Causal Inference with Panel Data

Lecture 3: Factor-Augmented Methods

Yiqing Xu (Stanford University)
Washington University in St. Louis

25 August 2021

This Lecture

- The interactive fixed effect model
- The matrix completion method
- Diagnostics
- Bayesian multi-factor models

Link with Synthetic Control

- Recall that ADH (2010) use a factor-augmented model to motivate the synthetic control method:

$$Y_{it}(0) = \theta_t' Z_i + \xi_t + \lambda_i' f_t + \varepsilon_{it}$$

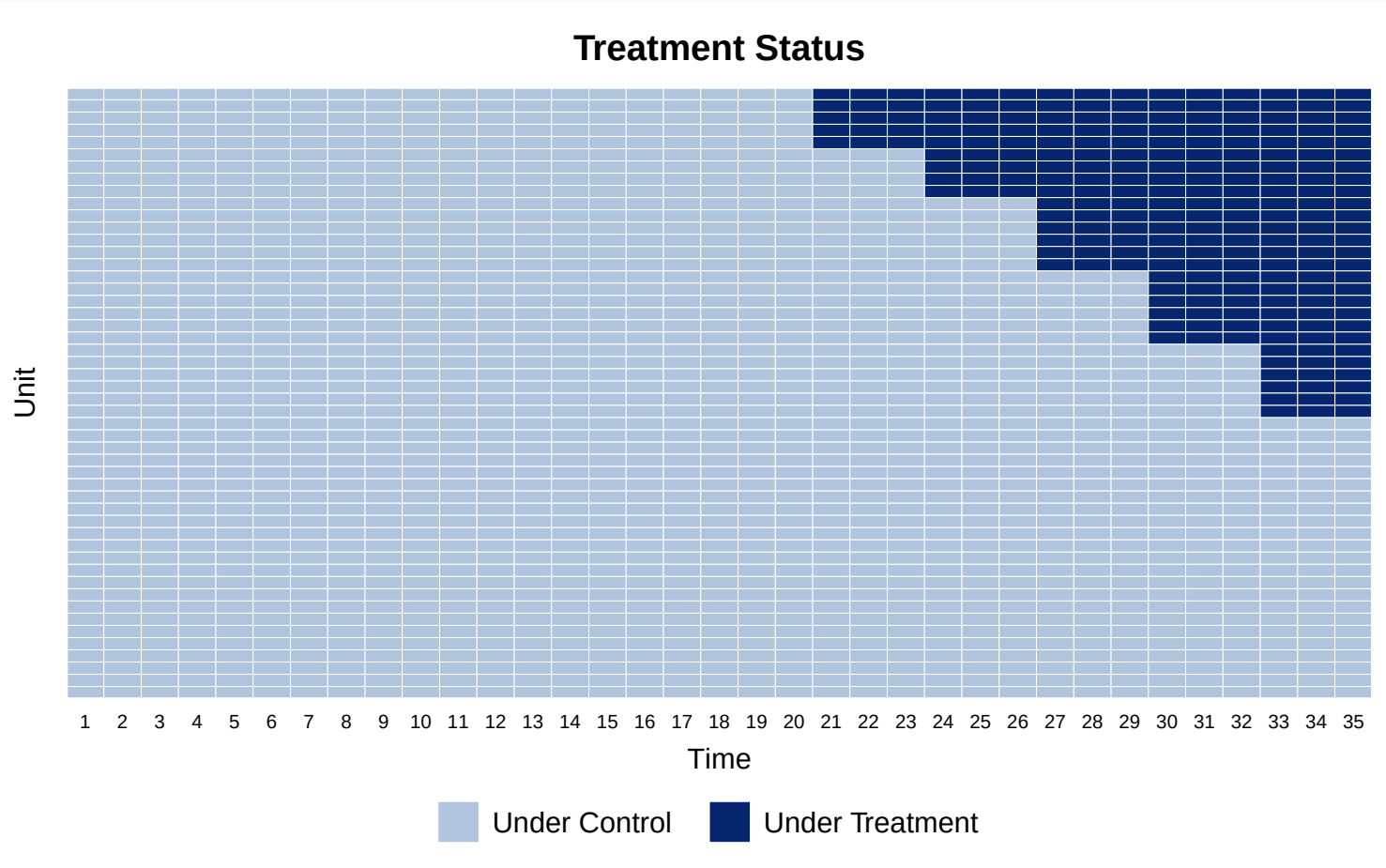
- What if we actually estimate the model using observations under the control condition only?
- Xu (2017) imports the so-called interactive fixed-effect (IFE) model to a DiD setting

$$Y_{it}(0) = X_{it}'\beta + \alpha_i + \xi_t + \lambda_i' f_t + \varepsilon_{it}$$

- Athey et al. (2021) extend it and introduce the matrix completion method
- Liu, Wang & Xu (2021) put these methods in a general framework — "the counterfactual estimators"
- No negative weighting!
- Limitations: needs large T and N; model dependency

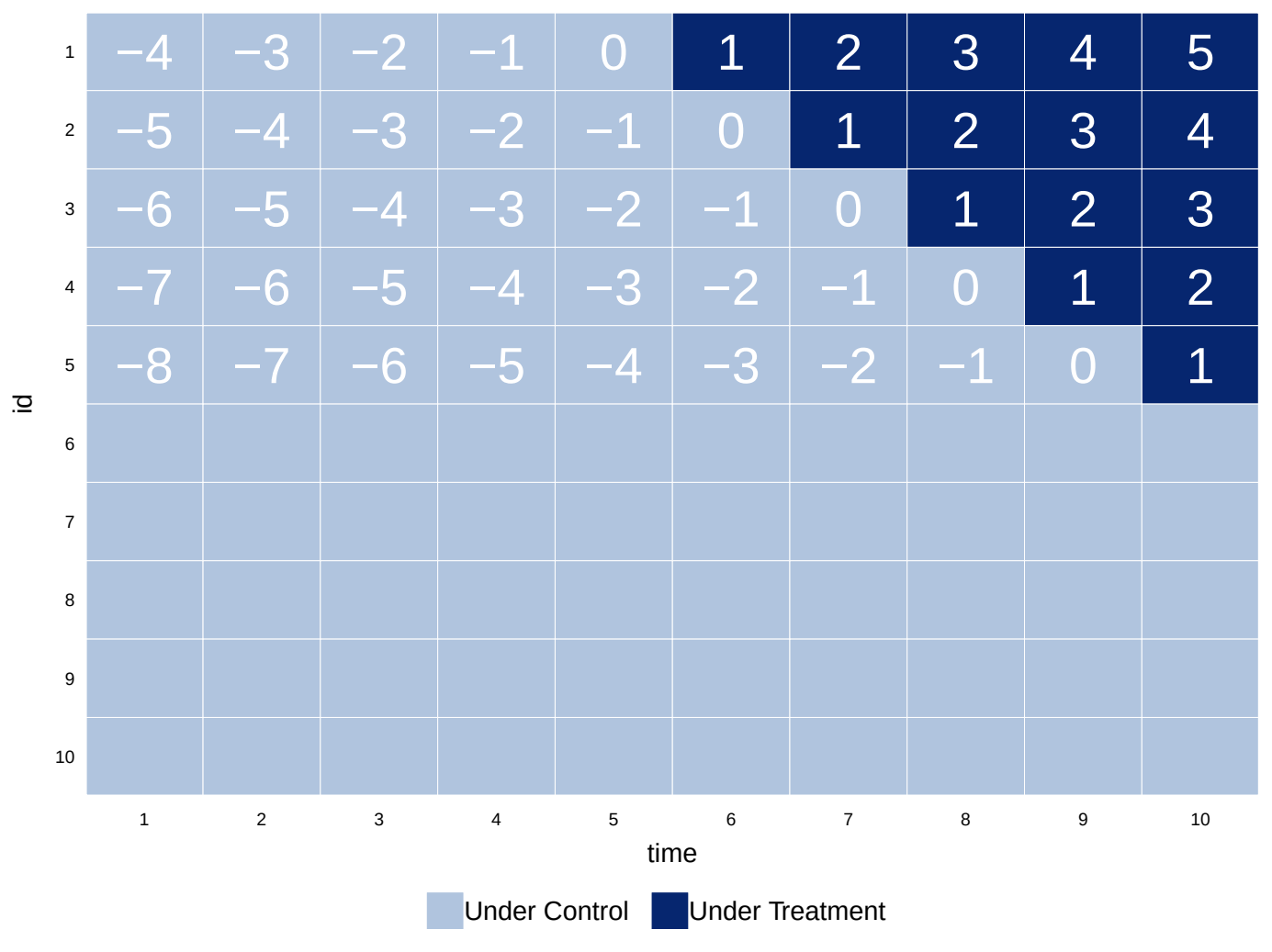
Basic Idea

- In a panel setting, treat $Y(1)$ as missing data
- Predict $Y(0)$ based on an outcome model
- (Use pre-treatment data for model selection)
- Estimate ATT by averaging differences between $Y(1)$ and $\hat{Y}(0)$

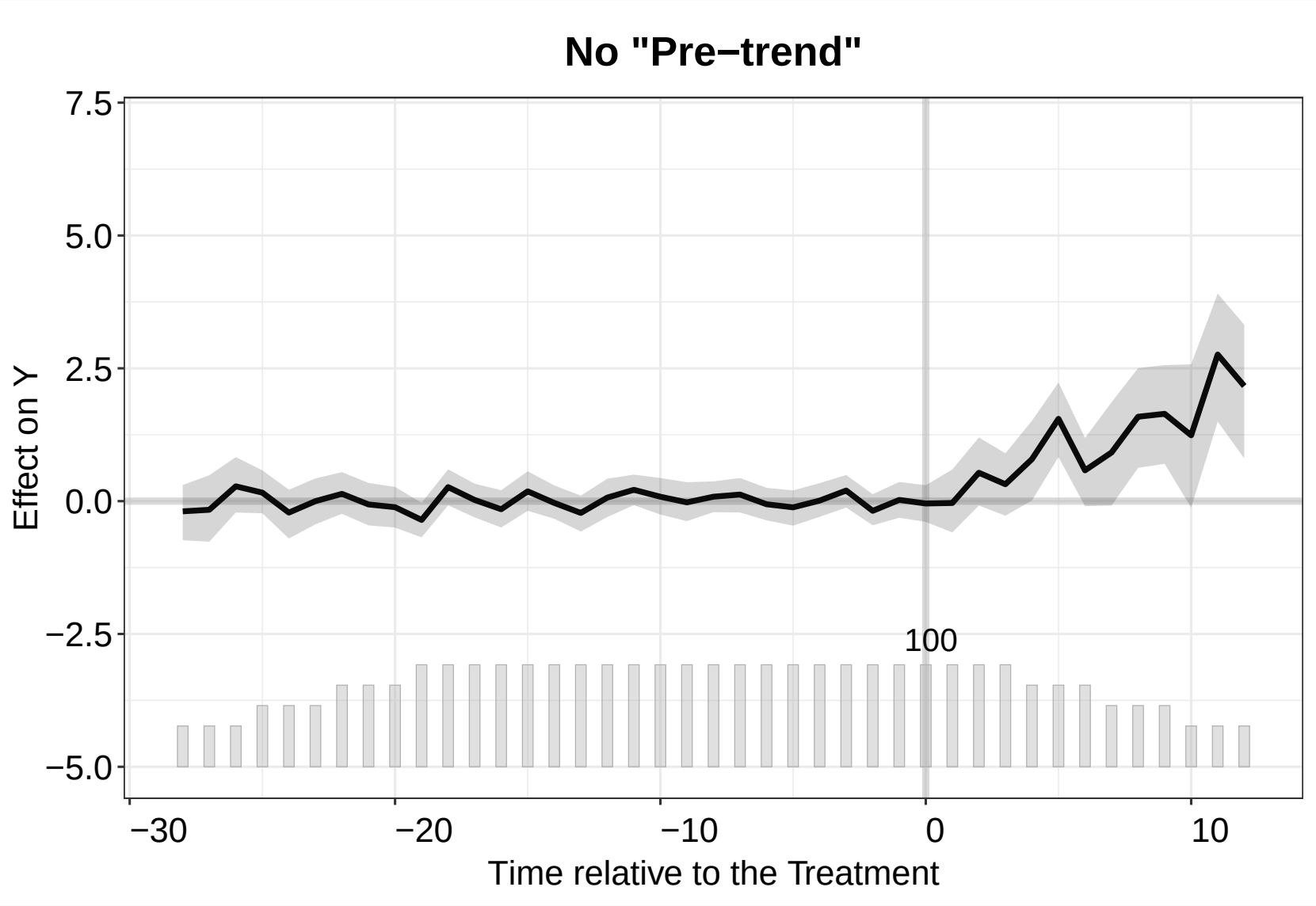


Basic Idea

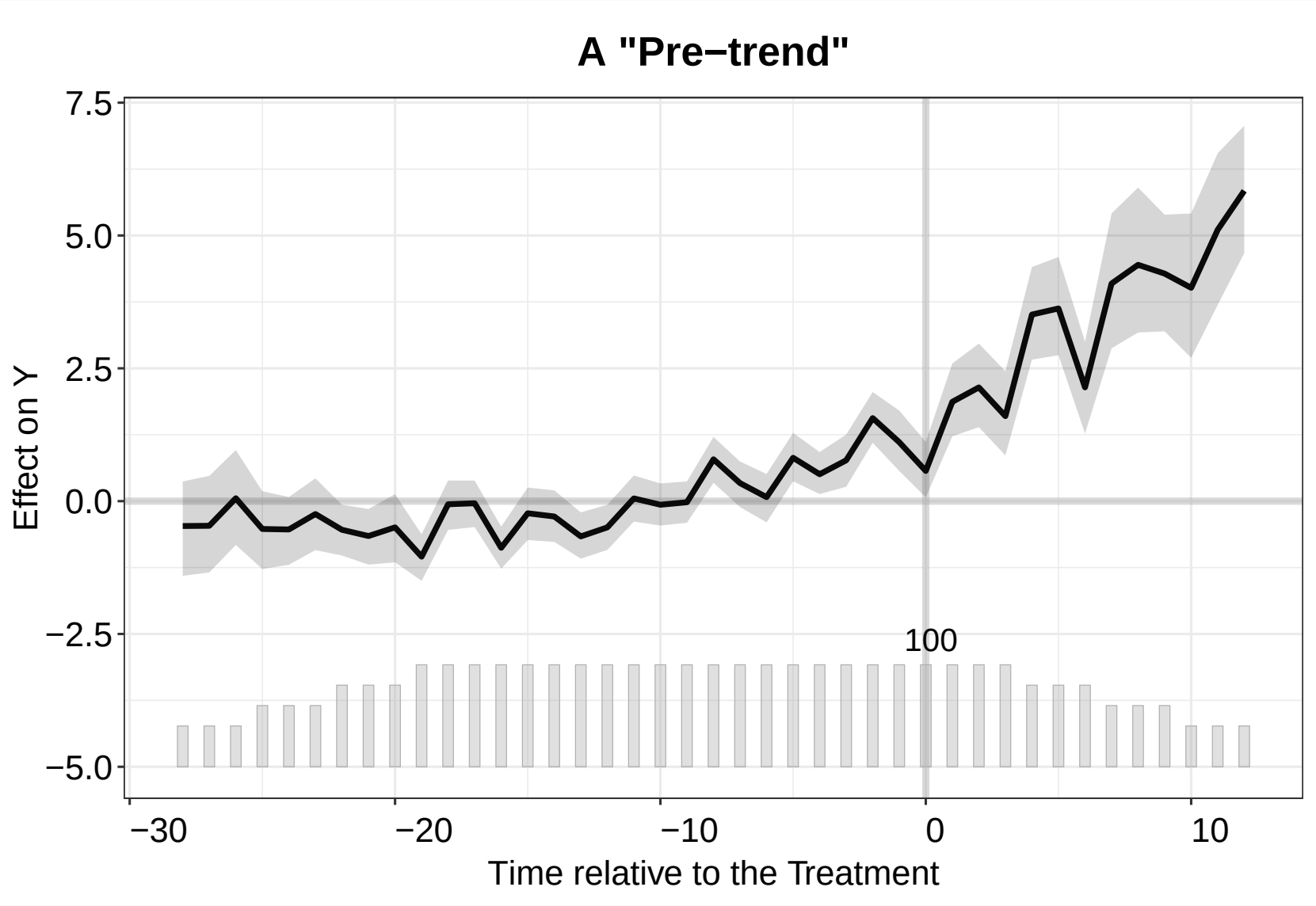
$$\widehat{ATT}_s = \hat{\mathbb{E}}[\hat{\tau}_{it} | D_{i,t-s} = 0, \underbrace{D_{i,t-s+1} = D_{i,t-s+2} = \cdots = D_{it} = 1}_{s \text{ periods}}, \forall i \in \mathcal{T}].$$



A New Plot for "Dynamic Treatment Effects"



A New Plot for "Dynamic Treatment Effects"



Current Practice: *Angrist & Pischke* Chapter 5

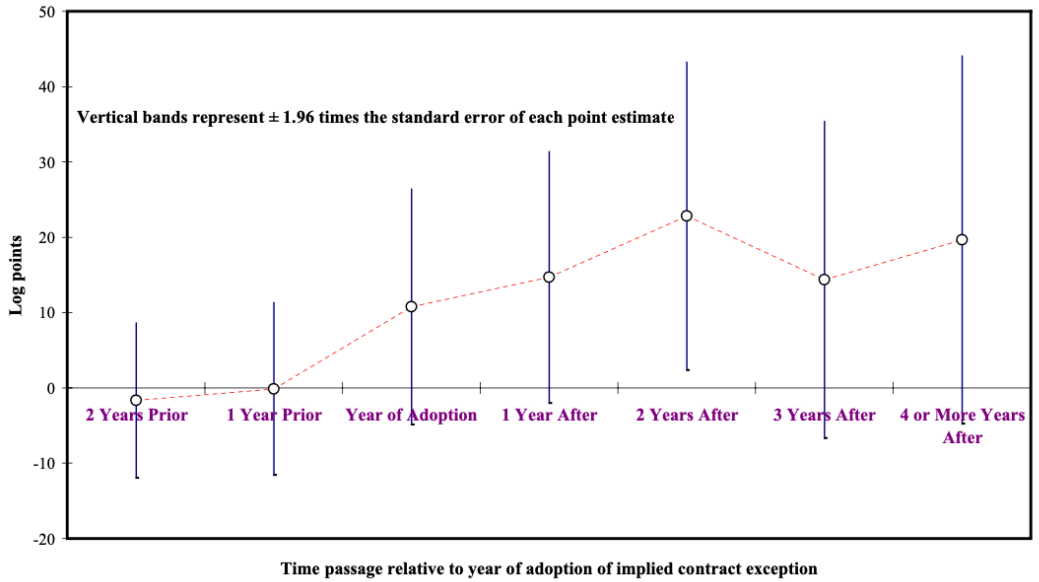


Figure 5.2.4: Estimated impact of state courts' adoption of an implied-contract exception to the employment-at-will doctrine on use of temporary workers (from Autor 2003). The dependent variable is the log of state temporary help employment in 1979 - 1995. Estimates are from a model that allows for effects before, during, and after adoption.

