Synthetic Control Methods through Predictive Synthesis

Masahiro Kato (The University of Tokyo)

Coauthors: Akira Fukuda, Kosaku Takanashi,

Kenichiro McAlinn, Akari Ohda, Masaaki Imaizumi

Paper 1: Synthetic Control Methods by Density Matching under Implicit Endogeneity (https://arxiv.org/abs/2307.11127)

Paper 2: Bayesian Predictive Synthetic Control Methods
(https://drive.google.com/file/d/1veWTQTuWTx2gAMyh7VSZnenxsVqs1nla/view)

Speaker Deck: https://speakerdeck.com/masakat0/synthetic-control-methods-through-predictive-synthesis?slide=25

Synthetic Control Methods

➢ Synthetic Control Methods (SCMs; Abadie et al. 2003).

■ **Core idea.**

- There are several units. One unit among them receives a policy intervention (treated unit).
- Policy intervention affects outcomes of the treated unit.
- We cannot observe outcomes when the treated unit does not receive the policy intervention
- Estimate counterfactual outcomes of the treated unit by using a weighted sum of observed outcomes of untreated units.
- Then, using the estimated outcome, estimate the causal effect of the treated unit.

Problem Setting

- $J+1$ units, $j \in \mathcal{J} := \{0,1,2,\dots,J\}$.
  - $j = 0$: **Treated unit** (a unit affected by the policy intervention).
  - $j \in \mathcal{J}^U$: $\mathcal{J} \setminus \{0\}$: **Untreated units**.
- T Periods, $t \in \mathcal{T} := \{1,2,\dots,T\}$.
  - Intervention occurs at $t = T_0 < T$.
  - $t \in \mathcal{T}_0 := \{1,2,\dots,T_0\}$: before the intervention.
  - $t \in \mathcal{T} \setminus \mathcal{T}_0$: after the intervention ($T_1 := |\mathcal{T}_1| = T - T_0$).

Problem Setting

- **Potential outcomes** (Neyman, 1923; Rubin, 1974):
  - For each unit $j \in \mathcal{J}$ and period $t \in \mathcal{T}$, define potential outcomes $\left(Y_{j,t}^I, Y_{j,t}^N\right) \in \mathbb{R}^2$.
    - Y_t^I and Y_t^N are potential outcomes with and without interventions.
    - $\mathbb{E}_{j,t}$: expectation over Y_t^I and Y_t^N.
- **Observations**:
  - Observe one of the outcomes, $Y_{j,t} \in \mathbb{R}$, corresponding to actual intervention; that is,

$$Y_{0,t} = \begin{cases} Y_{0,t}^I & if \ \ t \in \mathcal{T}_1 \\ Y_{0,t}^N & if \ \ t \in \mathcal{T}_0 \end{cases}, \qquad Y_{j,t} = Y_{j,t}^N \quad for\ j \in \mathcal{J}^U.$$

Problem Setting

- Causal effects:

$$\tau_{0,t} := \mathbb{E}_{0,t}\left[Y_{0,t}^{I} - Y_{0,t}^{N}\right] \quad for\ t \in \mathcal{T}_1.$$

