of outcomes on past outcomes and high-dimensional covariates.

Related Literature

- Dynamic Treatments' literature and marginal structural models [E.g., Robins, 1986, Robins et al., 2000, Bang and Robins 2005, Boruvka et al., 2018; Blackwell, 2013; Bojinov and Shephard 2019; Lewis and Syrgkanis, 2020; Bodory et al. 2020]
- Balancing and Dynamic Treatments: [Imai and Ratkovic (2005); Yiu and Su (2018); Kallus and Santacatterina (2018), Zhou and Wodtke (2018)]

Related Literature

- Dynamic Treatments' literature and marginal structural models [E.g., Robins, 1986, Robins et al., 2000, Bang and Robins 2005, Boruvka et al., 2018; Blackwell, 2013; Bojinov and Shephard 2019; Lewis and Syrgkanis, 2020; Bodory et al. 2020]
- Balancing and Dynamic Treatments: [Imai and Ratkovic (2005); Yiu and Su (2018); Kallus and Santacatterina (2018), Zhou and Wodtke (2018)]

$\Rightarrow$ None studies (approximate) residual balancing in dynamic settings.

$\Rightarrow$ We characterize the (high-dimensional) bias and balancing equations.

Related Literature

- Dynamic Treatments' literature and marginal structural models [E.g., Robins, 1986, Robins et al., 2000, Bang and Robins 2005, Boruvka et al., 2018; Blackwell, 2013; Bojinov and Shephard 2019; Lewis and Syrgkanis, 2020; Bodory et al. 2020]
- Balancing and Dynamic Treatments: [Imai and Ratkovic (2005); Yiu and Su (2018); Kallus and Santacatterina (2018), Zhou and Wodtke (2018)]

$\Rightarrow$ None studies (approximate) residual balancing in dynamic settings.

  $\Rightarrow$ We characterize the (high-dimensional) bias and balancing equations.

- Estimation and inference in time-series/(macro)econometrics [Jorda, 2005; Stock and Watson, 2018, Angrist et al. (2018); Rambachan and Shephard (2019)]
  - $\Rightarrow$ Exogeneity and independence of shocks (treatments).
- Balancing in *i.i.d.* settings [E.g., Zubizarreta (2015); Athey et al. (2018)]; difference-in-differences, synthetic controls, panel data [E.g., Ben-Michael et al.; Athey and Imbens (2022)); Arkhangelsky and Imbens (2019); ...]

$\Rightarrow$ Here estimation and inference under sequential exogeneity.

Content

Basic Model: Example

Basic model that we generalize in the following slides:

$$Y_{i,2} = \Big[Y_{i,1}, X_{i,1}, D_{i,1}, D_{i,2}\Big]\beta_2 + \varepsilon_{i,2}, \quad \varepsilon_{i,2} \perp D_{i,2} \Big| D_{i,1}, X_{i,1}, Y_{i,1}$$
$$Y_{i,1} = \Big[X_{i,1}, D_{i,1}\Big]\beta_1 + \varepsilon_{i,1}, \qquad (\varepsilon_{i,1}, \varepsilon_{i,2}) \perp D_{i,1} \Big| X_{i,1}.$$

Basic Model: Example

Basic model that we generalize in the following slides:

$$
\begin{aligned}
Y_{i,2} &= \Big[Y_{i,1}, X_{i,1}, D_{i,1}, D_{i,2}\Big]\beta_2 + \varepsilon_{i,2}, \quad && \varepsilon_{i,2} \perp D_{i,2} \Big| D_{i,1}, X_{i,1}, Y_{i,1} \\
Y_{i,1} &= \Big[X_{i,1}, D_{i,1}\Big]\beta_1 + \varepsilon_{i,1}, && (\varepsilon_{i,1}, \varepsilon_{i,2}) \perp D_{i,1} \Big| X_{i,1}.
\end{aligned}
$$