Model-based Counterfactual Estimators

A model-based counterfactual estimator proceeds in the following steps:

- Step 1. Train the model using observations under the control condition ($D_{it} = 0$).
- Step 2. Predict the counterfactual outcome $\hat{Y}_{it}(0)$ for each observation under the treatment condition ($D_{it} = 1$) and obtain an estimate of the individual treatment effect: $\hat{\tau}_{it} = Y_{it} - \hat{Y}_{it}(0)$.
- Step 3. Generate estimates for the causal quantities of interest

$$ATT = \mathbb{E}[\tau_{it} | D_{it} = 1, \forall i \in \mathcal{T}, \forall t], \qquad \text{or}$$

$$ATT_s = \mathbb{E}[\tau_{it} | D_{i,t-s} = 0, \underbrace{D_{i,t-s+1} = D_{i,t-s+2} = \cdots = D_{it} = 1}_{s \text{ periods}}, \forall i \in \mathcal{T}].$$

Review of Three Estimators

We review three estimation strategies:

- FEct:

$$\hat{Y}_{it}(0) = X_{it}\hat{\beta} + \hat{\alpha}_i + \hat{\xi}_t$$

- IFEct (Gobillon&Magnac 2016; Xu 2017):

$$\hat{Y}_{it}(0) = X_{it}\hat{\beta} + \hat{\lambda}_i'\hat{F}_t$$

- Matrix Completion (MC) (Athey et al. 2018):

$$\hat{Y}_{it}(0) = X_{it}\hat{\beta} + \hat{L}_{it},$$

  where matrix $\{L_{it}\}_{N\times T}$ is a lower-rank matrix approximation of $\{Y(0)\}_{N\times T}$ with missing values

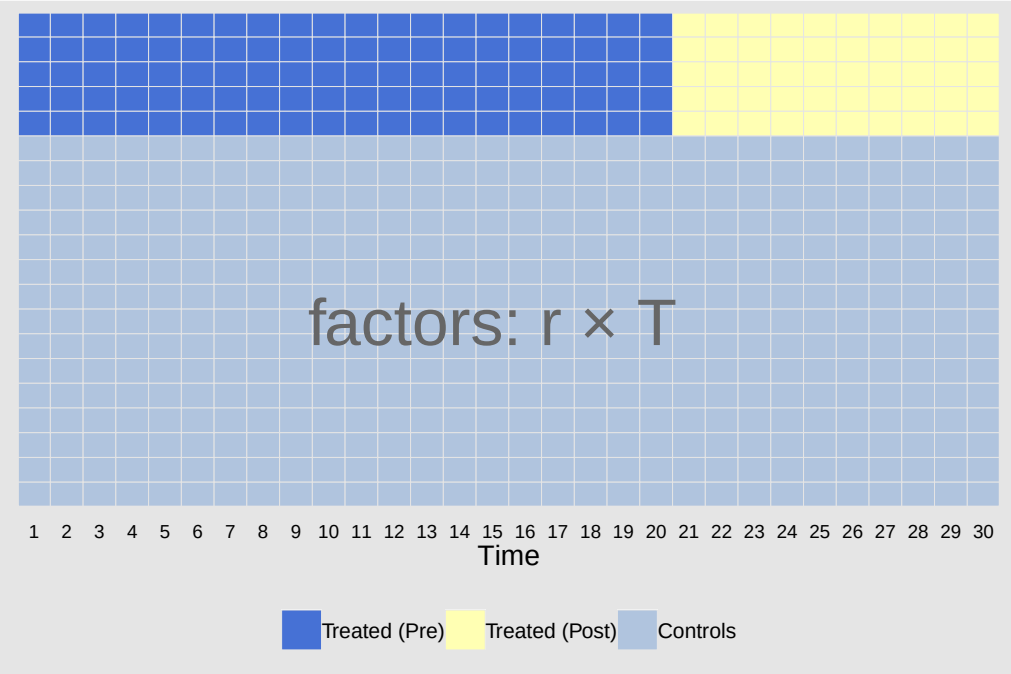
Remarks:

- DiD is a special case of FEct
- Both IFEct and MC are estimated via iterative algorithms
- Cross-validation to choose the tuning parameter

Xu (2017) proposes a three-step approach based on a latent factor model:

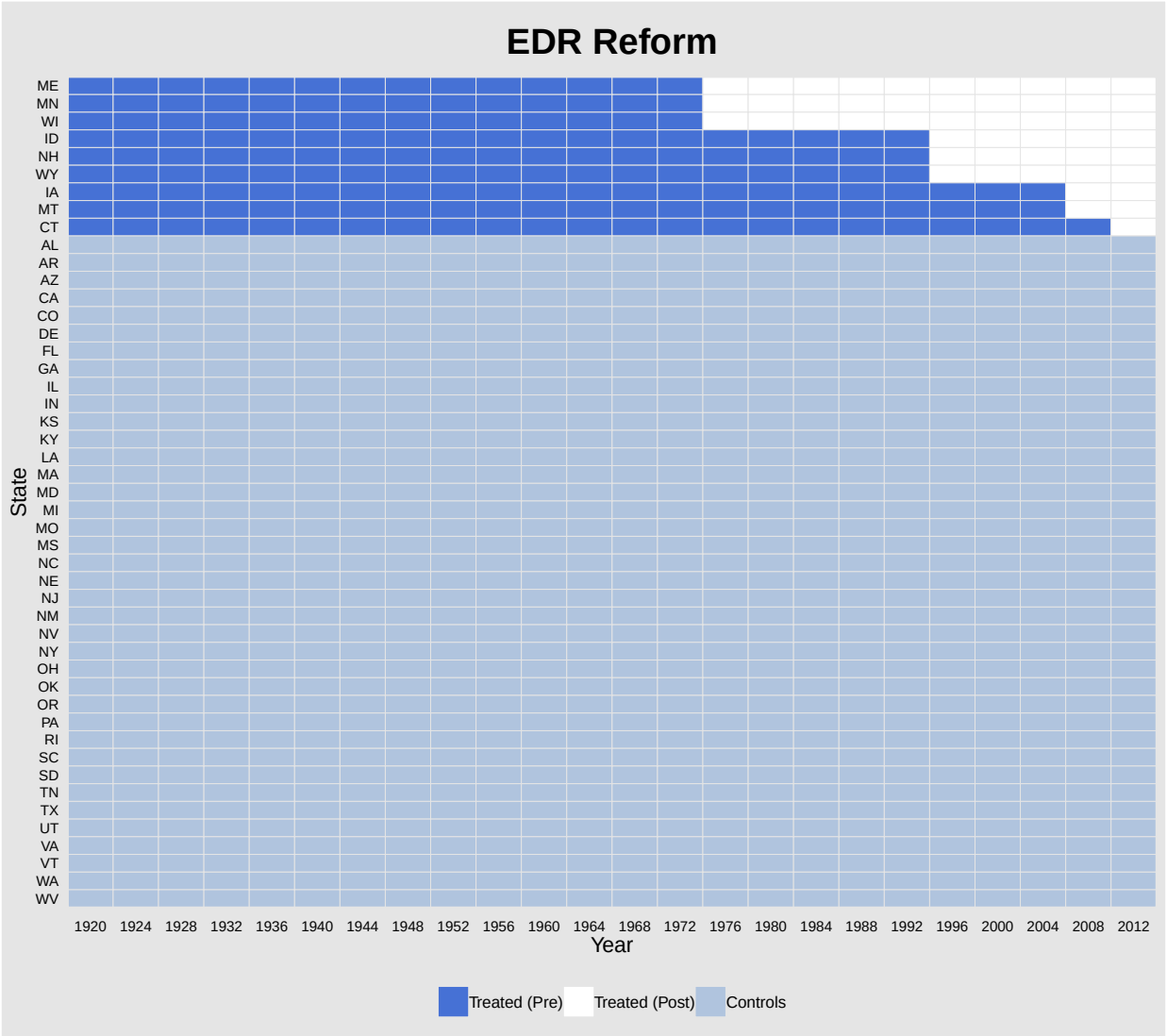
$$
\begin{aligned}
\text{Control} \quad & Y_{it}(0) = X_{it}'\beta + \alpha_i + \xi_t + \lambda_i' f_t + \varepsilon_{it} \\
\text{Treated} \quad & Y_{it}(0) = X_{it}'\beta + \alpha_i + \xi_t + \lambda_i' f_t + \varepsilon_{it} \qquad \text{(pre)} \\
& Y_{it}(1) = X_{it}'\beta + \alpha_i + \xi_t + \lambda_i' f_t + \varepsilon_{it} + \tau_{it} \qquad \text{(post)}
\end{aligned}
$$

1. Expectation-Maximization
   (Gobillon & Magnac 2016)

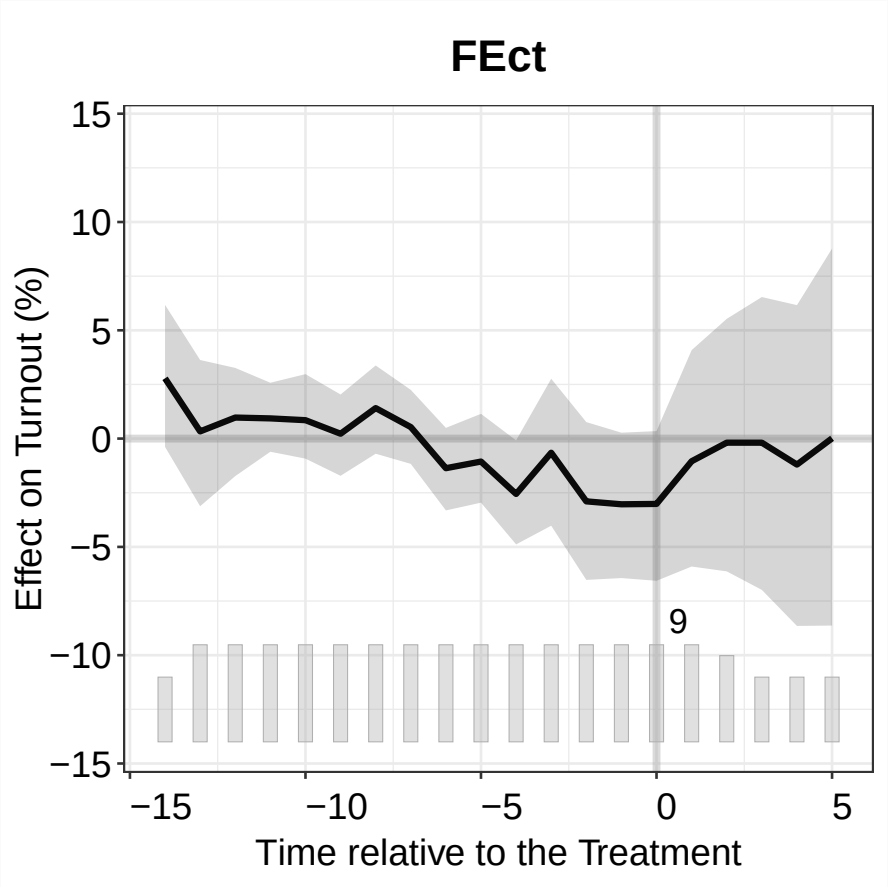


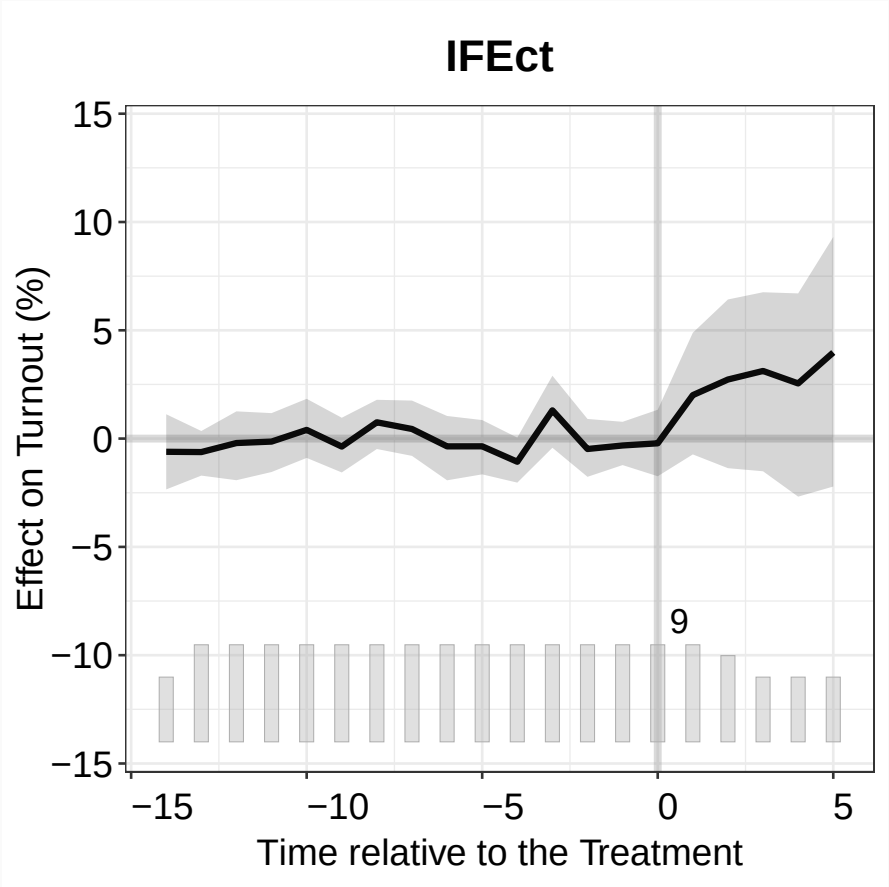
Election Day Registration (EDR) and Voter Turnout

Causal inference is a missing data problem.

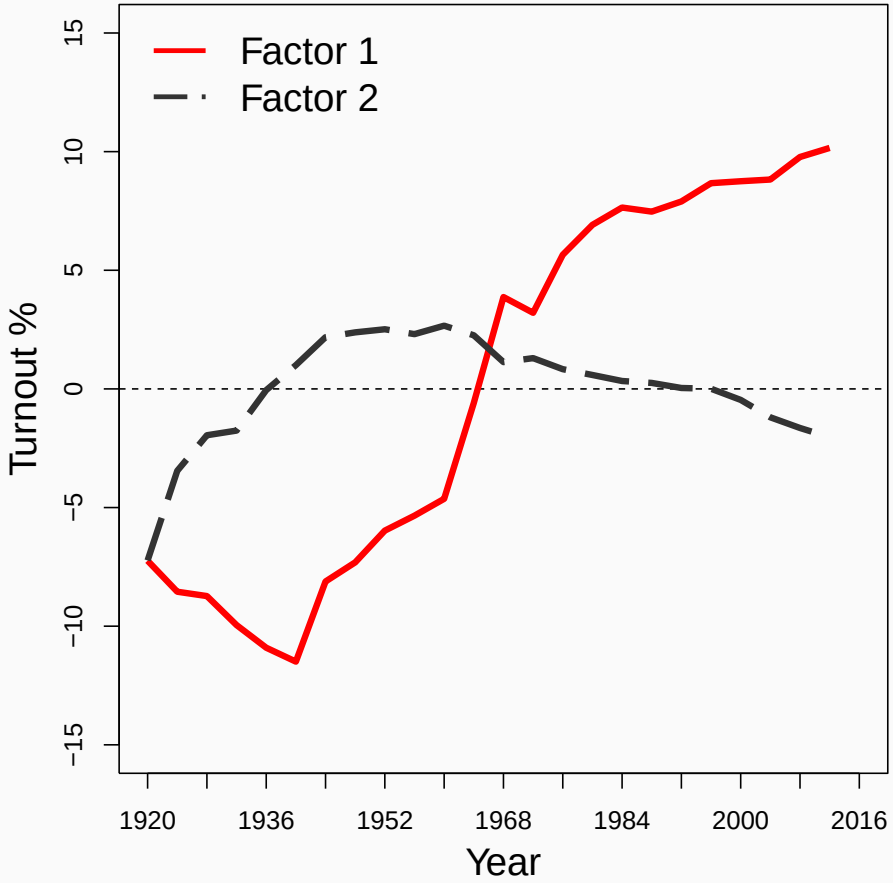


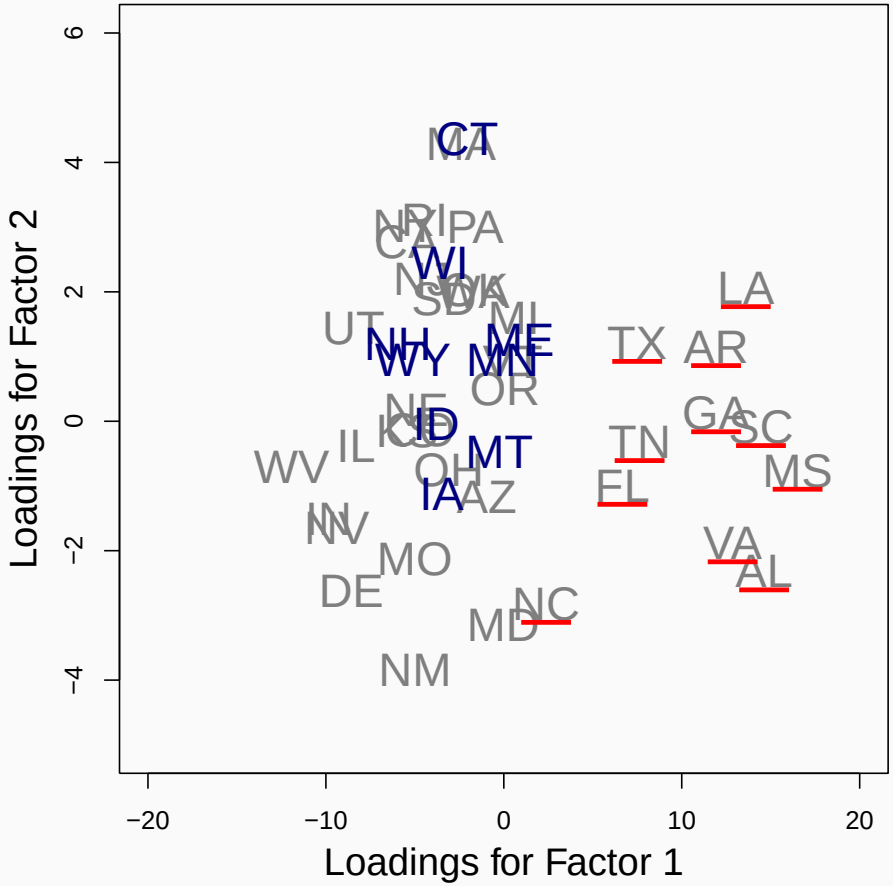
Main Results





Factors and Factor Loadings



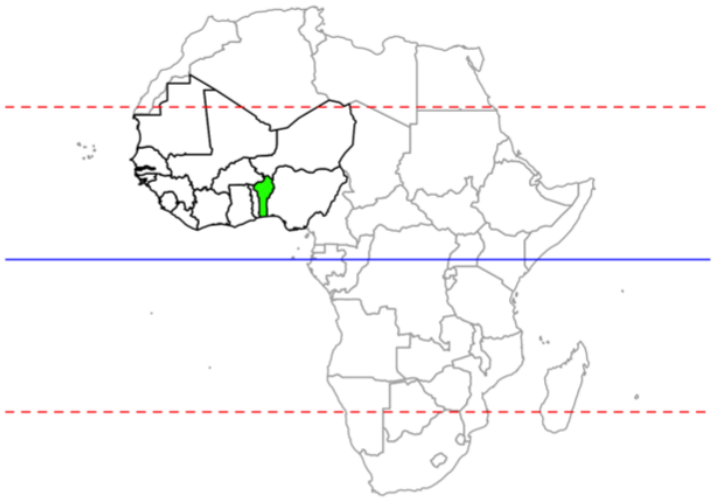


Example: Property Rights and Land Improvement

Sanford (2019)