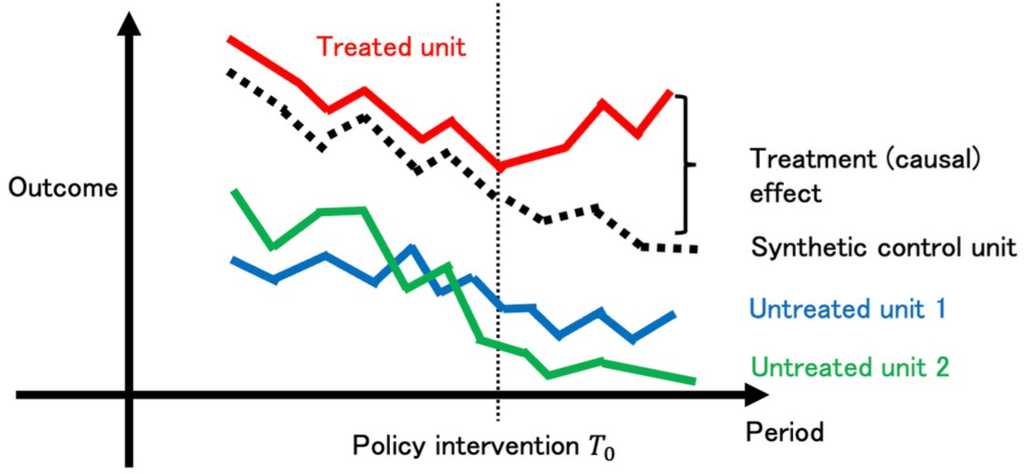
- Estimating the causal effect by predicting $Y_{0,t}^{N}$ for $t \in \mathcal{T}_1$.

- Core idea.

- Predict $Y_{0,t}^{N}$ by a weighted sum of $Y_{1,t}^{N}, \dots, Y_{J,t}^{N}$.

$$\hat{Y}_{0,t}^{N} = \sum_{j \in J^U} w_j Y_{j,t}^{N}.$$

  - $\hat{Y}_{0,t}^{N}$ is a counterfactual trend of the treated unit.
  - $\hat{Y}_{0,t}^{N}$ is called a synthetic control unit.



Contents

- Research questions mainly lie in estimation of the weights, $w_1, \ldots, w_J$.

- Paper 1: Synthetic Control Methods by Density Matching under Implicit Endogeneity.
  - Estimators in existing SCMs are not consistent (Ferman and Pinto, 2021).
  - We discuss the inconsistency problem from the viewpoint of endogeneity.
  - <u>Propose frequentist SCMs with the generalized method of moments (GMM).</u>

- Paper 2: Bayesian Predictive Synthetic Control Methods.
  - <u>Apply Bayesian predicative synthesis for SCMs.</u>
  - Flexible modeling with time-varying parameter, finite-sample analysis, and minimax optimality.

Synthetic Control Methods by Density Matching under Implicit Endogeneity

Least-Squares Estimator

- In standard SCM, we usually estimate the weights by constraint least squares.

- That is, we estimate w_j as

$$\left(\widehat{w}_j^{\mathrm{LS}}\right)_{j \in J^U} = \arg\min_{(w_j)_{j \in J^U}} \frac{1}{T} \sum_{t \in \mathcal{T}_0} \left(Y_{0,t}^N - w_j Y_{j,t}^N\right)^2 \quad \text{such that} \quad \sum_{j \in J^U} w_j = 1, \quad w_j \geq 0 \quad \forall j \in J^U.$$

- To justify the least squares (LS) estimator, we assume the linearity in the expected outcomes:

$$\mathbb{E}\left[Y_{0,t}^N\right] = \sum_{j \in J^U} w_j^* \mathbb{E}\left[Y_{j,t}^N\right].$$

Inconsistency of the LS Estimator

- Ferman and Pinto (2021) shows that the LS estimator is inconsistent; that is,

$$\widehat{w}_j^{\mathrm{LS}} \xrightarrow{p} \widetilde{w}_j \neq w_{\mathrm{j}}^*.$$

  - They propose another LS-based estimator that reduces the bias.
  - However, the estimator is still biased.

- Their results imply that the LS estimator is incompatible to SCMs under the linearity assumption, $\mathbb{E}\left[Y_{0,t}^N\right] = \sum_{j \in J^U} w_j^* \mathbb{E}\left[Y_{j,t}^N\right]$.