We can write

$$
\begin{aligned}
Y_{i,2}(d_1, d_2) &= \Big[Y_{i,1}(d_1), X_{i,1}, d_1, d_2\Big]\beta_2 + \varepsilon_{i,2} \\
Y_{i,1}(d_1) &= \Big[X_{i,1}, d_1\Big]\beta_1 + \varepsilon_{i,1}
\end{aligned}
$$

Basic Model: Example

Basic model that we generalize in the following slides:

$$Y_{i,2} = \Big[Y_{i,1}, X_{i,1}, D_{i,1}, D_{i,2}\Big]\beta_2 + \varepsilon_{i,2}, \quad \varepsilon_{i,2} \perp D_{i,2}\Big| D_{i,1}, X_{i,1}, Y_{i,1}$$

$$Y_{i,1} = \Big[X_{i,1}, D_{i,1}\Big]\beta_1 + \varepsilon_{i,1}, \quad (\varepsilon_{i,1}, \varepsilon_{i,2}) \perp D_{i,1}\Big| X_{i,1}.$$

We can write

$$Y_{i,2}(d_1, d_2) = \Big[Y_{i,1}(d_1), X_{i,1}, d_1, d_2\Big]\beta_2 + \varepsilon_{i,2}$$

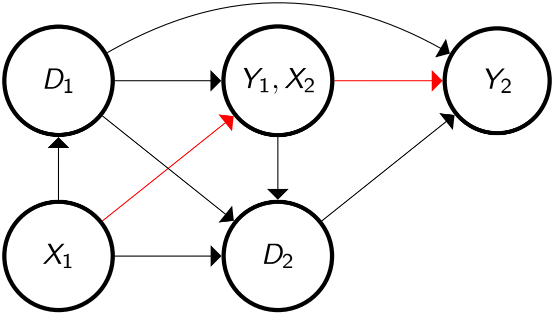
$$Y_{i,1}(d_1) = \Big[X_{i,1}, d_1\Big]\beta_1 + \varepsilon_{i,1}$$

$\Rightarrow$ $Y_{i,2}(d_1, d_2)$ is linear in $X_{i,1}$ *unconditionally* on $Y_{i,1}(d_1)$:

$$Y_{i,2}(d_1, d_2) = \Big[1, 1, X_{i,1}\Big]\beta_{d_1,d_2} + \nu_{i,1}^{d_1,d_2}, \quad \mathbb{E}\Big[\nu_{i,1}^{d_1,d_2}\Big| X_{i,1}, D_{i,1}\Big] = 0.$$

Example: Illustration

- Model on potential outcomes: linear dependencies?



- Remaining components are left unspecified.

Basic Model: Example

Recall the starting model:

$$Y_{i,2} = \Big[Y_{i,1}, X_{i,1}, D_{i,1}, D_{i,2}\Big]\beta_2 + \varepsilon_{i,2}, \quad \varepsilon_{i,2} \perp D_{i,2}\Big|D_{i,1}, X_{i,1}, Y_{i,1}$$

$$Y_{i,1} = \Big[X_{i,1}, D_{i,1}\Big]\beta_1 + \varepsilon_{i,1}, \quad (\varepsilon_{i,1}, \varepsilon_{i,2}) \perp D_{i,1}\Big|X_{i,1}.$$

$\Rightarrow$ Why not a linear model on *observed* outcomes?

Basic Model: Example

Recall the starting model:

$$\begin{aligned} Y_{i,2} &= \Big[Y_{i,1}, X_{i,1}, D_{i,1}, D_{i,2}\Big]\beta_2 + \varepsilon_{i,2}, \quad && \varepsilon_{i,2} \perp D_{i,2}\Big| D_{i,1}, X_{i,1}, Y_{i,1} \\ Y_{i,1} &= \Big[X_{i,1}, D_{i,1}\Big]\beta_1 + \varepsilon_{i,1}, && (\varepsilon_{i,1}, \varepsilon_{i,2}) \perp D_{i,1}\Big| X_{i,1}. \end{aligned}$$

$\Rightarrow$ Why not a linear model on *observed* outcomes?