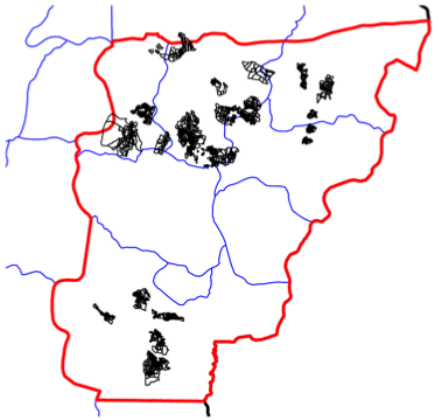
- Does property rights lead to improved land quality?
- "Experiment": giving peasants in Borgou, Benin land titles
- Use satellite (remote sensing) data to measure land improvement, i.e., switch from annual crops to perennial crops (bushes and trees)
- Use IFEct to construct counterfactuals

Benin

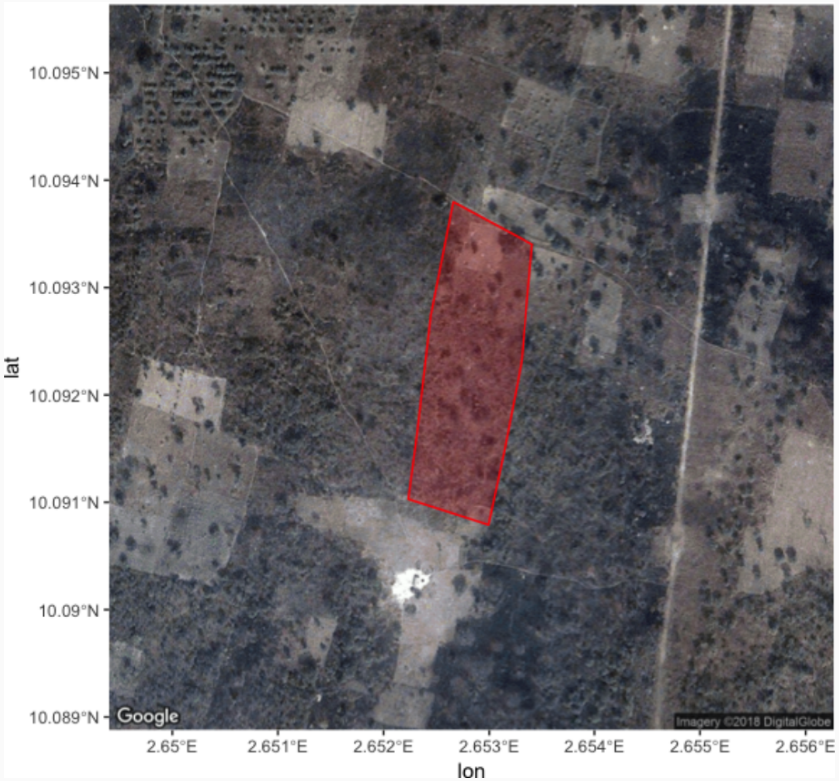


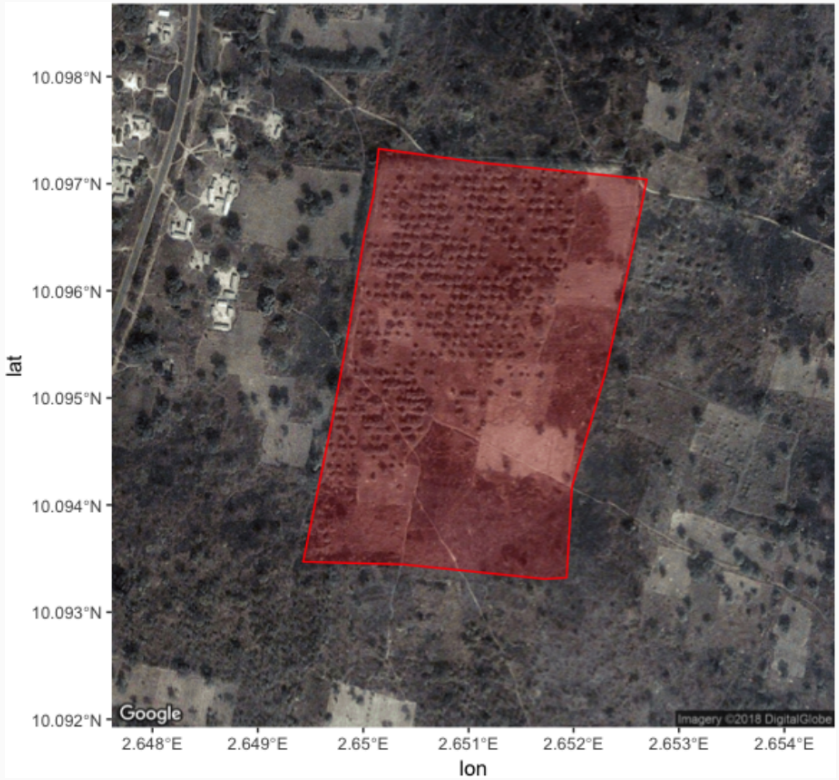
Borgou Department

Plots in Borgou

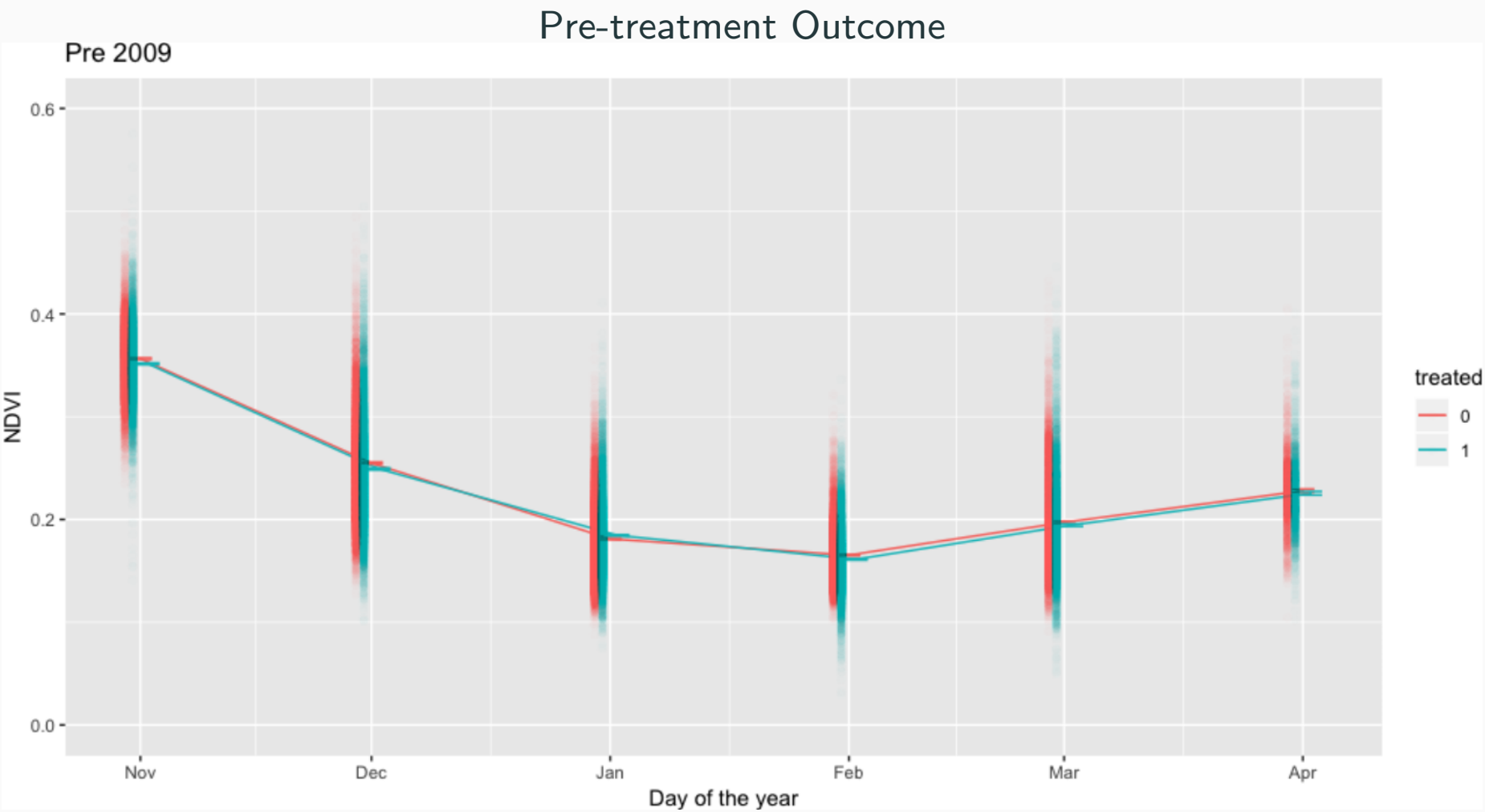


Original Satellite Images

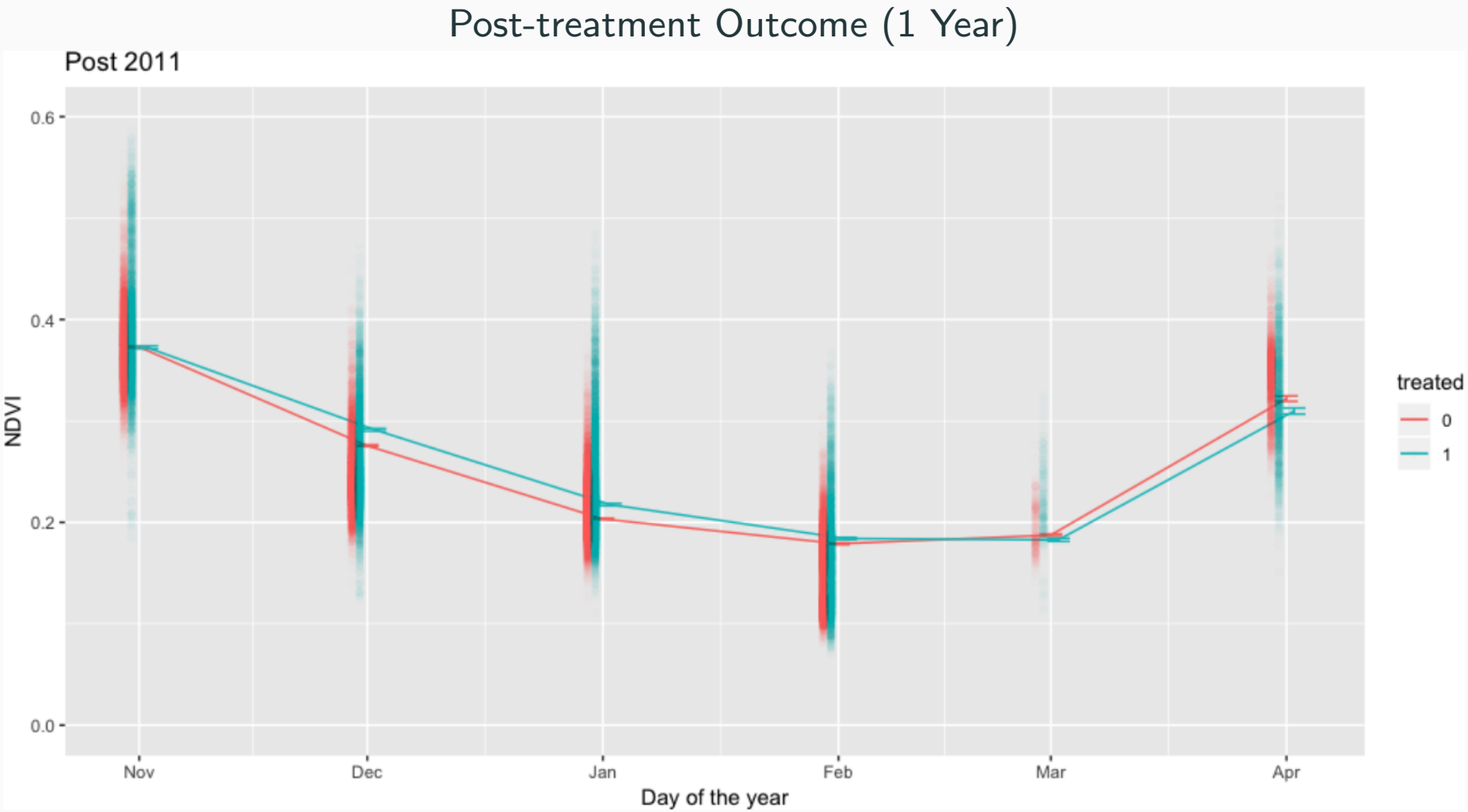




Treated and Counterfactual Averages

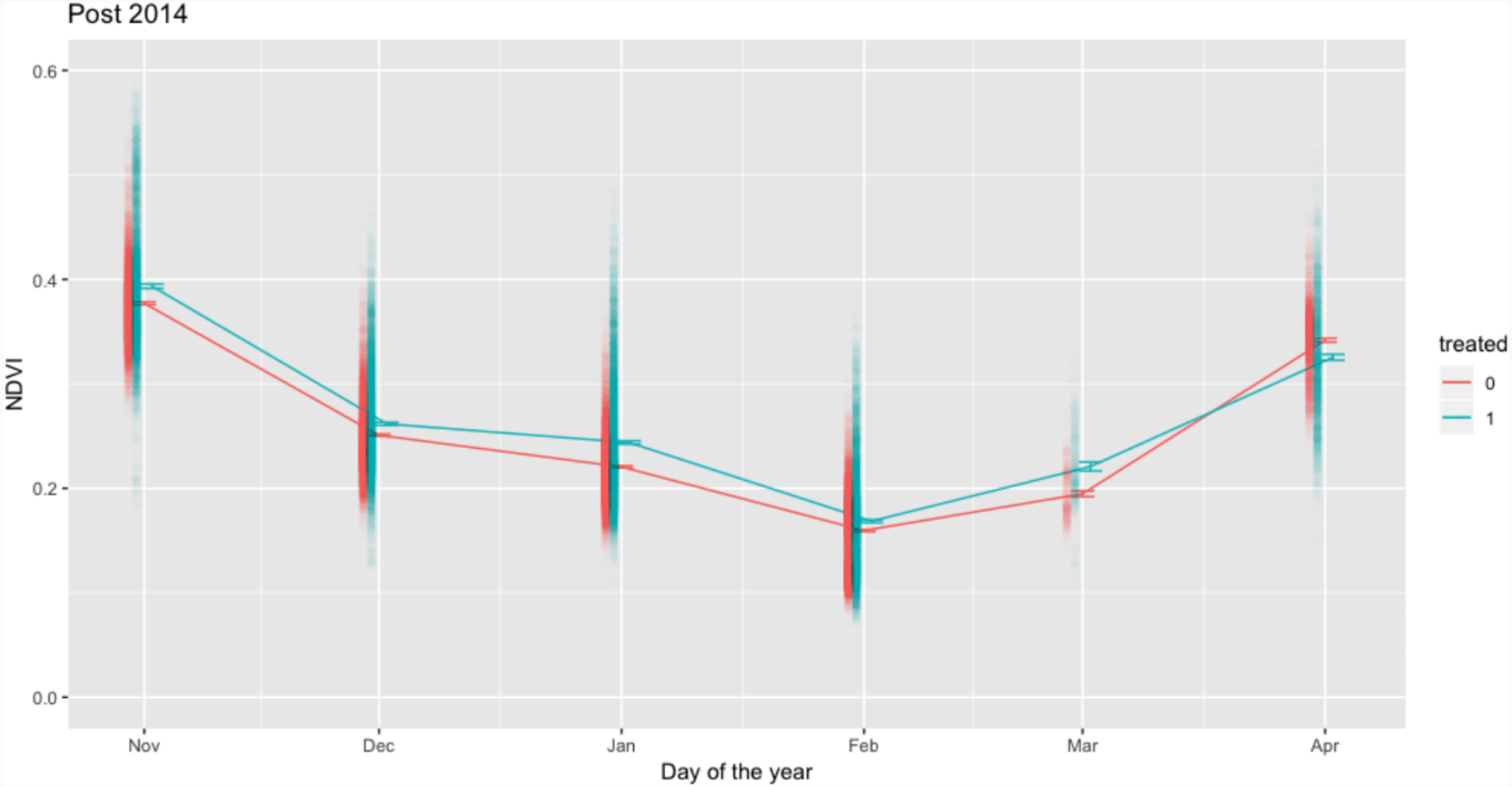


Treated and Counterfactual Averages

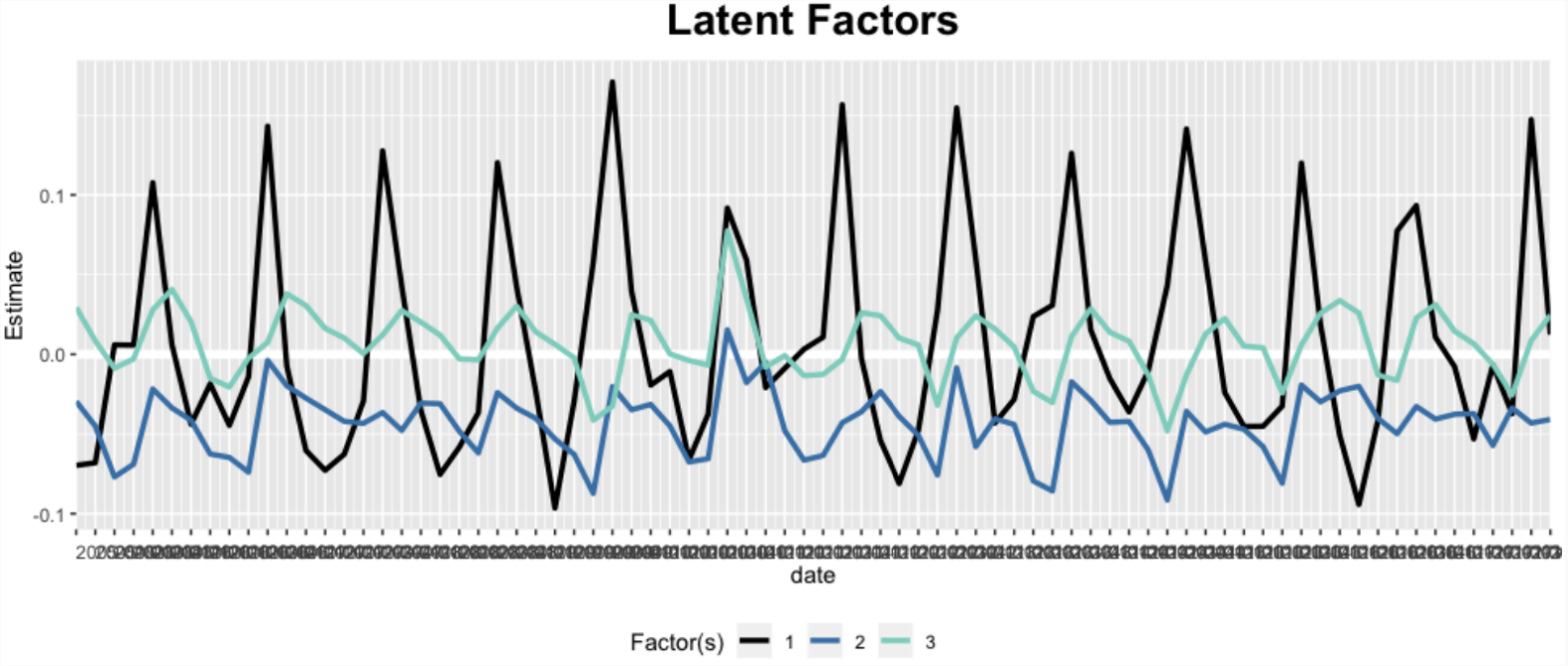


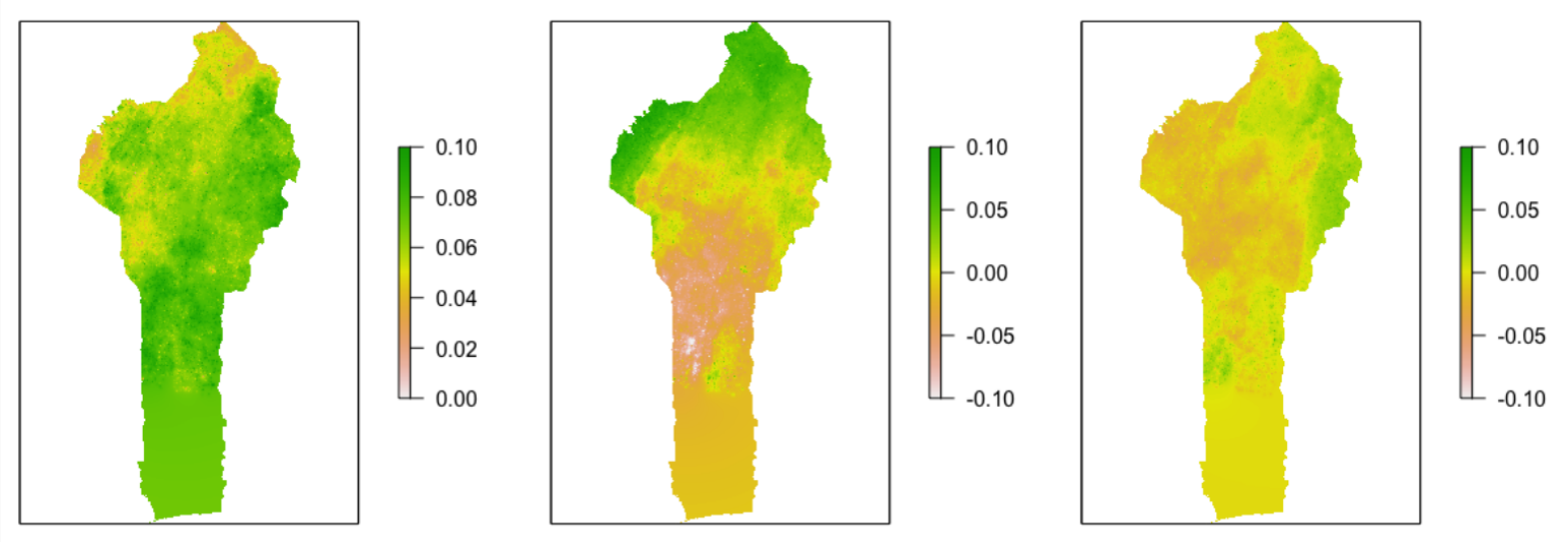
Treated and Counterfactual Averages

Post-treatment Outcome (4 Years)