Implicit Endogeneity

- We investigate this problem from the viewpoint of endogeneity.
  - Let $Y_{j,t}^N = \mathbb{E}_{j,t}\left[Y_{j,t}^N\right] + \varepsilon_{j,t}$.
  - Under $\mathbb{E}\left[Y_{0,t}^N\right] = \sum_{j\in J^U} w_j^* \mathbb{E}\left[Y_{j,t}^N\right]$, it holds that

$$Y_{0,t}^N = \sum_{j\in\mathcal{J}^U} w_j^* Y_{j,t}^N - \sum_{j\in\mathcal{J}^U} w_j^* \varepsilon_{j,t} + \varepsilon_{0,t} = \sum_{j\in\mathcal{J}^U} w_j^* Y_{j,t}^N + v_t.$$

- **Implicit endogeneity** (measurement error bias): correlation between $Y_{j,t}^N$ and v_t.
  - There is an (implicit) endogeneity between the explanatory variable and the error term.
  - This is a reason why the LS estimator $\widehat{w}_j^{\mathrm{LS}}$ is biased; that is, $\widehat{w}_j^{\mathrm{LS}} \xrightarrow{p} \widetilde{w}_j \neq w_j^*$.

Mixture Models

- The implicit endogeneity implies that the LS estimator is incompatible to SCMs.

➢ Consider another estimation strategy.

- Assume **mixture models** and estimate the weights by the **GMM**.

- $p_{j,t}(y)$: density of $Y_{j,t}^N$

- Mixture models between $p_{0,t}(y)$ and $\{p_{j,t}(y)\}_{j\in\mathcal{J}^U}$:

$$p_{0,t}(y) = \sum_{j\in\mathcal{J}^U} w_j^* \, p_{j,t}(y).$$

Fine-Grained Models.

- Assuming mixture models is stronger than assuming $\mathbb{E}\left[Y_{0,t}^{N}\right] = \sum_{j \in J^U} w_j^* \mathbb{E}\left[Y_{j,t}^{N}\right]$.

➢ **Mixture models can be justified from the viewpoint of fine-grained models (Shi et al., 2021).**

- Linear factor models are usually assumed in SCMs:

$$Y_{j,t}^{N} = c_j + \delta_t + \lambda_t \mu_j + \varepsilon_{j,t}, \qquad Y_{j,t}^{I} = \tau_{0,t} + Y_{j,t}^{N}$$

- Shi et al., (2021) finds that mixture models imply factor models under some assumptions.

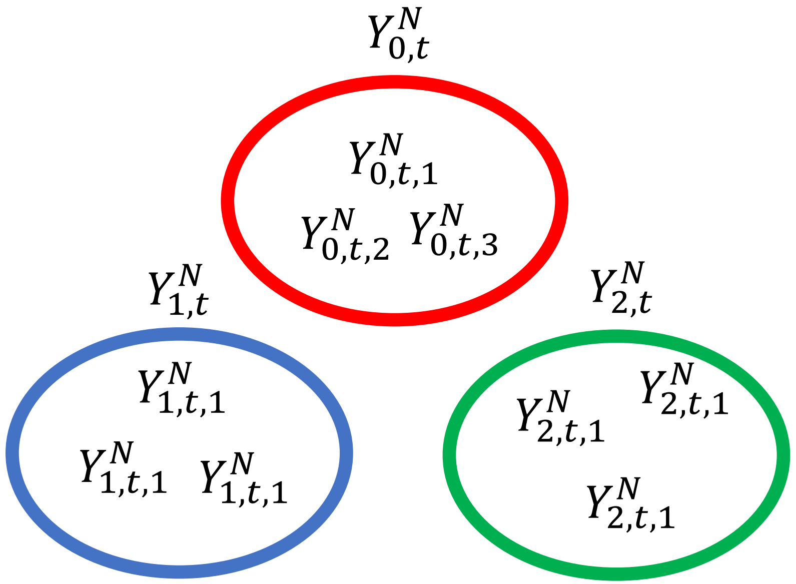
Fine–Grained Models.

- **Fine-grained models** (Shi et al., 2021).
  - Assume that $Y_{j,t}^N$ represents a group-level outcome.
  - In each unit j, there are unobserved small units $Y_{j,t1}^N, Y_{j,t2}^N, \ldots$.