$$\mathbb{E}\Big[Y_{i,2}\Big|X_{i,1}, D_{i,1}\Big] = \Big[X_{i,1}, D_{i,1}\Big]\gamma + \underbrace{\beta_2 \mathbb{E}\Big[D_{i,2}|X_{i,1}, D_{i,1}\Big]}_{\textit{problematic}}$$

Basic Model: Example

Recall the starting model:

$$Y_{i,2} = \Big[Y_{i,1}, X_{i,1}, D_{i,1}, D_{i,2}\Big]\beta_2 + \varepsilon_{i,2}, \quad \varepsilon_{i,2} \perp D_{i,2}\Big| D_{i,1}, X_{i,1}, Y_{i,1}$$

$$Y_{i,1} = \Big[X_{i,1}, D_{i,1}\Big]\beta_1 + \varepsilon_{i,1}, \quad (\varepsilon_{i,1}, \varepsilon_{i,2}) \perp D_{i,1}\Big| X_{i,1}.$$

$\Rightarrow$ Why not a linear model on *observed* outcomes?

$$\mathbb{E}\Big[Y_{i,2}\Big| X_{i,1}, D_{i,1}\Big] = \Big[X_{i,1}, D_{i,1}\Big]\gamma + \underbrace{\beta_2 \mathbb{E}\Big[D_{i,2}|X_{i,1}, D_{i,1}\Big]}_{problematic}$$

$\Rightarrow$ Model on the observed outcomes also depend on $(X_1, D_1) \to D_2$;

$\Rightarrow$ Model on potential outcomes is more flexible.

Model Specification: General Case

Define $H_{i,2}(d_1) = \Big[X_{i,1}, X_{i,2}(d_1), Y_{i,1}(d_1)\Big]$ (also with intercepts).

$$\mathbb{E}\Big[Y_{i,2}(d_1, d_2)\Big|H_{i,2}, D_{i,1} = d_1\Big] = H_{i,2}(d_1)\beta^2_{d_1,d_2},$$
$$\mathbb{E}\Big[Y_{i,2}(d_1, d_2)\Big|X_{i,1}\Big] = X_{i,1}\beta^1_{d_1,d_2}.$$

Model Specification: General Case

Define $H_{i,2}(d_1) = \Big[X_{i,1}, X_{i,2}(d_1), Y_{i,1}(d_1)\Big]$ (also with intercepts).

$$\mathbb{E}\Big[Y_{i,2}(d_1, d_2)\Big|H_{i,2}, D_{i,1} = d_1\Big] = H_{i,2}(d_1)\beta^2_{d_1,d_2},$$
$$\mathbb{E}\Big[Y_{i,2}(d_1, d_2)\Big|X_{i,1}\Big] = X_{i,1}\beta^1_{d_1,d_2}.$$

- It holds if $H_{i,2}(d_1)$ is linear in $X_{i,1}$;
- It does not impose structural assumptions on $(D_{i,1}, D_{i,2})$;
- High dimensional model can be interpreted as approximate linearity (up-to small estimation error)

Model Specification: General Case

Define $H_{i,2}(d_1) = \Big[X_{i,1}, X_{i,2}(d_1), Y_{i,1}(d_1)\Big]$ (also with intercepts).

$$\mathbb{E}\Big[Y_{i,2}(d_1, d_2)\Big|H_{i,2}, D_{i,1} = d_1\Big] = H_{i,2}(d_1)\beta^2_{d_1,d_2},$$
$$\mathbb{E}\Big[Y_{i,2}(d_1, d_2)\Big|X_{i,1}\Big] = X_{i,1}\beta^1_{d_1,d_2}.$$

- It holds if $H_{i,2}(d_1)$ is linear in $X_{i,1}$;
- It does not impose structural assumptions on $(D_{i,1}, D_{i,2})$;
- High dimensional model can be interpreted as approximate linearity (up-to small estimation error)

Sequential ignorability:

$$Y_{i,2}(d_1, d_2) \perp D_{i,2}\Big|H_{i,2}, D_{i,1}, \quad \Big(Y_{i,2}(d_1, d_2), H_{i,1}(d_1)\Big) \perp D_{i,1}\Big|X_{i,1}.$$