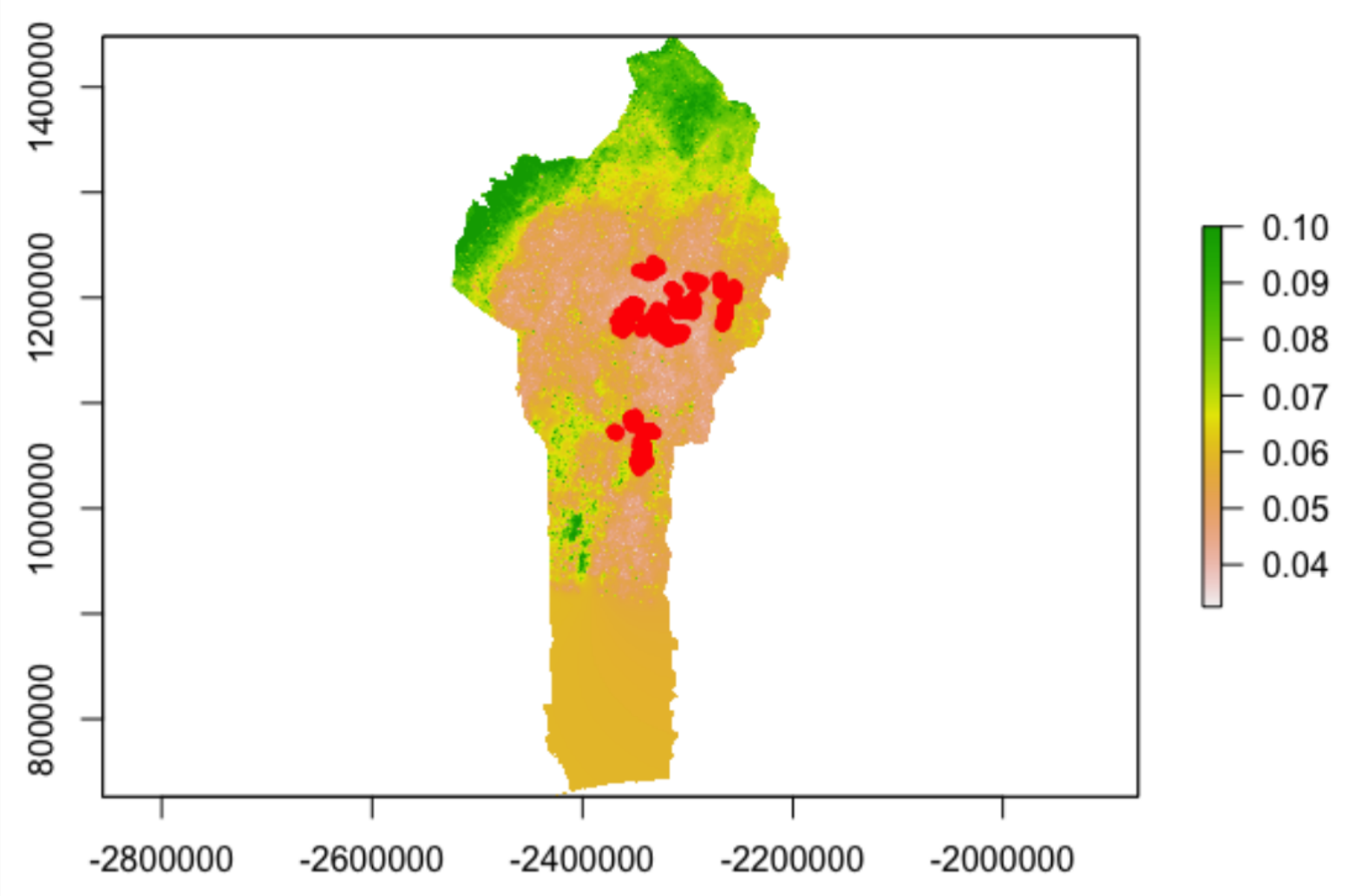
Factors and Loadings



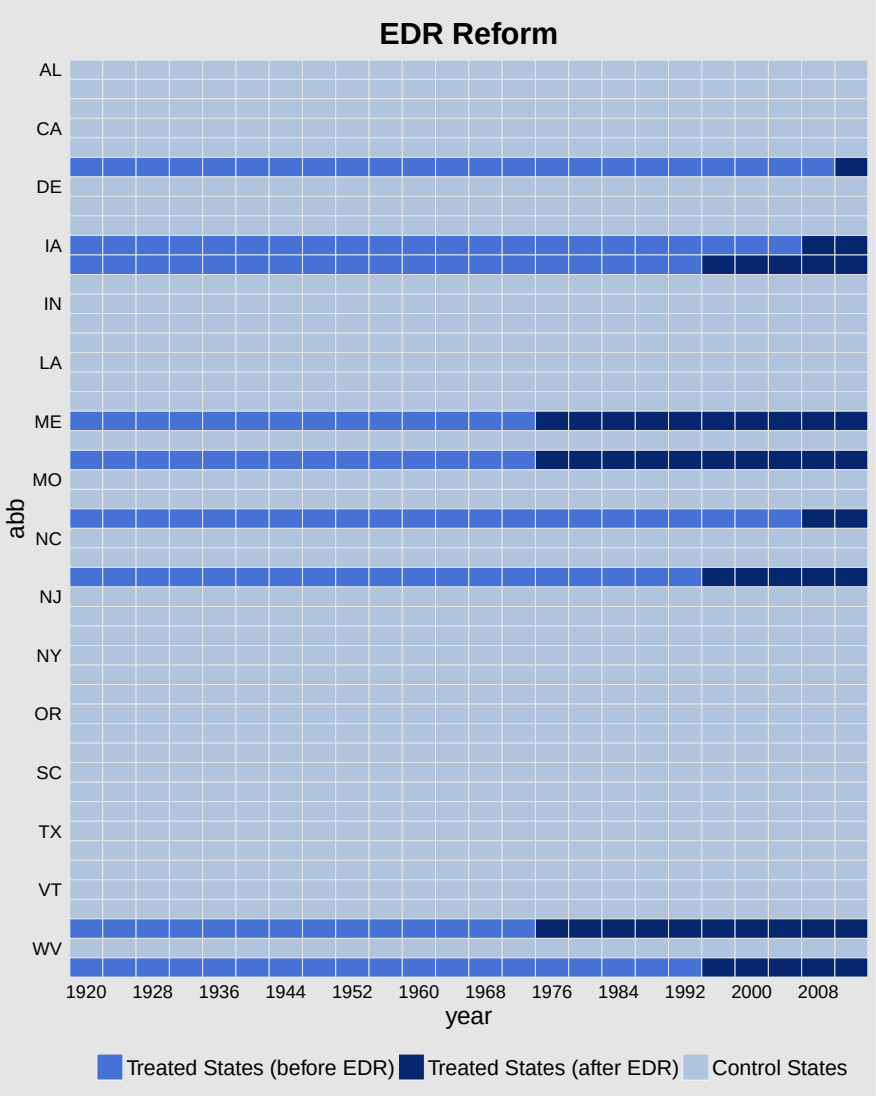


Geographic Distribution of Heavily-weighted Controls

Bigger number represents higher dissimilarity



Matrix Completion Methods



- Recall that our main goal is to predict treated counterfactuals
- Taking advantage of the matrix structure, matrix completion methods use non-treated data to achieve this goal
- The basic idea to find a lower-rank representation of the matrix to impute the "missing data"
- Xu (2017) is a special case of this approach

Matrix Completion Methods

- Recall in the baseline DiD setup:

$$\mathbf{Y} = \begin{pmatrix} Y^0_{\mathcal{T},pre} & ?? \\ Y^0_{\mathcal{C},pre} & Y^0_{\mathcal{C},post} \end{pmatrix}$$

- Matrix completion (MC) methods attempt to find a lower-rank representation of $\mathbf{Y}$, which we call $\mathbf{L}$, that makes predictions of missing values in $\mathbf{Y}$
- Athey et al. (2021) generalize Xu (2017) with different ways of constructing $\mathbf{L}$
- Plus, missingness can be arbitrary $\rightarrow$ accommodate reversible treatments (note: strict exogeneity)

Matrix Completion Methods

- Mathematically,

$$Y_{it} = L_{it} + \alpha_i + \xi_t + X_{it}'\beta + \varepsilon_{it}$$

  in which L_{it} is an element of $\mathbf{L}$, an $(N \times T)$ matrix

- We need regularization on $\mathbf{L}$ because of too many parameters:

$$\min_{\mathbf{L}} \frac{1}{\#Controls} \sum_{D_{it}=0} (Y_{it} - L_{it})^2 + \lambda_L \|\mathbf{L}\|_*$$

- The nuclear norm $\|.\|_*$ generally leads to a low-rank solution for $\mathbf{L}$

$$\|\mathbf{L}\|_* = \sum_{i=1}^{\min(N,T)} \sigma_i(\mathbf{L})$$

  in which $\sigma_i(\mathbf{L})$ represents the i'th singular values of $\mathbf{L}$

IFEct vs. MC

- Singular value decomposition of L

$$\mathbf{L}_{N\times T} = \mathbf{S}_{N\times N}\mathbf{\Sigma}_{N\times T}\mathbf{R}_{T\times T}$$

- Difference in how $\mathbf{\Sigma}_{N\times T}$ is regularized

IFE

best subset

$$\begin{pmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

MC

nuclear norm

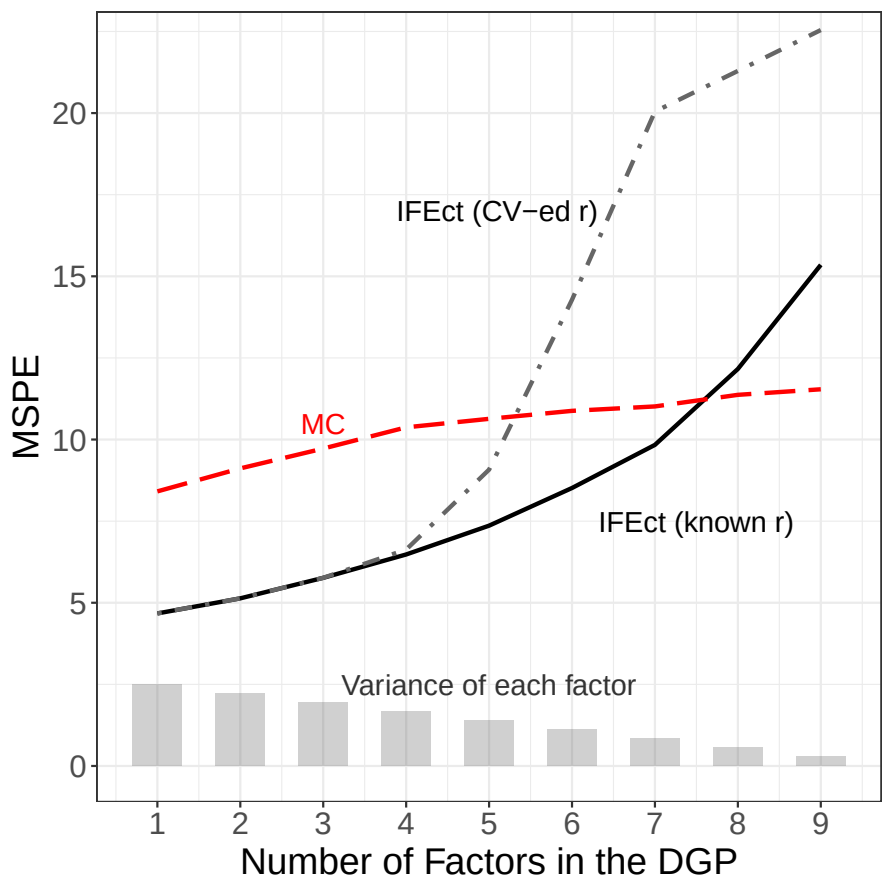
$$\begin{pmatrix} |\sigma_1 - \lambda_L|_+ & 0 & 0 & \cdots & 0 \\ 0 & |\sigma_2 - \lambda_L|_+ & 0 & \cdots & 0 \\ 0 & 0 & |\sigma_3 - \lambda_L|_+ & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \\ & & & & 0 \\ & & & & |\sigma_T - \lambda_L|_+ \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

in which $|a|_+ = \max(a, 0)$

IFEct vs. MC

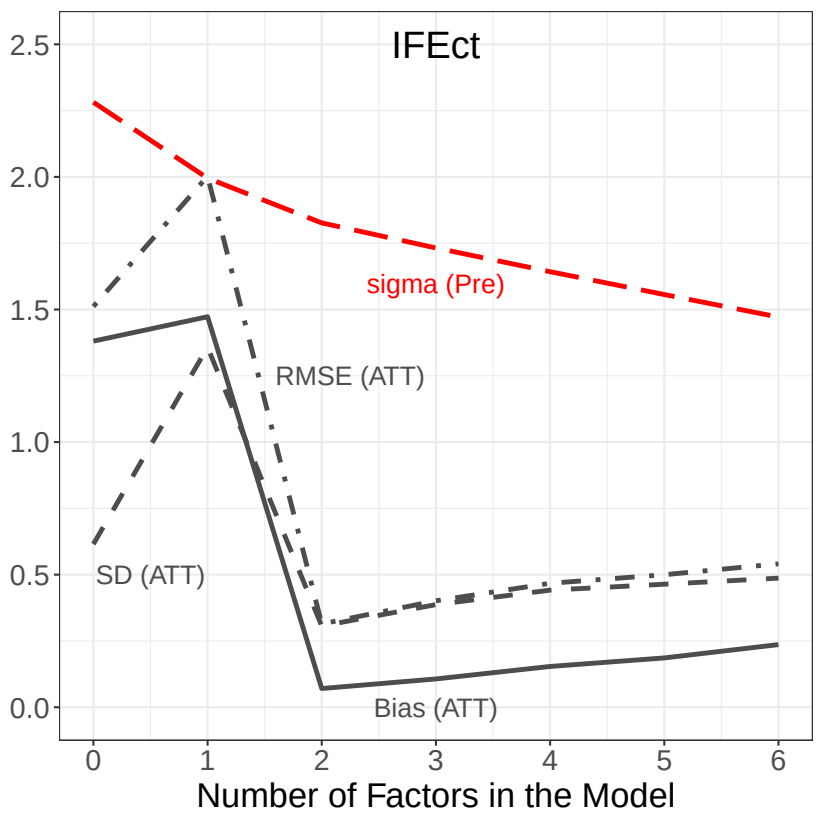
The pros and cons of IFEct and MC:

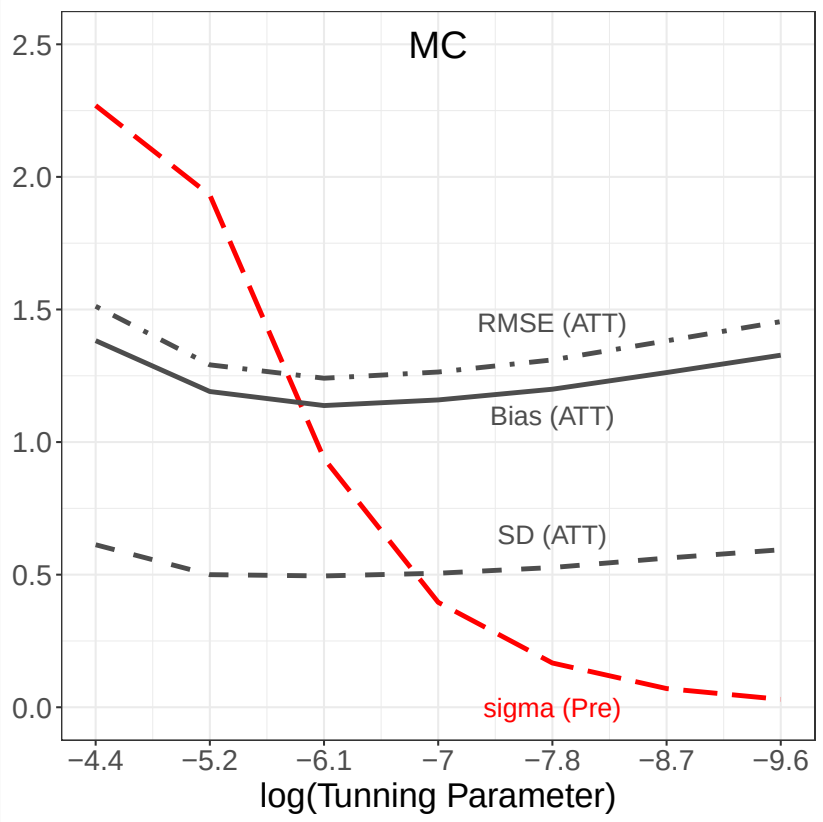
- IFEct works better with a small number of strong factors
- MC works better with a large number of weak factors



Over-fitting of IFEct and MC

When the true DGP is a two-factor IFE model:





Inferential Methods

- Non-parametric block bootstrap
  - sample with replacement across units
  - valid when N is large, $\frac{N_{tr}}{N}$ is fixed

- A permutation-based test for Sharp Nulls (Chernozhukov et al 2019)
  - e.g. $Y_{it}(1) = Y_{it}(0), \forall i \in \mathcal{T}, t > T_{0i}$
  - randomization over time (by blocks) instead of across units
  - valid if T is large, errors are stationary weekly dependent, and estimators are consistent or stable
  - exact if errors are i.i.d.