= In each unit, there are unobserved units that constitute $Y_{j,t}^N$.

→ Under some assumptions,

- each $p_{j,t}(y)$ can be linked to the linear factor model, and
- $p_{0,t}(y) = \sum_{j \in \mathcal{J}^U} w_j^* \, p_{j,t}(y)$ holds.



Moment Conditions

➢ **Moment conditions.**

■ Under the mixture models, the following moment conditions hold:

$$\mathbb{E}_{0,t}\left[\left(Y_{0,t}^{N}\right)^{\gamma}\right] = \sum_{j\in\mathcal{J}^U} w_j^* \, \mathbb{E}_{j,t}\left[\left(Y_{j,t}^{N}\right)^{\gamma}\right] \qquad \forall\, \gamma \in \mathbb{R}^+.$$

■ Empirical approximation of $\mathbb{E}_{0,t}\left[\left(Y_{0,t}^{N}\right)^{\gamma}\right] - \sum_{j\in\mathcal{J}^U} w_j^* \, \mathbb{E}_{j,t}\left[\left(Y_{j,t}^{N}\right)^{\gamma}\right]$:

$$\widehat{m}_{\gamma}(w) := \frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left\{\left(Y_{0,t}^{N}\right)^{\gamma} - \sum_{j\in\mathcal{J}^U} w_j \left(Y_{j,t}^{N}\right)^{\gamma}\right\}.$$

- We estimate w to achieve $\widehat{m}_{\gamma}(w) \approx 0$.

GMM

- A set of positive values $\Gamma \coloneqq \{1,2,3,\dots,G\}$, e.x., $\Gamma = \{1,2,3,4,5\}$.

- Estimate w_j^* as

$$\left(\widehat{w}_j^{\mathrm{GMM}}\right)_{j\in\mathcal{J}^U} \coloneqq \arg \min_{(w_j):\sum_{j\in\mathcal{J}^U} w_j=1} \sum_{\gamma\in\Gamma} \left(\widehat{m}_\gamma(w)\right)^2.$$

  - We can weight each empirical moment condition; that is, by using some weight $v_\gamma \in \mathbb{R}^+$,

$$\left(\widehat{w}_j^{\mathrm{GMM}}\right)_{j\in\mathcal{J}^U} \coloneqq \arg \min_{(w_j):\sum_{j\in\mathcal{J}^U} w_j=1} \sum_{\gamma\in\Gamma} v_\gamma \left(\widehat{m}_\gamma(w)\right)^2.$$