Identification and Estimation of the Linear model

! $\mathbb{E}\Big[Y_{i,2}\Big|X_{i,1}, D_{i,1} = 1, D_{i,2} = 1\Big] \neq \mathbb{E}\Big[Y_{i,2}(1,1)\Big|X_{i,1}\Big]$

Identification and Estimation of the Linear model

- ! $\mathbb{E}\Big[Y_{i,2}\Big|X_{i,1}, D_{i,1} = 1, D_{i,2} = 1\Big] \neq \mathbb{E}\Big[Y_{i,2}(1,1)\Big|X_{i,1}\Big]$

Identification

$$\mathbb{E}\Big[\underbrace{\mathbb{E}\Big[Y_{i,2}\Big|H_{i,2}, D_{i,1} = d_1, D_{i,2} = d_2\Big]}_{=\mathbb{E}[Y_{i,2}(d_1,d_2)|H_{i,2},D_{i,1}=d_1]}\Big|X_{i,1}, D_{i,1} = d_1\Big] = \mathbb{E}\Big[Y_{i,2}(d_1, d_2)\Big|X_{i,1}\Big].$$

Identification and Estimation of the Linear model

! $\mathbb{E}\Big[Y_{i,2}\Big|X_{i,1}, D_{i,1}=1, D_{i,2}=1\Big] \neq \mathbb{E}\Big[Y_{i,2}(1,1)\Big|X_{i,1}\Big]$

Identification

$$\mathbb{E}\Big[\underbrace{\mathbb{E}\Big[Y_{i,2}\Big|H_{i,2}, D_{i,1}=d_1, D_{i,2}=d_2\Big]}_{=\mathbb{E}[Y_{i,2}(d_1,d_2)|H_{i,2},D_{i,1}=d_1]}\Big|X_{i,1}, D_{i,1}=d_1\Big] = \mathbb{E}\Big[Y_{i,2}(d_1,d_2)\Big|X_{i,1}\Big].$$

Coefficients' Estimation (with Lasso):

$$Y_2 \rightarrow \Big[X_2, X_1, D_1=d_1, D_2=d_2\Big] \rightarrow H_2\hat{\beta}^2_{d_1,d_2}$$
$$H_2\hat{\beta}_{d_1,d_2} \rightarrow \Big[X_1, D_1=d_1\Big] \rightarrow X_1\hat{\beta}^1_{d_1,d_2}$$

Identification and Estimation of the Linear model

! $\mathbb{E}\Big[Y_{i,2}\Big|X_{i,1}, D_{i,1}=1, D_{i,2}=1\Big] \neq \mathbb{E}\Big[Y_{i,2}(1,1)\Big|X_{i,1}\Big]$

Identification

$$\mathbb{E}\Big[\underbrace{\mathbb{E}\Big[Y_{i,2}\Big|H_{i,2}, D_{i,1}=d_1, D_{i,2}=d_2\Big]}_{=\mathbb{E}[Y_{i,2}(d_1,d_2)|H_{i,2},D_{i,1}=d_1]}\Big|X_{i,1}, D_{i,1}=d_1\Big] = \mathbb{E}\Big[Y_{i,2}(d_1,d_2)\Big|X_{i,1}\Big].$$

Coefficients' Estimation (with Lasso):

$$Y_2 \rightarrow \Big[X_2, X_1, D_1=d_1, D_2=d_2\Big] \rightarrow H_2\hat{\beta}^2_{d_1,d_2}$$

$$H_2\hat{\beta}_{d_1,d_2} \rightarrow \Big[X_1, D_1=d_1\Big] \rightarrow X_1\hat{\beta}^1_{d_1,d_2}$$

- Problem: bias due to high-dimensionality;
- Unknown propensity score: (a)IPW can be prone to misspecification and poor overlap

Estimating $\mathbb{E}[Y_2(1,1)]$:

	$D_2 = 1$	$D_2 = 0$
$D_1 = 1$		
$D_1 = 0$		

Balancing conditions:

$$\hat{\gamma}_1 : \left|\left|\bar{X}_1 - \hat{\gamma}_1^\top X_1\right|\right|_\infty \leq \delta_n, \quad \hat{\gamma}_2 : \left|\left|\hat{\gamma}_2^\top H_2 - \hat{\gamma}_1^\top H_2\right|\right|_\infty \leq \delta_n$$

Balancing: Illustration

Estimating $\mathbb{E}[Y_2(1,1)]$:

	$D_2 = 1$	$D_2 = 0$
$D_1 = 1$		
$D_1 = 0$	$\hat{\gamma}_1 = 0$	$\hat{\gamma}_1 = 0$

Balancing conditions:

$$\hat{\gamma}_1 : \left|\left|\bar{X}_1 - \hat{\gamma}_1^\top X_1\right|\right|_\infty \leq \delta_n, \quad \hat{\gamma}_2 : \left|\left|\hat{\gamma}_2^\top H_2 - \hat{\gamma}_1^\top H_2\right|\right|_\infty \leq \delta_n$$

Balancing: Illustration

Estimating $\mathbb{E}[Y_2(1,1)]$:

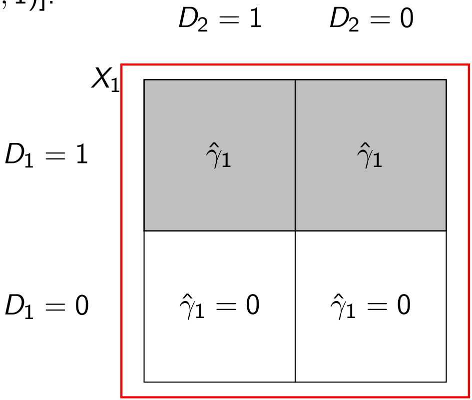
	$D_2 = 1$	$D_2 = 0$
$D_1 = 1$	$\hat{\gamma}_1$	$\hat{\gamma}_1$
$D_1 = 0$	$\hat{\gamma}_1 = 0$	$\hat{\gamma}_1 = 0$

Balancing conditions:

$$\hat{\gamma}_1 : \left|\left|\bar{X}_1 - \hat{\gamma}_1^\top X_1\right|\right|_\infty \leq \delta_n, \quad \hat{\gamma}_2 : \left|\left|\hat{\gamma}_2^\top H_2 - \hat{\gamma}_1^\top H_2\right|\right|_\infty \leq \delta_n$$

Balancing: Illustration

Estimating $\mathbb{E}[Y_2(1,1)]$:



Balancing conditions:

$$\hat{\gamma}_1 : \left|\left|\bar{X}_1 - \hat{\gamma}_1^\top X_1\right|\right|_\infty \leq \delta_n, \quad \hat{\gamma}_2 : \left|\left|\hat{\gamma}_2^\top H_2 - \hat{\gamma}_1^\top H_2\right|\right|_\infty \leq \delta_n$$

Balancing: Illustration

Estimating $\mathbb{E}[Y_2(1,1)]$:

	$D_2 = 1$	$D_2 = 0$
$D_1 = 1$		$\hat{\gamma}_2 = 0$
$D_1 = 0$	$\hat{\gamma}_2 = 0$	$\hat{\gamma}_2 = 0$

Balancing conditions:

$$\hat{\gamma}_1 : \left|\left|\bar{X}_1 - \hat{\gamma}_1^\top X_1\right|\right|_\infty \leq \delta_n, \quad \hat{\gamma}_2 : \left|\left|\hat{\gamma}_2^\top H_2 - \hat{\gamma}_1^\top H_2\right|\right|_\infty \leq \delta_n$$

Balancing: Illustration

Estimating $\mathbb{E}[Y_2(1,1)]$:

H_2

	$D_2 = 1$	$D_2 = 0$
$D_1 = 1$	$\hat{\gamma}_2$	$\hat{\gamma}_2 = 0$
$D_1 = 0$	$\hat{\gamma}_2 = 0$	$\hat{\gamma}_2 = 0$