Diagnostic Tests

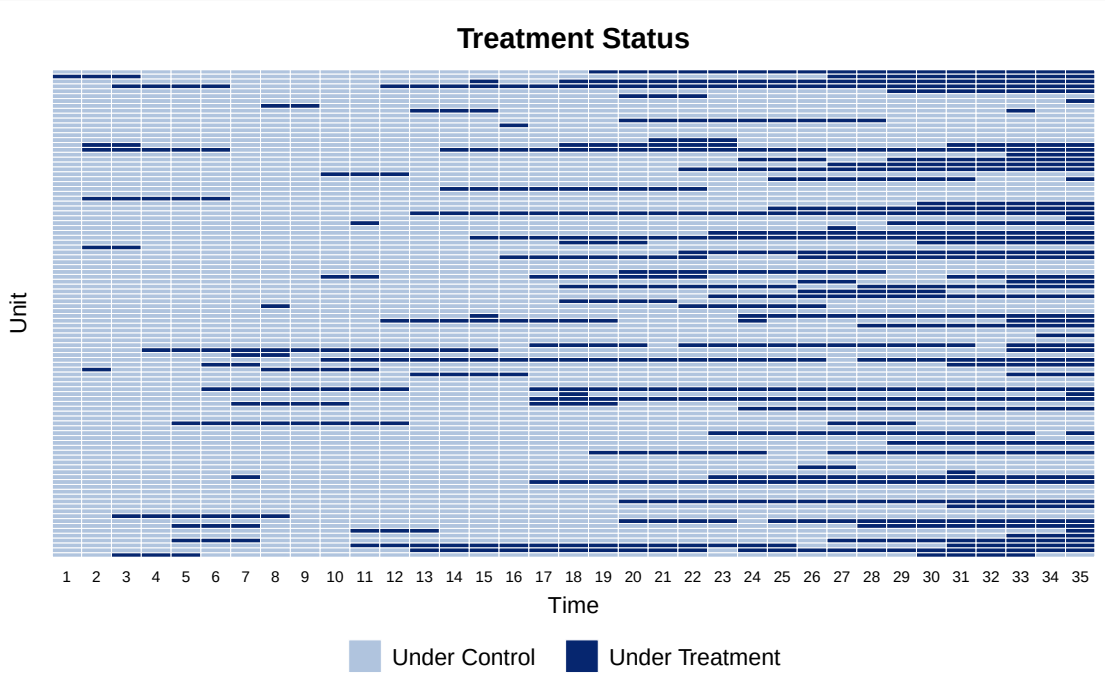
A Simulated Example

Data Generating Process:

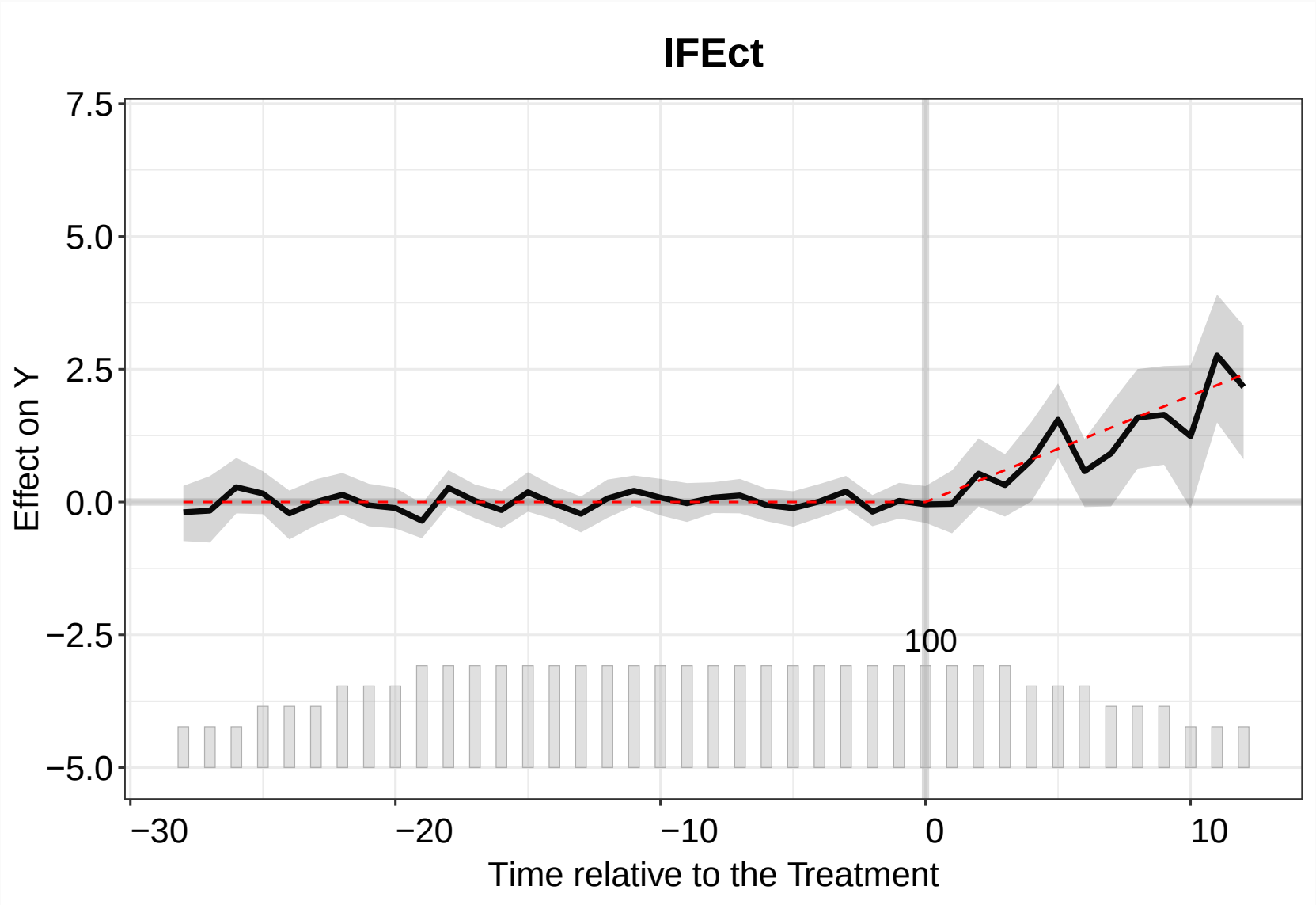
- $T = 35,\ N = 200$
- **Outcome model**: a linear interactive fixed effect model with two factors: one drift process and one white noise.

$$Y_{it} = \tau_{it} D_{it} + 5 + 1 \cdot X_{it,1} + 3 \cdot X_{it,2} + \lambda_{i1} \cdot f_{1t} + \lambda_{i2} \cdot f_{2t} + \alpha_i + \xi_t + \varepsilon_{it}$$

- **Treatment assignment**: timing of the treatment correlated with additive and interactive fixed effect.
- **Treatment effects**: $\tau_{i,t>T_{0i}} = 0.2(t - T_{0i}) + e_{it}$

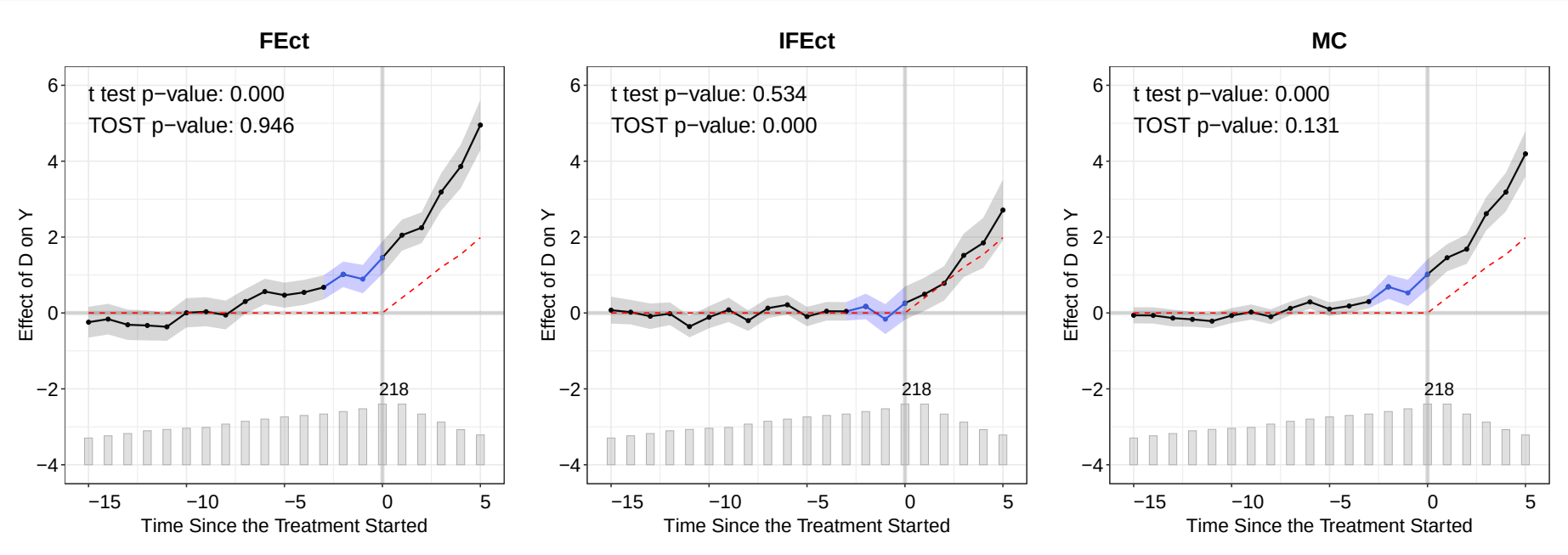


Dynamic Treatment Effects



Placebo Test

- Drop S periods before the treatment's onset, and estimate the average treatment effect in these periods.
- Test whether the average effect is significant
- Robust to model misspecification



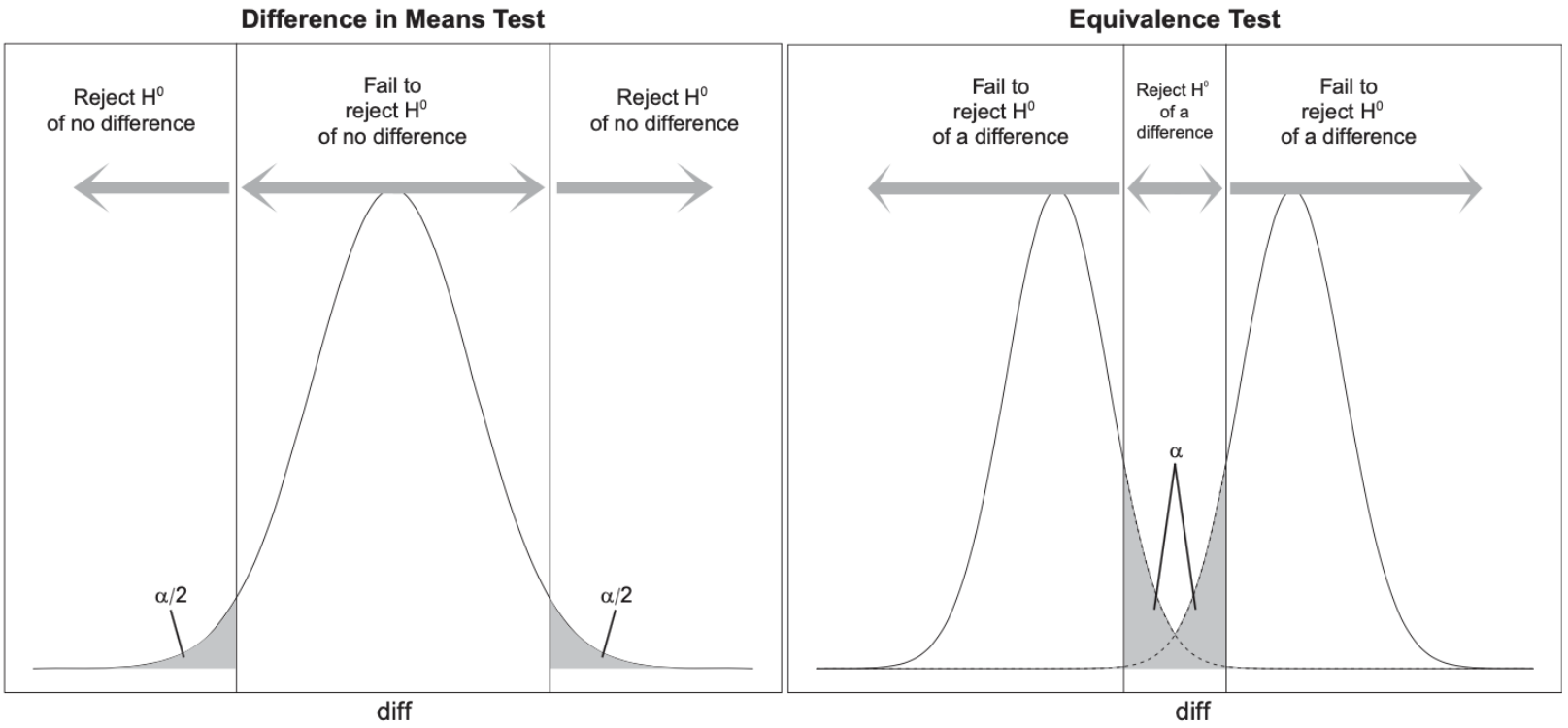
Equivalence Test

- Extension to Hartman and Hidalgo (2018) in a TSCS setting
- H0: $|ATT^p| > \theta$ vs. H1: $|ATT_t| \leq \theta$
- We calculate the maximal possible θ and compare it with pre-specified threshold: $0.36 * sd(\tilde{Y}_{it}|D_{it} = 0)$
- It has more power when the sample size grows larger, and is more likely to reject the Null (hence, equivalence holds) when a confounder is trivial

- **Drawback**: setting the threshold requires user discretion

Equivalence Test

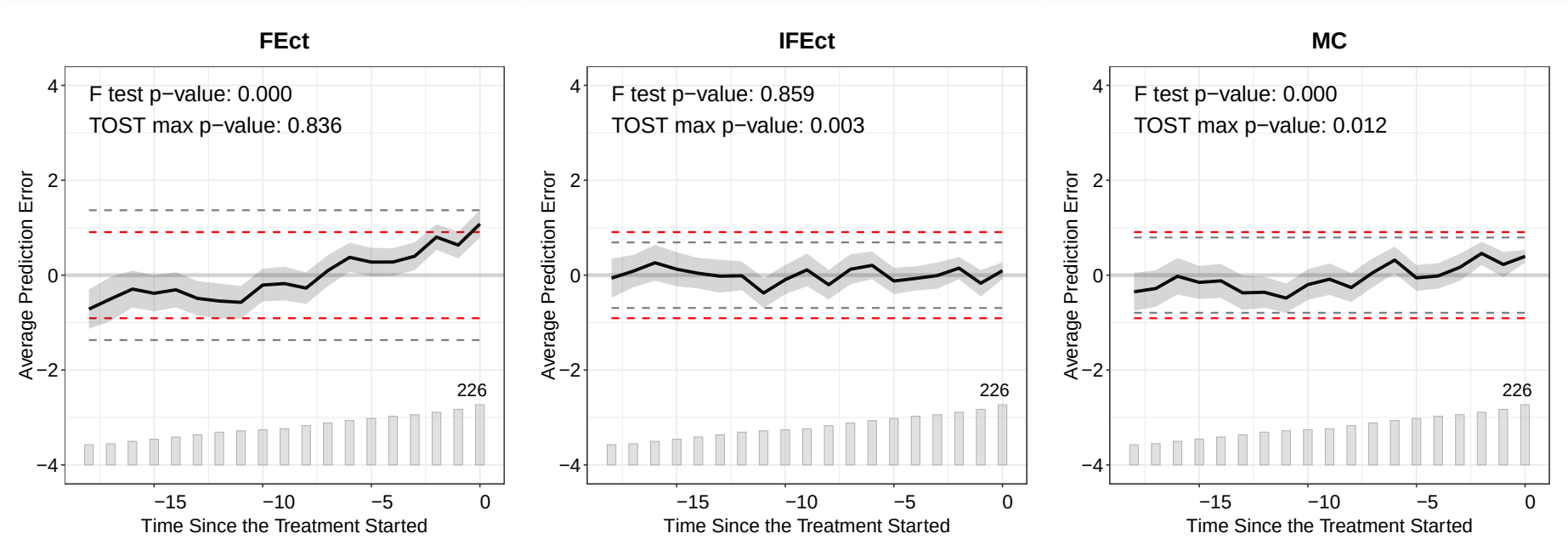
From Hartman and Hidalgo (2018)



Note: The left panel depicts the logic of tests of difference under the null hypothesis of no difference. The right panel depicts the logic of one type of equivalence test—the two one-sided t-test (TOST)—under the null hypothesis of difference.

Extension 1. Test for No Pre-Trend

- Drop one pre-treatment period at a time (leave-one-out) and collect the residual averages
- Test whether the residual averages equal to 0 collectively
- Use both the difference-in-means approach and the equivalence approach
- Can be too lenient for IFE or MC

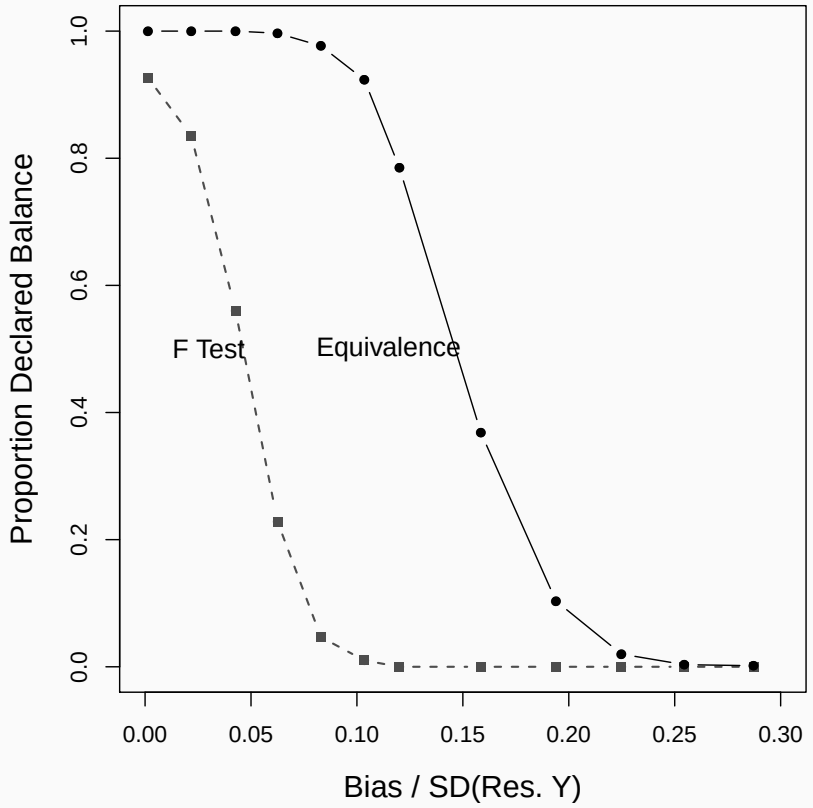


Why Not a Wald test?