- We can show that the GMM estimator is asymptotically unbiased; that is,

$$\widehat{w}_j^{\mathrm{GMM}} \xrightarrow{p} w_j^*.$$

Inference

- Hypothesis testing about the sharp null

$$H_0: \tau_{0,t} = 0 \quad \text{for} \quad t \in \mathcal{T}_1.$$

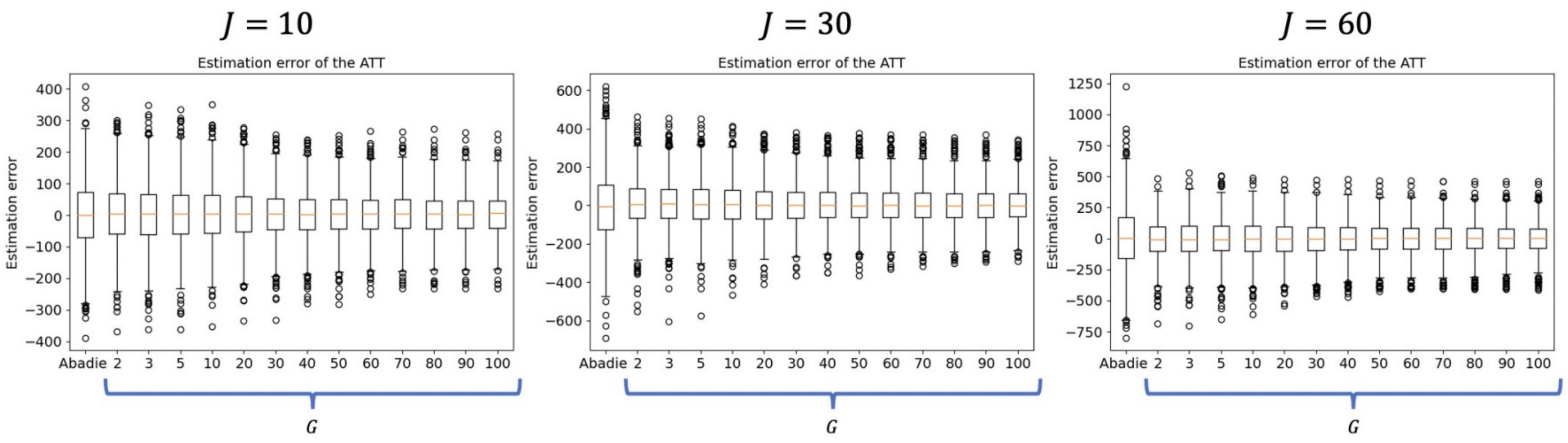
  - Note that $\tau_{0,t} = Y_{0,t}^I - Y_{0,t}^N$ under the linear factor model.

- We usually employ the **conformal inference** for testing the hypothesis.

- Nonparametrically test the sharp null.
- Computational costs.

Simulation Studies

- G is chosen from $\{2,3,5,10,20,30,40,50,60,70,80,90,100\}$. J is chosen from $\{10, 30, 60\}$.
  - Recall that $\left(\widehat{w}_j^{\mathrm{GMM}}\right)_{j\in\mathcal{J}^U} \coloneqq \arg\min_{(w_j):\sum_{j\in\mathcal{J}^U} w_j=1} \sum_{\gamma\in\{1,2,\ldots,G\}} \left(\frac{1}{T_0}\sum_{t\in\mathcal{T}_0}\left\{\left(Y_{0,t}^N\right)^\gamma - \sum_{j\in\mathcal{J}^U} w_j \left(Y_{j,t}^N\right)^\gamma\right\}\right)^2.$
- Generate $Y_{j,t}^N$ from gaussian distributions.
- The y–axis denotes the estimation error, and the x–axis denotes G.

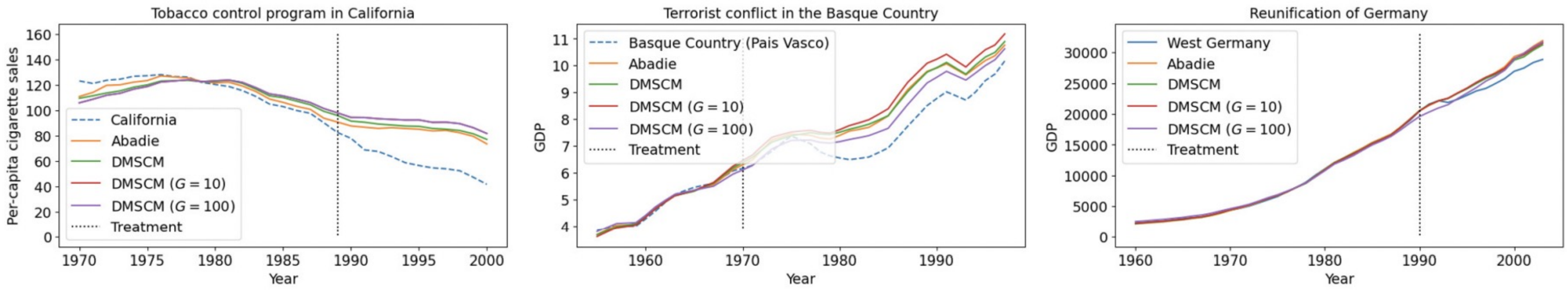


Empirical Studies

- Empirical analysis using case studies in existing studies.
  - Tobacco control in California (Abadie, Diamond and Hainmueller, 2010).
  - Basque conflict in the Basque country (Abadie and Gardeazabal, 2003).
  - Reunification of Germany (Abadie, Diamond and Hainmueller, 2015).