Balancing conditions:

$$\hat{\gamma}_1 : \left|\left|\bar{X}_1 - \hat{\gamma}_1^\top X_1\right|\right|_\infty \le \delta_n, \quad \hat{\gamma}_2 : \left|\left|\hat{\gamma}_2^\top H_2 - \hat{\gamma}_1^\top H_2\right|\right|_\infty \le \delta_n$$

Estimators

- ⇒ Balancing to guarantee neglegible ($o_p(1/\sqrt{n})$) estimation error;
- ⇒ It does not require the specification (and estimation) of the propensity score.

Two periods: $\hat{\mu}_{1,1} = \hat{\gamma}_2^\top \underbrace{(Y_2 - H_2\hat{\beta}^2_{1,1})}_{Y_2 - Pred2} + \hat{\gamma}_1^\top \underbrace{(H_2\hat{\beta}^2_{1,1} - X_1\hat{\beta}^1_{1,1})}_{Pred2 - Pred1} + \underbrace{\bar{X}_1\hat{\beta}^1_{1,1}}_{Est1}$

Estimators

- ⇒ Balancing to guarantee neglegible ($o_p(1/\sqrt{n})$) estimation error;
- ⇒ It does not require the specification (and estimation) of the propensity score.

Two periods: $$\hat{\mu}_{1,1} = \hat{\gamma}_2^\top \underbrace{\left(Y_2 - H_2\hat{\beta}^2_{1,1}\right)}_{Y_2 - Pred2} + \hat{\gamma}_1^\top \underbrace{\left(H_2\hat{\beta}^2_{1,1} - X_1\hat{\beta}^1_{1,1}\right)}_{Pred2 - Pred1} + \underbrace{\bar{X}_1\hat{\beta}^1_{1,1}}_{Est1}$$

Step t (Sequential)

$$\hat{\gamma}_t = \text{argmin}_{\gamma_t} ||\gamma_t||^2$$

$$\Big|\Big| \sum_{i=1}^n \hat{\gamma}_{i,t-1} H_{i,t} - \sum_{i=1}^n \gamma_{i,t} H_{i,t} \Big|\Big|_\infty \le \delta(n,p), \quad \sum_{i=1}^n \gamma_{i,t} = 1 .$$

$$||\gamma_t||_\infty \le \log(n) n^{-2/3}, \quad 0 \le \gamma_{i,t} \le \prod_{s=1}^t 1\{D_{i,s} = d_s\}$$

Rationale

Lemma 1

For any $\hat{\gamma}_1$, and $\hat{\gamma}_2$, where $\hat{\gamma}_{i,2} = 0$ if $(D_{i,1}, D_{i,2}) \neq (1,1)$, we have

$$\hat{\mu}(1,1) - \mu(1,1) = \underbrace{(\beta^1_{1,1} - \hat{\beta}^1_{1,1})^\top \left(\bar{X}_1 - \hat{\gamma}_1^\top X_1\right)}_{BIAS1} + \underbrace{(\beta^2_{1,1} - \hat{\beta}^2_{1,1})^\top \left(\hat{\gamma}_2^\top H_2 - \hat{\gamma}_1^\top H_2\right)}_{BIAS2}$$
$$+ \underbrace{\hat{\gamma}_2^\top \left(Y_2 - H_2\beta^2_{1,1}\right)}_{\text{Error 1}} + \underbrace{\hat{\gamma}_1^\top H_2\beta^2_{1,1} - \hat{\gamma}_1^\top X_1\beta^1_{1,1}}_{\text{Error 2}}. \quad (1)$$

⇒ Balancing control Bias1 and Bias2. Bias2 arises due to dynamics

? Why should we estimate γ_1, γ_2 sequentially?