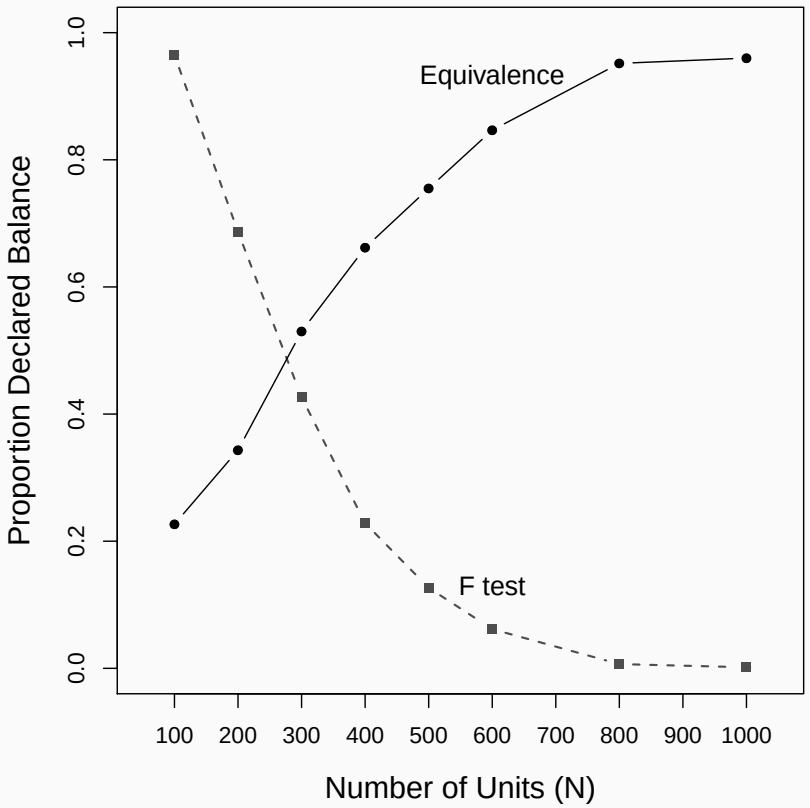
- A simpler test: conduct an Wald (F) test on pre-treatment residual averages
- Power issue
- When there exists a small confounders which induces a neglectable bias compared with the ATT, a Wald test will almost always reject the Null (that equivalence holds) when there are enough data
- The equivalence test approach avoids this problem

Why Not a Wald Test?

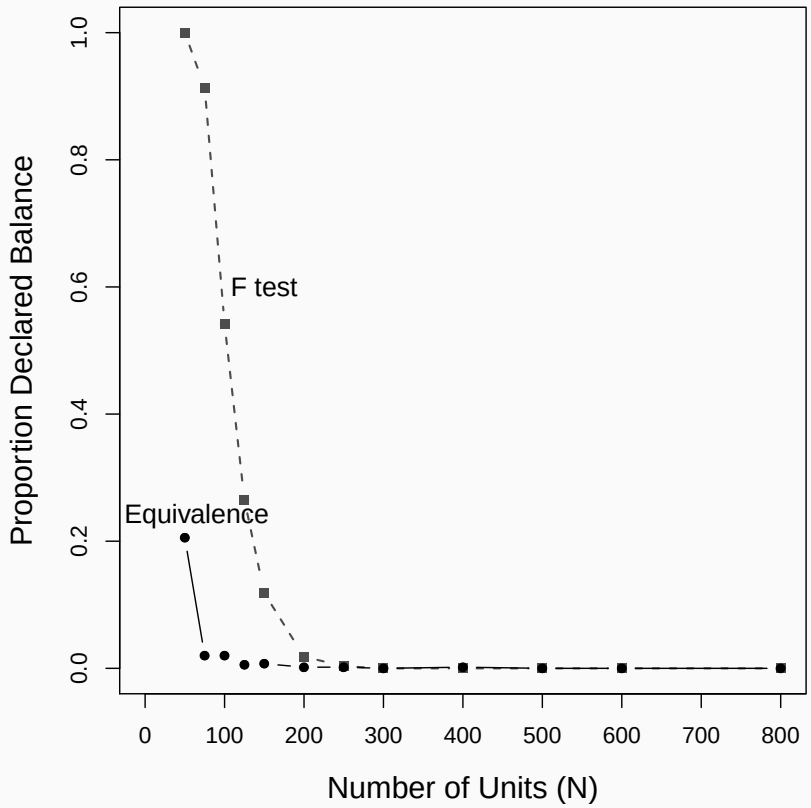
- There exist a time-varying confounder
- We vary its influence on the bias in the ATT



Why Not a Wald Test?



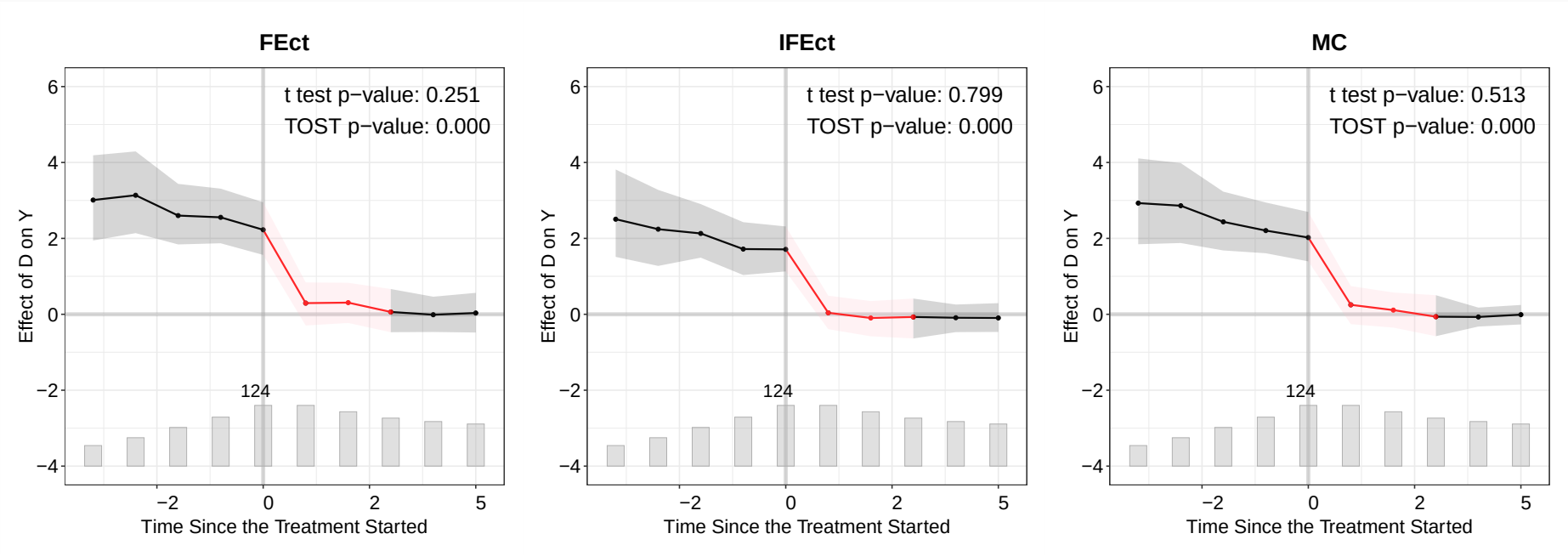
$Bias = 0.08SD(Y)$



$Bias = 0.28SD(Y)$

Extension 2. Test for No Carryover Effects

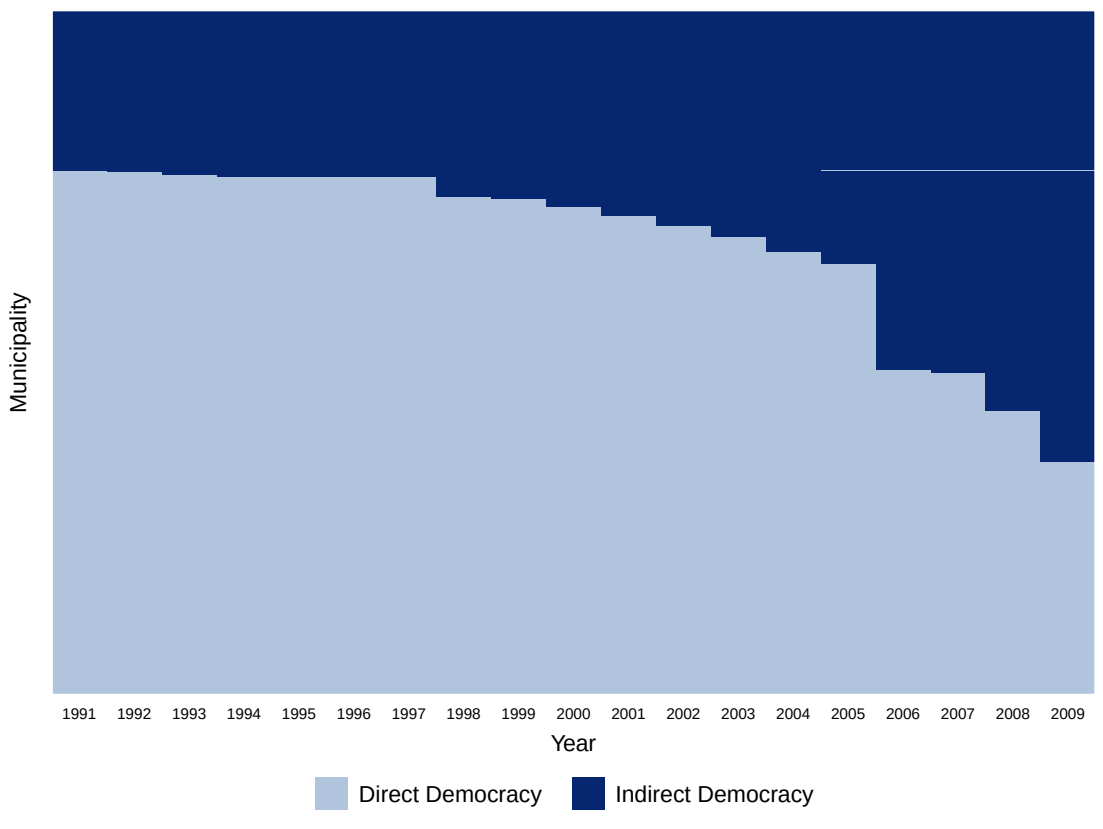
- Drop several periods after the treatment ends
- Test whether the average carryover effect is significant
- Can use both the difference-in-means approach and the equivalence approach



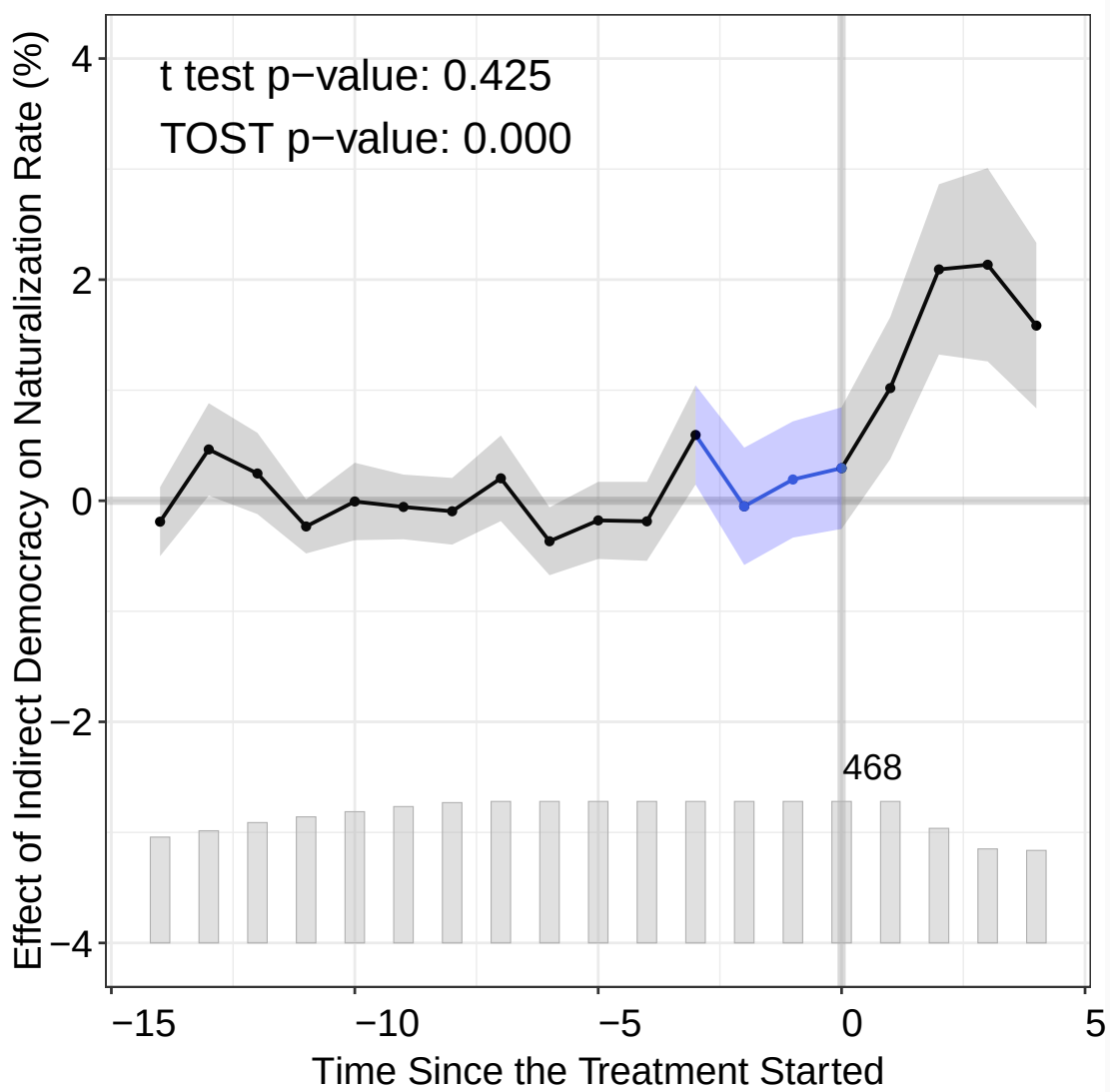
Hainmueller & Hangatner (2019)

Does indirect democracy benefit immigrant minorities?

- Unit of analysis: 1400 Swiss municipalities from 1991-2009
- Treatment: Indirect (vs. direct) democracy
- Outcome: Naturalization rate



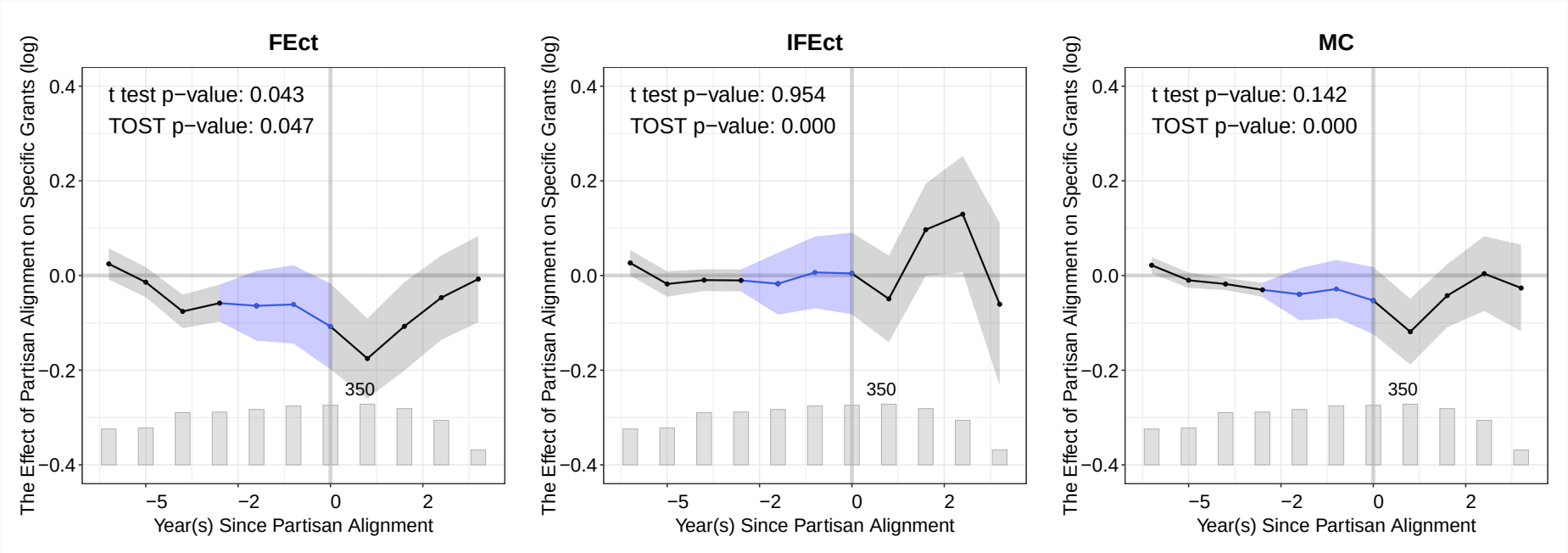
Hainmueller & Hangatner (2019)



Fouirnaies and Mutlu-Eren (2015)

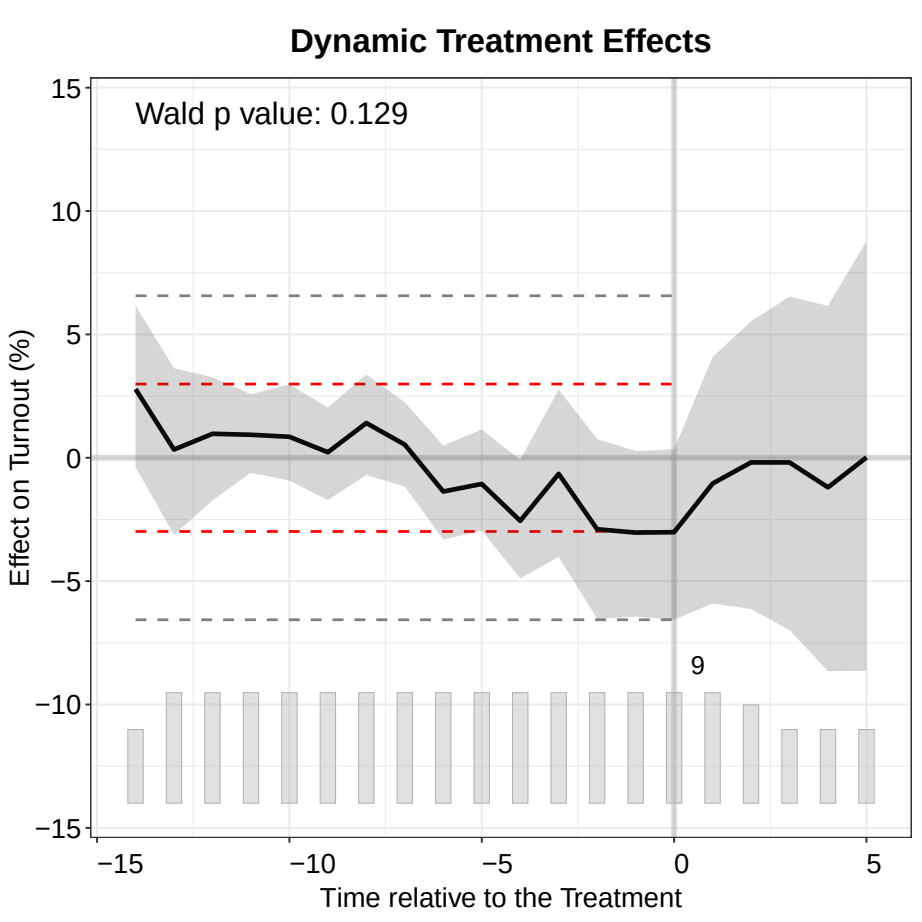
Does partisan alignment brings special grant in UK?