- Pretreatment fit: Predictive ability for outcomes for $t \in T_0$.

Figure 2: Results of simulation studies. Estimation errors of ATT estimation.



Bayesian Predictive Synthetic Control Methods

Bayesian SCMs

- We introduced frequentist method for SCMs.
- Frequentist SCMs require
  - Large samples for showing the convergence of the weight estimators.
  - Special inference methods, such as conformal inference.
  - Distance minimization to employ covariates, which is not easy to be justified.
- Consider <u>Bayesian</u> approach for SCMs.
  - Works with finite samples.
  - Inference with posterior distribution.

Bayesian Predictive Synthesis

- Our Bayesian SCMs are based on the formulation of Bayesian predictive synthesis (BPS).
- BPS: a method for synthesizing predictive models (McAlinn and West, 2019).
  - Synthesize predictive models with reflecting the model uncertainty.
  - A generalization of Bayesian model averaging.
  - Incorporating various predictive models with weighting them time-varying parameters.
- We regard untreated outcomes and predictive models for the outcomes using covariates as predictors of $Y_{0,t}^{N}$
  - We first predict outcomes using covariates.
  - Then, we incorporate the predictors using the BPS.

BPSCM

- We propose SCMs with the **BPS**, referred to as the **BPSCMs.**

➢ **BPSCM.**

- Φ_t: a set of time−varying parameters at t. Φ_t depends on $\{Y_{0,t+1}^N\}_{t\in[1:t]}$.

- The conditional density function of $Y_{0,t+1}^N$ is referred to as the **synthesis function, denoted by**

$$\alpha\left(y\middle|\{Y_{j,t}^N\}_{j\in\mathcal{J}^U},\Phi_\mathrm{t}\right).$$

- Bayesian decision maker predicts $Y_{0,t+1}^N$ using the **posterior distribution** defined as

$$p^N(y|\{Y_{0,t+1}^N\}_{t\in[1:t]},\Phi_t) \coloneqq \int \alpha(y|\{y_{j,t}^N\}_{j\in\mathcal{J}^U},\Phi_t)\prod_{j\in\mathcal{J}^U} p_{j,t}(y_{j,t}^N)\mathrm{d}y_{j,t}^N.$$

Dynamic Latent Factor Linear Regression Models

- There are several specifications for the synthesis function.
- Ex. Latent factor dynamic linear model:
  - Set the synthesis function as $\alpha\left(y_{0,t}^N \middle| \left\{Y_{j,t}^N\right\}_{j\in\mathcal{J}^U}, \Phi_t\right) = \phi\left(y_{0,t}^N, ; w_{0,t} + \sum_{j=1}^J w_{j,t} Y_{j,t}^N, \nu_t\right)$.
    - $\phi(\cdot; a, b^2)$: a univariate normal density with mean a and variance b^2.
    - v_t are unobserved error terms.
  - Specify the process of $Y_{0,t}^N$ and $w_{t,j}$ as

$$Y_{0,t}^N = w_{0,t} + \sum_{j\in\mathcal{J}^U} w_{t,j} Y_{j,t}^N + \epsilon_t, \quad \epsilon_t \sim N(0, \nu_t), \qquad w_{t,j} = w_{t-1.j} + \eta_{t,j}, \qquad \eta_{t,j} \sim N(0, \nu_t \boldsymbol{W}_t),$$

Auxiliary Covariates

➢ The BPSCM can use covariates by predicting outcomes using various predictive models.

- $X_{j,t}$: Covariates for the unit j.

■ Define L predictors for $Y_{j,t}^N$ by $\{\hat{f}^l(X_{j,t})\}_{l=1}^L$.

  - These predictors can be constructed from machine learning methods.
  - We can use covariates in the predictive models.