Residuals from balancing

Lemma 2

Let the sigma algebra $\sigma(\hat{\gamma}_1) \subseteq \sigma(X_1, D_1)$ and $\hat{\gamma}_{i,1} = 0$ if $D_{i,1} \neq 1$. Then

$$\mathbb{E}\Big[\hat{\gamma}_{i,1} H_{i,2} \beta^2_{1,1} - \hat{\gamma}_{i,1} X_{i,1} \beta^1_{1,1} \Big| X_{i,1}, D_{i,1}\Big] = 0.$$

Intuition:

- If $D_{i,1} \neq 1$ then the expression is zero by construction;
- If $D_{i,1} = 1$, we can use the law of iterated expectations;
- If $\hat{\gamma}_1$ depends on future observations, or $\hat{\gamma}_{i,1} = 1$ for untreated units, the expression is not necessarily zero.

Guarantees

Under regularity assumptions:

Thm1 For n large enough, the optimization problem admits a feasible solution which includes stabilized inverse-probability weights;

Guarantees

Under regularity assumptions:

Thm1 For n large enough, the optimization problem admits a feasible solution which includes stabilized inverse-probability weights;

Thm2 For $\log(Tpn)/n^{1/4} \to 0$ (high-dimensions)

$$\hat{\mu}_T(d_{1:T}) - \mu_T(d_{1:T}) = \mathcal{O}_p(n^{-1/2}).$$

Also, $\sqrt{n}||\hat{\gamma}_t||_2 = O_p(1)$;

Guarantees

Under regularity assumptions:

Thm1 For n large enough, the optimization problem admits a feasible solution which includes stabilized inverse-probability weights;

Thm2 For $\log(Tpn)/n^{1/4} \to 0$ (high-dimensions)

$$\hat{\mu}_T(d_{1:T}) - \mu_T(d_{1:T}) = \mathcal{O}_p(n^{-1/2}).$$

Also, $\sqrt{n}||\hat{\gamma}_t||_2 = O_p(1)$;

Thm3 Inference with chi-squared ($\chi_T(\alpha)$) and Gaussian critical values:

$$\lim_{n\to\infty} P\Big(\Big|\frac{\sqrt{n}\Big(\hat{\mu}_T(d_{1:T}) - \mu_T(d_{1:T})\Big)}{\hat{V}_T(d_{1:T})^{1/2}}\Big| > \sqrt{\chi_T(\alpha)}\Big) \leq \alpha.$$

Guarantees

Under regularity assumptions:

Thm1 For n large enough, the optimization problem admits a feasible solution which includes stabilized inverse-probability weights;

Thm2 For $\log(Tpn)/n^{1/4} \to 0$ (high-dimensions)

$$\hat{\mu}_T(d_{1:T}) - \mu_T(d_{1:T}) = \mathcal{O}_p(n^{-1/2}).$$

Also, $\sqrt{n}||\hat{\gamma}_t||_2 = O_p(1)$;

Thm3 Inference with chi-squared ($\chi_T(\alpha)$) and Gaussian critical values:

$$\lim_{n\to\infty} P\Big(\Big|\frac{\sqrt{n}\Big(\hat{\mu}_T(d_{1:T}) - \mu_T(d_{1:T})\Big)}{\hat{V}_T(d_{1:T})^{1/2}}\Big| > \sqrt{\chi_T(\alpha)}\Big) \leq \alpha.$$

$\Rightarrow$ Confidence with Gaussian critical values if $\sqrt{n}||\hat{\gamma}_t||_2$ converges almost surely (e.g., for Bernoulli design).

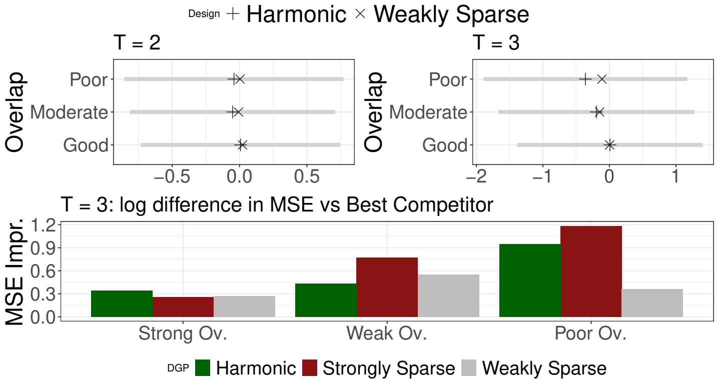
Remarks

- Trade-offs for δ_n: larger δ_n guarantees feasibility, but increases bias;
- Bias is of order $||\hat{\beta} - \beta||_1 \delta_n$. Longer T implies larger estimation error unless we assume limited carry-over effects;
- In the paper: algorithmic procedure to find the smallest feasible δ_n.