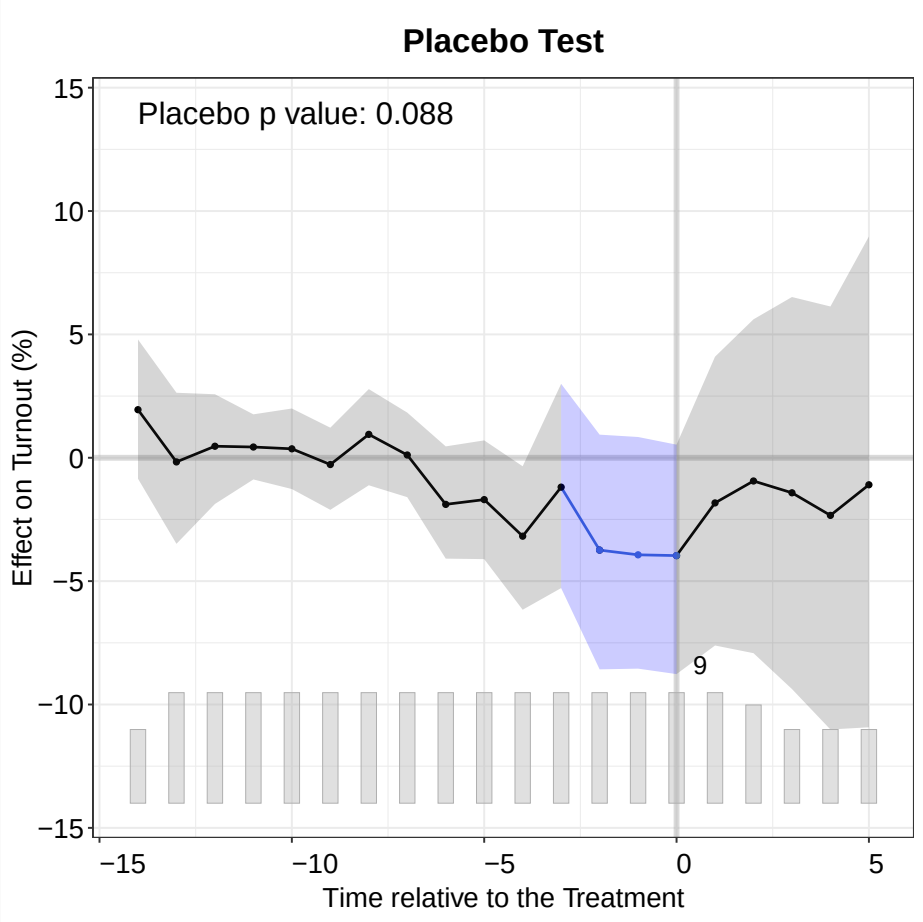
- Unit of analysis: 466 local councils from 1992 to 2012
- Treatment: Partisan alignment between locality and central government
- Outcome: Amount of special grant



EDR on Turnout

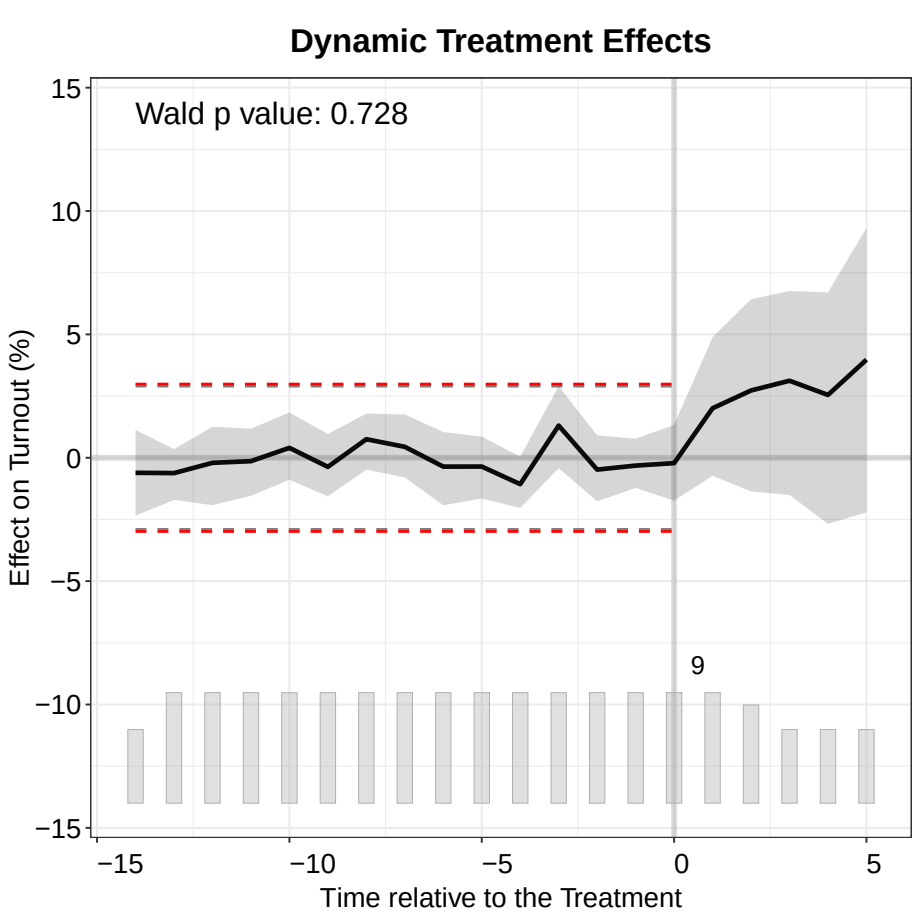
FEct

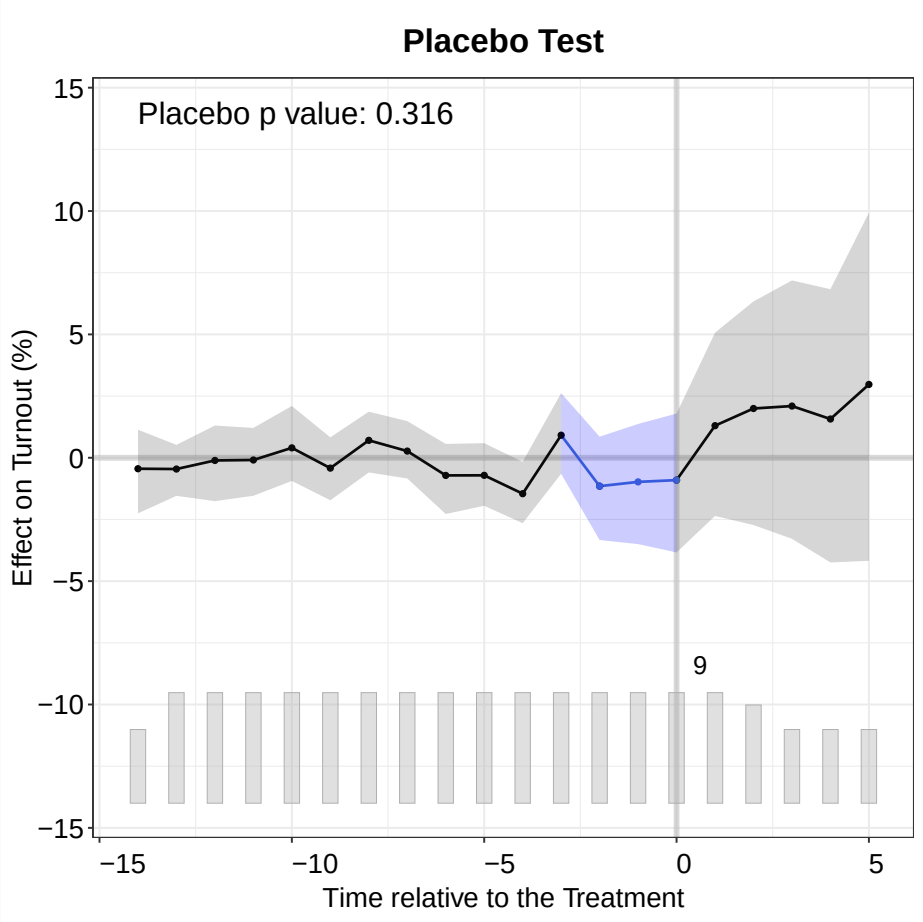




EDR on Turnout

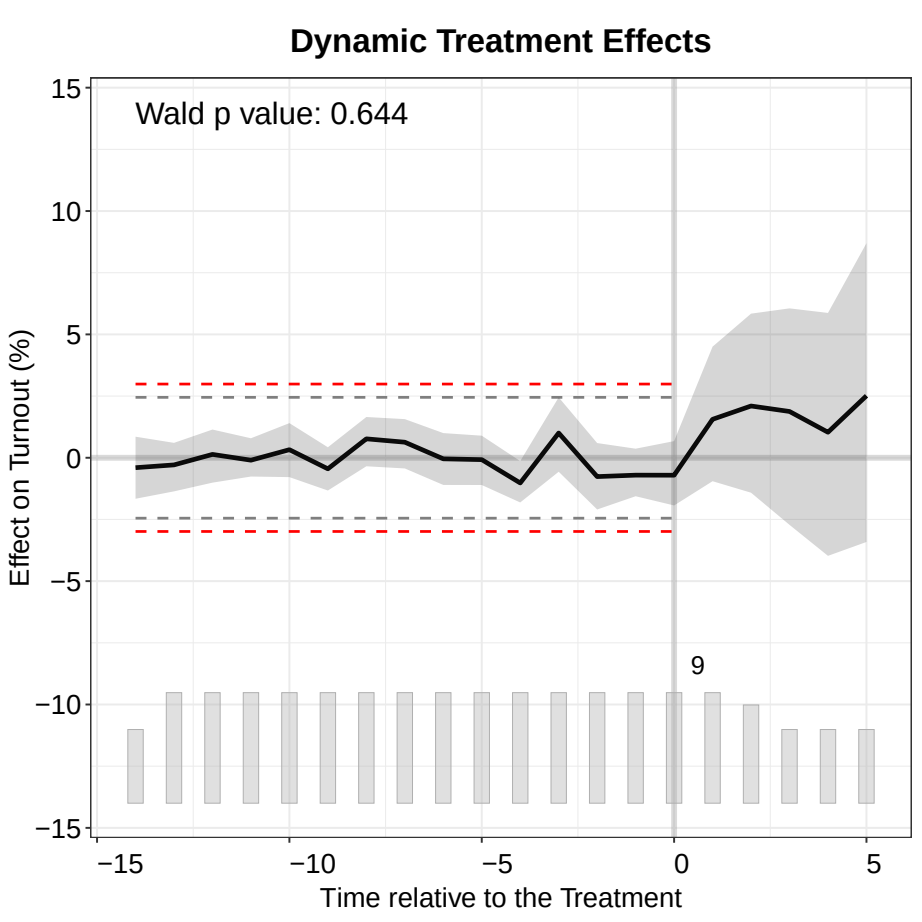
IFEct

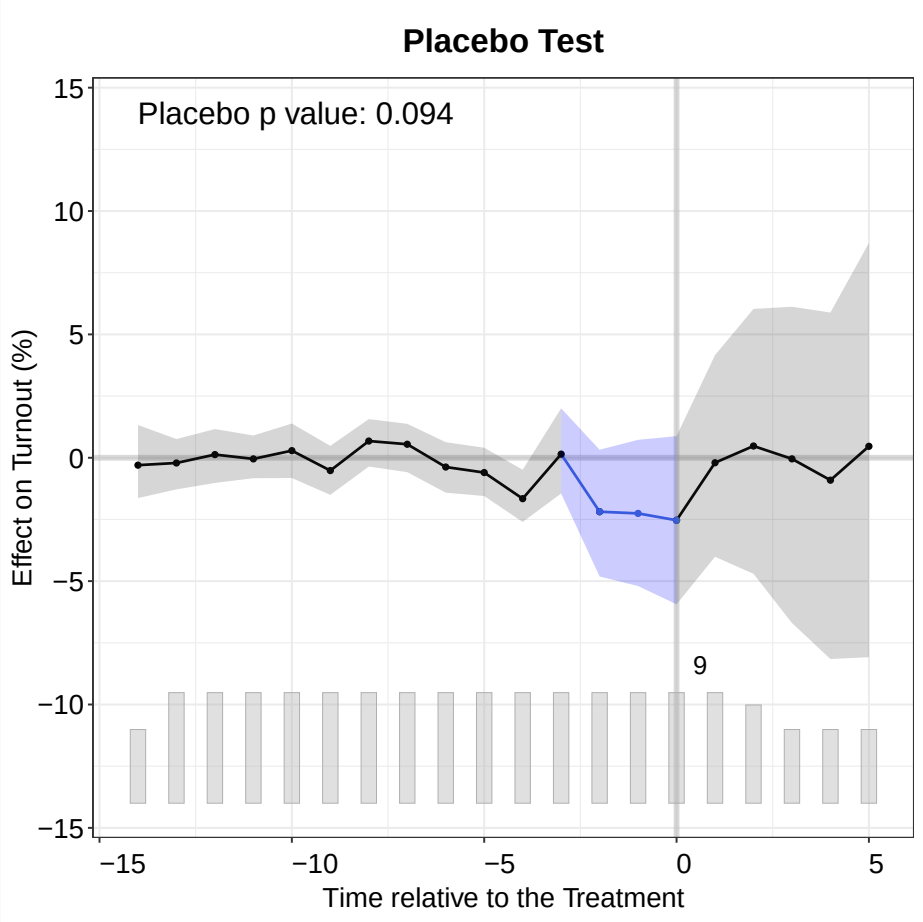




EDR on Turnout

MC





Bayesian Multi-Factor Models

Why A Bayesian Approach

- *Challenge:* Valid inference for individual or average treatment effects in comparative case studies remains difficult

- *Solution:* A fully Bayesian approach — theory, estimation, prediction, and inference

- Key features of Pang, Liu & Xu (2021)

  1. Bayesian causal inference: Treated counterfactuals = Missing data (under MNAR)
     $\Rightarrow$ Inference based on posterior distributions of imputed counterfactuals

  2. Semi-parametric:
     A multi-factor state-space model + Stochastic model specification search

  3. Dimension reduction:
     Dense modeling (factor analysis) + Sparse modeling (Bayesian Lasso)

- Other work:
  Carlson (2018); Samartsidis (2020); Kim, Lee & Gupta (2020); Feller et al. (2021)

Basic Idea: Bayesian Prediction

Like counterfactual estimators, Bayesian causal inference takes three steps:

1. **Bayesian Model Search**
   using MCMC to obtain a model for the non-treated potential outcome:
   $f(y_{it}(0)|X_{it}, \theta_{it})$

2. **Bayesian Prediction**
   for treated counterfactuals based on posterior draws of the parameters and observed covariates

3. **Summarization and Averaging**
   based on observed treated outcomes and the posterior draws of their counterfactuals

1-2 *SUTVA & Staggered Adoption*

3 *Latent Ignorablity* (under MNAR):

- Treatment assignment (missingness) is ignorable when we condition on observed data and an unobserved latent variable $\mathbf{U}$:

$$\Pr(D_i|\mathbf{X}_i, \mathbf{Y}_i(\mathbf{0}), \mathbf{U}_i) = \Pr(D_i|\mathbf{X}_i, \mathbf{Y}_i(\mathbf{0})^{obs}, \mathbf{U}_i) = \Pr(D_i|\mathbf{X}_i, \mathbf{U}_i)$$

4 *Feasible Data Extraction*:

- $\mathbf{U}_i$ can be learned from observed data $(\mathbf{X}, \mathbf{Y}(\mathbf{0})^{obs})$
- $\mathbf{U}_{(n\times T)}$ can be approximated by two lower-rank matrices, $\mathbf{U} = \mathbf{\Gamma}'\mathbf{F}$ in which $\mathbf{F} = (\mathbf{f}_1, ..., \mathbf{f}_T)$

◦ Then we have:

$$\Pr(\mathbf{Y}(\mathbf{0})^m|\mathcal{D}, \mathbf{Y}(\mathbf{0})^o, \mathbf{X}, \mathbf{U}) \propto \Pr(\mathbf{Y}(\mathbf{0}), \mathbf{X}, \mathbf{U})$$

$$\propto \int \underbrace{\left(\prod f(y(0)^m_{it}|\mathbf{X}_{it}, \mathbf{U}_i, \boldsymbol{\theta})\right)}_{\text{posterior predictive distribution}} \times \underbrace{\left(\prod f(y(0)^o_{it}|\mathbf{X}_{it}, \mathbf{U}_i, \boldsymbol{\theta})\right)}_{\text{likelihood}} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

A Hierarchical Dynamic Factor Model

Assume the following DGP for $y_{it}(0)$

$$
\begin{aligned}
y_{it}(0) &= \mathbf{x}_{it}\boldsymbol{\beta}_{it} + \boldsymbol{\gamma}_i'\mathbf{f}_t + \epsilon_{it}, \\
\boldsymbol{\beta}_{it} &= \boldsymbol{\beta} + \boldsymbol{\alpha}_i + \boldsymbol{\xi}_t \\
\boldsymbol{\xi}_t &= \Phi_\xi \boldsymbol{\xi}_{t-1} + \mathbf{e}_t, \\
\mathbf{f}_t &= \Phi_f \mathbf{f}_{t-1} + \boldsymbol{\nu}_t.
\end{aligned}
$$

- $\mathbf{x}_{it}$: observed covariates, and could be time-invariant or unit-invariant
- $\boldsymbol{\beta}_{it}$: unit-time-specific (individual) relationships between covariates and outcome
- $\boldsymbol{\gamma}_i'\mathbf{f}_t$: the latent multifactor term (a.k.a., interactive fixed effects)

Correlation and Heteroskedasticity

- The reduced and matrix (estimation) form of the model:

$$y_{it}(0) = \mathbf{X}_{it}\boldsymbol{\beta} + \mathbf{Z}_{it}\boldsymbol{\alpha}_i + \mathbf{A}_{it}\boldsymbol{\xi}_t + \mathbf{f}_t\boldsymbol{\gamma}_i + \epsilon_{it}$$
$$\boldsymbol{\xi}_t = \Phi_\xi \boldsymbol{\xi}_{t-1} + \mathbf{e}_t,\ \mathbf{f}_t = \Phi_f \mathbf{f}_{t-1} + \boldsymbol{\nu}_t$$

  where $\mathbf{Z}_{it}$, $\mathbf{A}_{it} \subseteq \mathbf{X}_{it}$ (*not all covariates have to have varying coefficients*)

- Correlated and Heteroskedastic Errors (variance-covariance matrix):

$$\boldsymbol{\Omega}_{\mathbf{y}_i} = \mathbf{Z}_i'\boldsymbol{\Sigma}_{\alpha_i}\mathbf{Z}_i + \mathbf{A}_i'\boldsymbol{\Sigma}_\xi\mathbf{A}_i + (\mathbf{F}\boldsymbol{\gamma}_i)'\mathbf{F}\boldsymbol{\gamma}_i + \sigma_\epsilon^2\mathbf{I}$$