■ With the original untreated outcomes $Y_{j,t}^N$, there are $K = (1 + L)L$ predictrs $\left\{Y_{j,t}^N, \{\hat{f}^l(x_{j,t})\}_{l=1}^L\right\}_{j=1}^J$.

■ We incorporate them by using the BPS.

Auxiliary Covariates

- A set of predictors are denoted by

$$\boldsymbol{Z}_t = \{Y_{1,t}^N, \ldots, Y_{J,t}^N, \hat{f}^1(X_{1,t}), \ldots, \hat{f}^L(X_{1,t}), \hat{f}^1(X_{2,t}), \ldots, \hat{f}^L(X_{J-1,t}), \hat{f}^1(X_{J,t}), \ldots, \hat{f}^L(X_{J,t})\}$$

- Conduct BPSCM as if there are $J + JL$ untreated units that can be used for SCMs:

$$p^N\left(y \mid \Phi_t, \{Y_{0,t}^N\}_{t\in[1:t]}\right) = \int \alpha\left(y \mid \boldsymbol{z}_t, \Phi_t\right) \prod_{j\in\{1,2,\ldots,(1+L)J\}} p_{j,t}\left(z_{j,t}\right) \mathrm{d}z_{j,t}\,.$$

Ex. Synthesize predictive models such as linear regression and random forest.

Advantages of the BPSCM

✓ Time-varying parameters.

✓ Incorporate uncertainty of each untreated outcome’s outcome.

✓ Minimax optimality

- Even under model misspecification, predictor of the BPSCM is minimax optimal in terms of KL divergence (Takanashi and McAlinn, 2021).
- Avoid the implicit endogeneity problem?

✓ Works with finite samples.

✓ Inference (posterior distribution).

Empirical Analysis

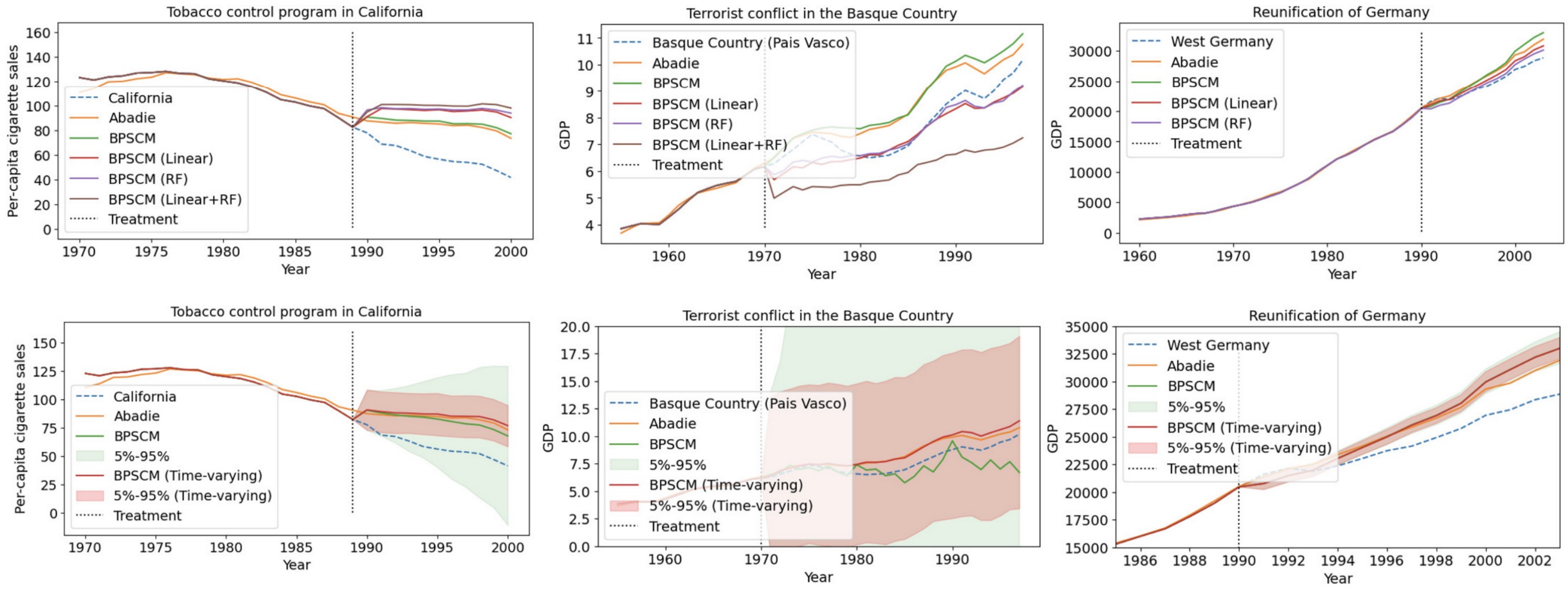
- Empirical studies using the same case studies in the previous slide.