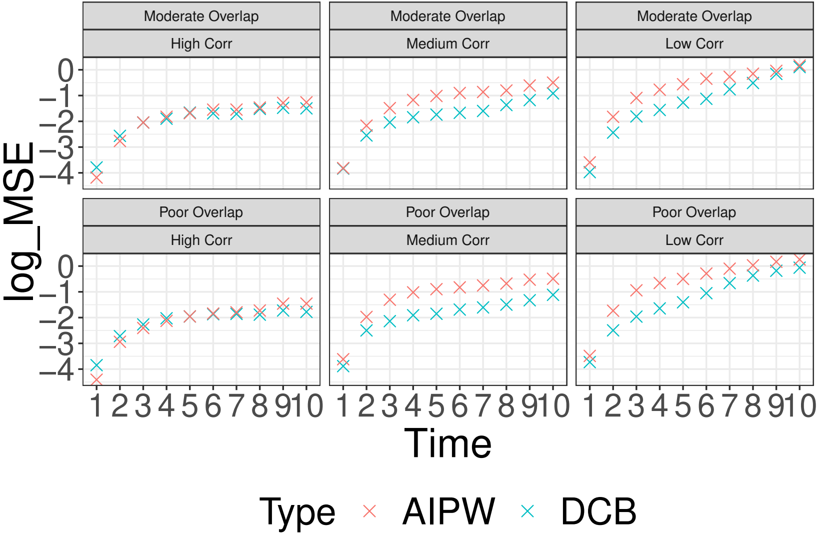
Content

Numerical Study

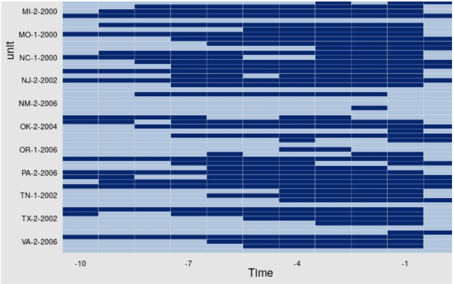
- Designs: $\beta^{(j)} = 1/j, \beta^{(j)} = 1/j^2$
- Linear model, different level of overlap.



Comparison with AIPW

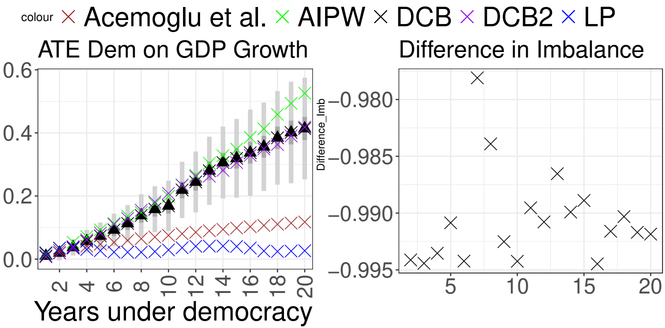


Advertisment and Elections



	ATE DCB	ATE AIPW
S1	-1.767	-0.57
	(0.70)	(1.45)
S2	-0.493	-0.22
	(0.764)	(1.33)

Democracy and Growth



Content

Conclusion

- We propose a method for estimating dynamic causal effects;
- We provide an estimation and inferential procedure using novel covariate balancing conditions;
- We characterize asymptotic properties of the estimator, and study its finite sample properties in numerical studies and empirical applications.

Questions?

Link `dviviano.github.io/projects`