- Prior Distributions of all parameters and hyperparameters

Bayesian Stochastic Model Specification Search

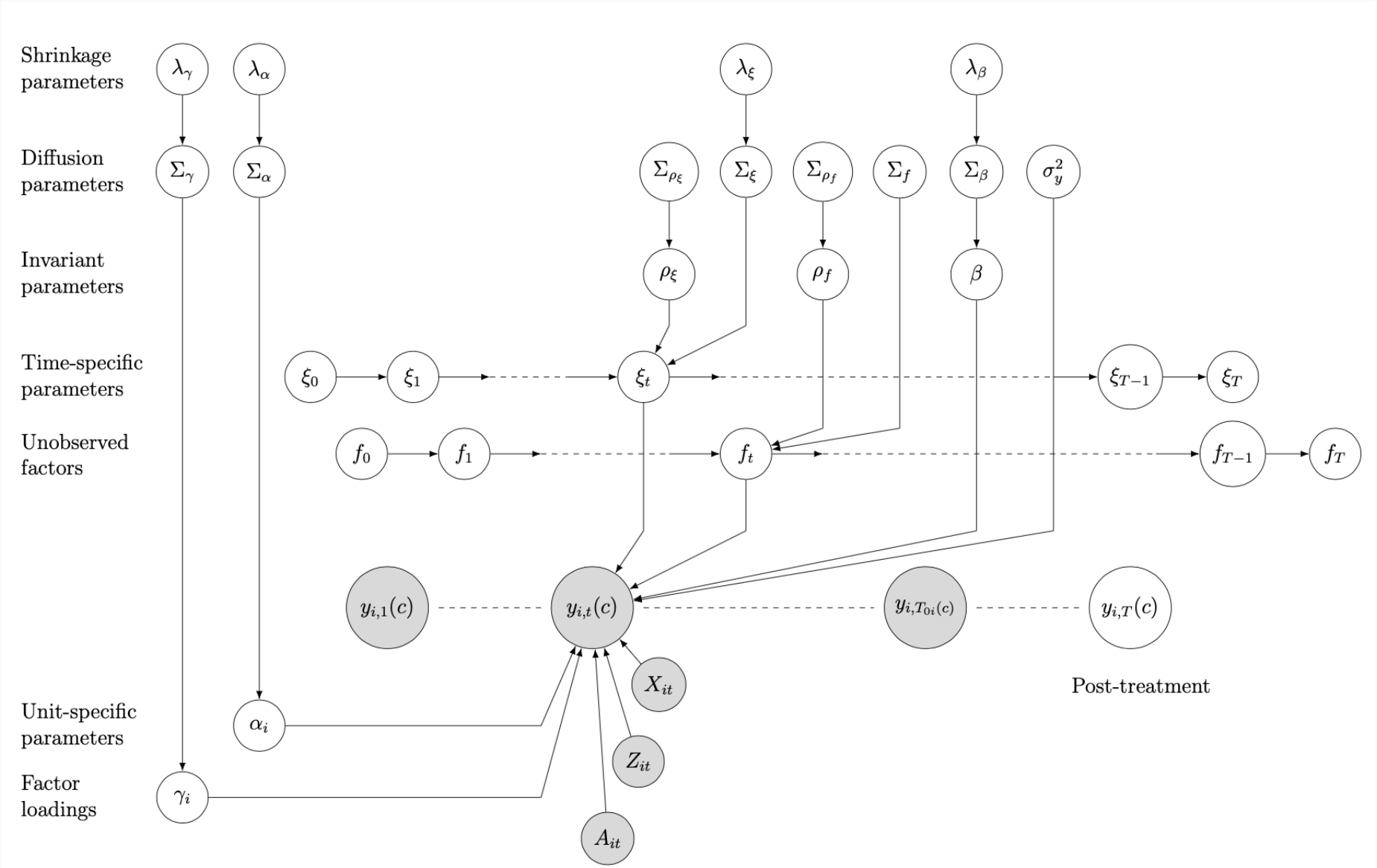
- Dense Modeling + Sparse Modeling
  - *Dense Modeling*: factor analysis for feature extraction and variable aggregation
  - *Sparse Modeling*: Bayesian shrinkage on factors and varying parameters
- Bayesian Lasso and Lasso-like hierarchical shrinkage
  - Lasso: $\min_{\boldsymbol{\beta}}(\mathbf{y}-\boldsymbol{\beta}\mathbf{X})'(\mathbf{y}-\boldsymbol{\beta}\mathbf{X})+\lambda\sum_{j=1}^{J}|\beta_j|$
  - Bayesian Lasso: Bayesian posterior mode with independent Laplace priors to $\boldsymbol{\beta}$ is equivalent to Lasso, and the Laplacian is a Gaussion Scale mixture
  - For variable selection on the observed confounders $\mathbf{X}$, Bayesian Lasso is directly applied on $\boldsymbol{\beta}$

$$\beta_k|\tau^2_{\beta_k} \sim \mathcal{N}(0,\tau^2_{\beta_k}), \forall 1 \leq k \leq p_1$$

$$\tau^2_{\beta_k}|\lambda_\beta \sim Exp(\frac{\lambda^2_\beta}{2})$$

$$\lambda^2_\beta \sim \mathcal{G}(a_1, a_2)$$

Model Setup


Shrinkage parameters
Diffusion parameters
Invariant parameters
Time-specific parameters
Unobserved factors
Unit-specific parameters
Factor loadings
λ_γ
λ_α
λ_ξ
λ_β
Σ_γ
Σ_α
Σ_{ρ_ξ}
Σ_ξ
Σ_{ρ_f}
Σ_f
Σ_β
σ_y^2
ρ_ξ
ρ_f
β
ξ_0
ξ_1
ξ_t
ξ_{T-1}
ξ_T
f_0
f_1
f_t
f_{T-1}
f_T
$y_{i,1}(c)$
$y_{i,t}(c)$
$y_{i,T_{0i}(c)}$
$y_{i,T}(c)$
Post-treatment
X_{it}
α_i
Z_{it}
γ_i
A_{it}

Bayesian Stochastic Model Searching

- Re-parameterized model:

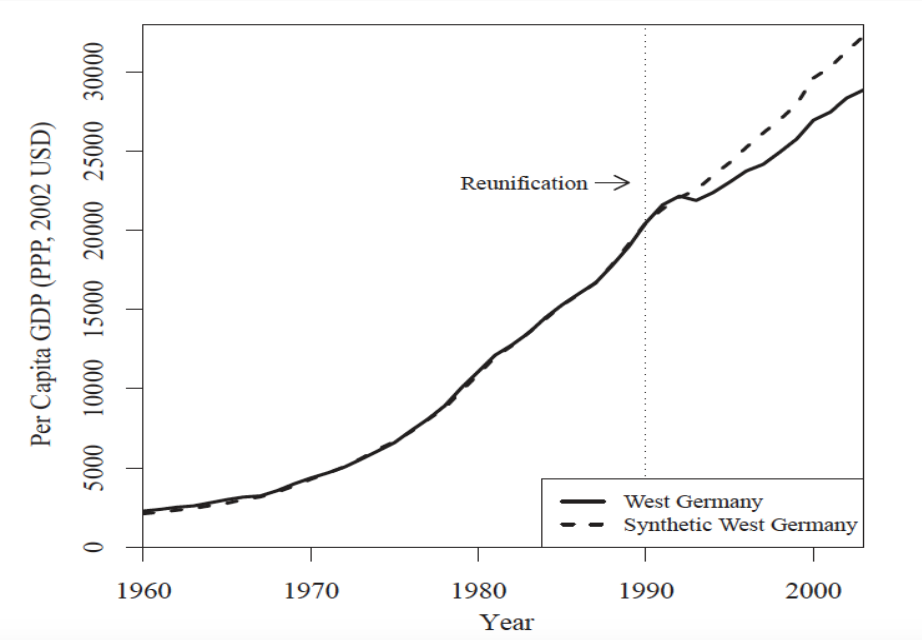
$$y_{it} = \mathbf{X}_{it}'\boldsymbol{\beta} + \mathbf{Z}_{it}'(\boldsymbol{\omega}_{\boldsymbol{\alpha}} \cdot \tilde{\boldsymbol{\alpha}}_i) + \mathbf{A}_{it}'(\boldsymbol{\omega}_{\boldsymbol{\xi}} \cdot \tilde{\boldsymbol{\xi}}_t) + (\boldsymbol{\omega}_{\boldsymbol{\gamma}} \cdot \tilde{\boldsymbol{\gamma}}_i)'\mathbf{f}_t + \varepsilon_{it},$$

- A summary of stochastic model searching:
  1. $\boldsymbol{\beta}$: when $\beta_{j_1} \approx 0$, then covariate X_{it,j_1} does not have an intercept.
  2. $\boldsymbol{\alpha}_i$: when $\omega_{\alpha_{j_2}} \approx 0$, then covariate Z_{it,j_2} does not have a random effect at unit level.
  3. $\boldsymbol{\xi}_t$: when $\omega_{\xi_{j_3}} \approx 0$, then covariate A_{it,j_3} does not have a time-varying effect.
  4. $\mathbf{f}_t$: when $\omega_{\gamma_{j_4}} \approx 0$, then the j_4^{th} unobserved factor will not be included.

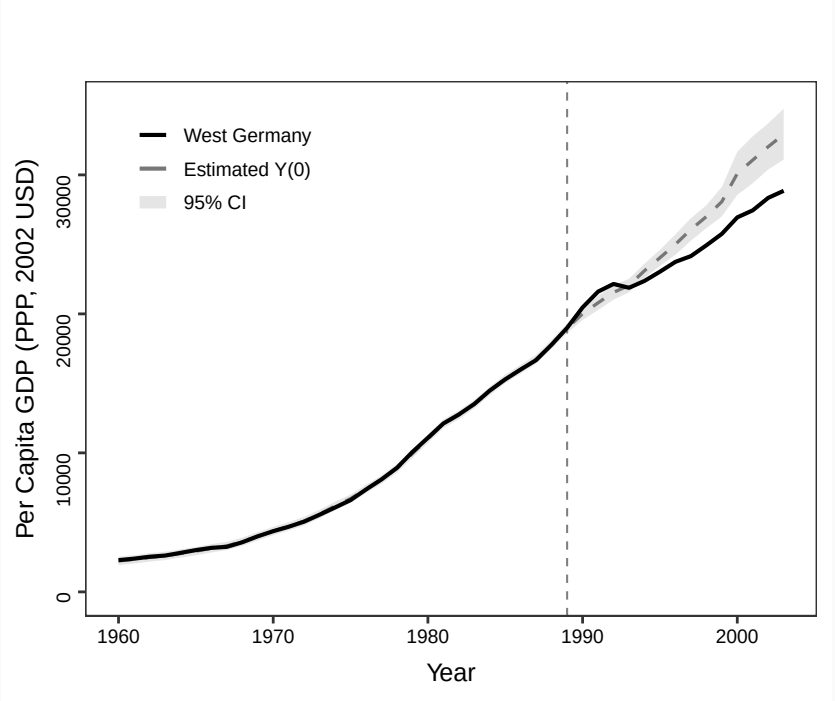
- In each MCMC iteration, a model is sampled with some covariates (and factors) included in the model, but the others are virtually zeroed out.

ADH (2015): The Effect of German Re-Unification

Figure 1: Predicted Counterfactuals



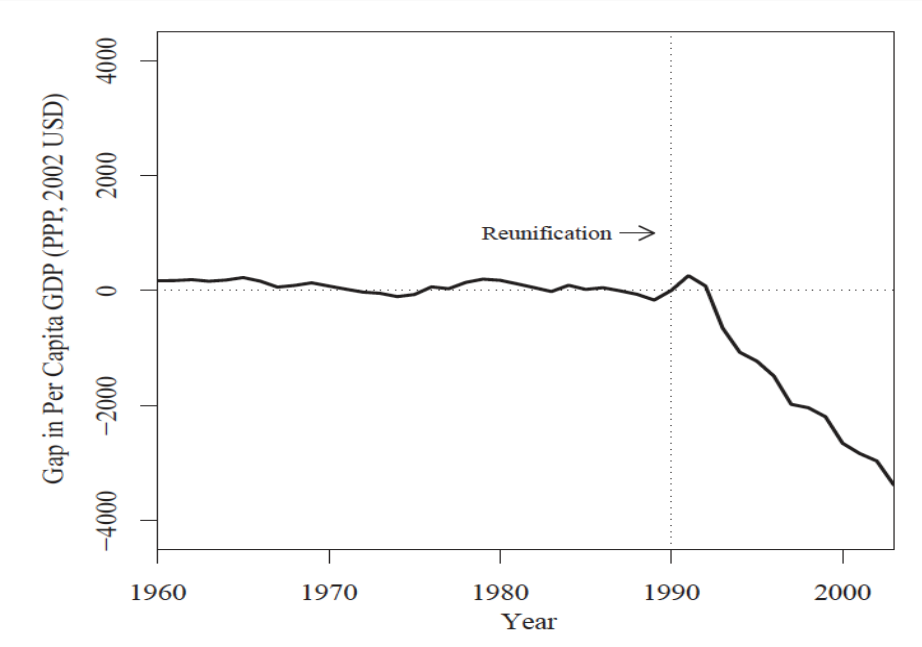
SCM



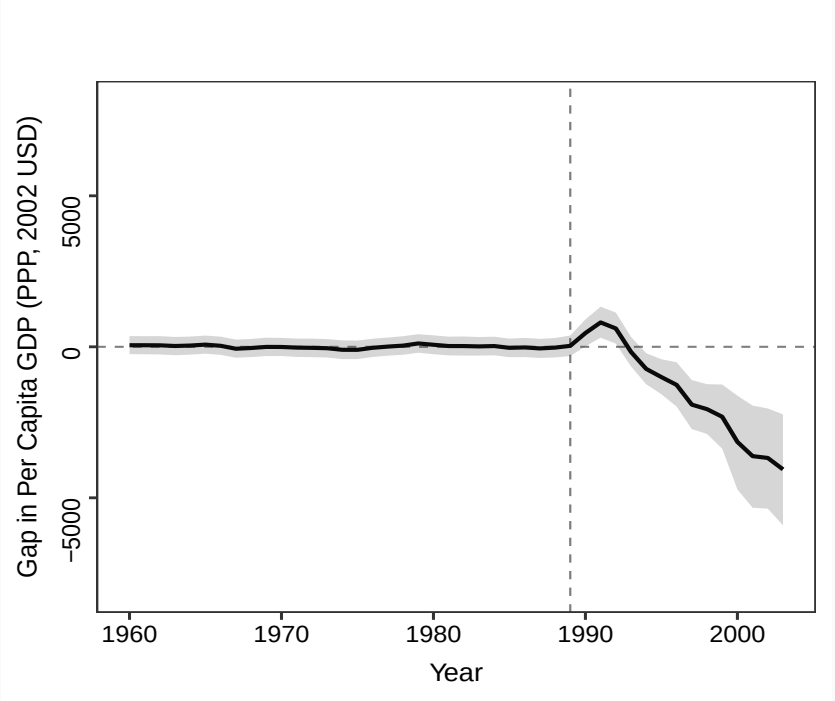
Bayesian DM-LFM

ADH (2015): The Effect of German Re-Unification

Figure 2: Estimated Treatment Effect



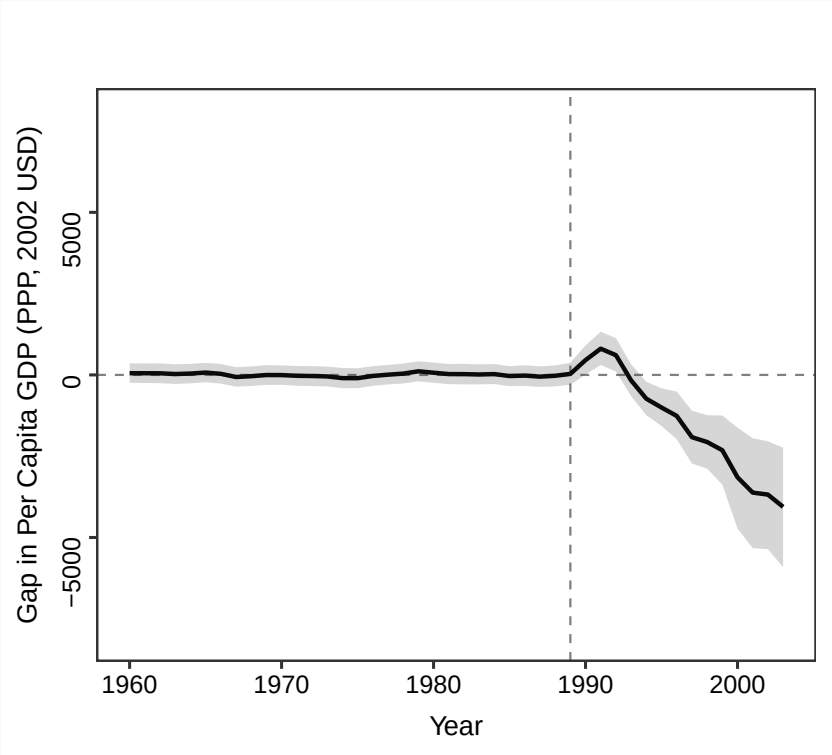
SCM



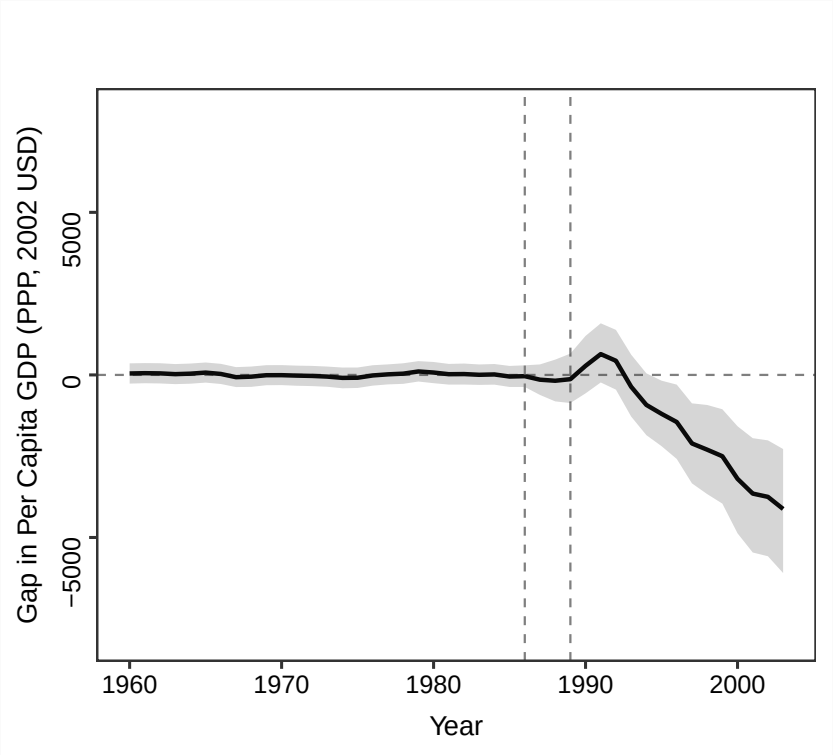
Bayesian DM-LFM

ADH (2015): A Placebo Test

w/o Placebo Periods



w/ Placebo Periods



Summary

- Compared with existing methods, Bayesian methods provide more interpretable uncertainty estimates (Bayesian credibility intervals)
- Bayesian model search avoids arbitrary choices of model specifications and efficient model searching
- Limitations
  - Model dependency
  - Requires strict exogeneity (albeit conditional on latent factors)
  - Requires stationarity of time series data
- Future work: pay more attention to treatment assignment mechanisms; other modeling choices: e.g., Gaussian process

References

- Xu, Yiqing (2017). "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models" *Political Analysis*, Vol. 25, Iss. 1, January 2017, pp. 57-76.
- Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. 2021. "Matrix Completion Methods for Causal Panel Data Models." *Journal of the American Statistical Association*, forthcoming. https://doi.org/10.3386/w25132.
- Sanford, Luke (2019) "Geospatial Synthetic Controls for Agricultural Impact Evaluation." Working Paper.
- Liu, Licheng, Ye Wang, Yiqing Xu (2021). "A Practical Guide to Counterfacutal Estimators for Causal Inference with Time-Series Cross-Sectional Data." Working Paper, Stanford University.
- Xun, Pang, Licheng Liu, and Yiqing Xu (2021). "A Bayesian Alternative to Synthetic Control for Comparative Case Studies". *Political Analysis*, forthcoming.