- Compare following five prediction models.

	Time-varying coef.s	Using covariates	Synthesized predictive models
Abadie		✓	–
BPSCM	✓		–
BPSCM (Linear)	✓	✓	Least squares
BPSCM (RF)	✓	✓	Random forests
BPSCM (Linear + RF)	✓	✓	Least squares + random forests

Empirical Analysis

- We mainly check the pretreatment fit and posterior distribution of the BPSCMs.



Conclusion

- **SCMs suffer from the issue of inconsistency.**
  - The LS estimator is incompatible to the assumption, $\mathbb{E}\left[Y_{0,t}^{N}\right] = \sum_{j \in J^U} w_j^* \mathbb{E}\left[Y_{j,t}^{N}\right]$.

  → Implicit endogeneity (measurement error bias).
  - $Y_{0,t}^{N} = \sum_{j \in J^U} w_j^* Y_{j,t}^{N}$ is not realistic…?

- **Frequentist density matching (Mixture model + GMM).**
  - Mixture model $p_{0,t}(y) = \sum_{j \in \mathcal{J}^U} w_j^* \, p_{j,t}(y)$, a stronger assumption than $\mathbb{E}\left[Y_{0,t}^{N}\right] = \sum_{j \in J^U} w_j^* \mathbb{E}\left[Y_{j,t}^{N}\right]$.
  - By using the GMM under the assumption, we can estimate the weight consistently.

- **BPSCM.**
  - By using the Bayesian method, we can obtain the minimax optimal predictor without assuming the mixture models without assuming mixture models.
  - Advantages such as flexible modeling and finite sample inference.

Reference

- Abadie, A. and Gardeazabal, J. “The economic costs of conflict: A case study of the basque country.” American Economic Review, 2003.
- Abadie, A., Diamond, A., and Hainmueller, J. “Synthetic control methods for comparative case studies: Estimating the effect of california’s tobacco control program.” Journal of the American Statistical Association, 2010
- Abadie, A., Diamond, A., and Hainmueller, J. “Comparative politics and the synthetic control method.” American Journal of Political Science, 2015
- Ferman, B. and Pinto, C. Synthetic controls with imperfect pretreatment fit. Quantitative Economics, 12(4):1197–1221, 2021.
- McAlinn, K. and West, M., “Dynamic Bayesian predictive synthesis in time series forecasting,” Journal of econometrics, 2019
- McAlinn, K., Aastveit, K. A., Nakajima, J., and West, M. “Multivariate Bayesian predictive synthesis in macroeconomic forecasting.” Journal of the American Statistical Association, 2020
- Shi, C., Sridhar, D., Misra, V., and Blei, D. On the assumptions of synthetic control methods. In AISTATS, pp. 7163–7175, 2022
- Takanashi, K. and McAlinn, K. “Predictions with dynamic bayesian predictive synthesis are exact minimax”, 2021
- West, M. and Harrison, P. J. “Bayesian Forecasting & Dynamic Models.” Springer Verlag, 2nd edition, 1997