

Bayesian Synthetic Control with a Soft Simplex Constraint

Yihong Xu¹ Quan Zhou²

¹Dept. of Economics, Texas A&M University

²Dept. of Statistics, Texas A&M University

2025 Southern Economic Association

Nov 22, 2025

Overview

- The Synthetic Control Method (SCM) has at least two limitations;
 - High-dimensional problem: the number of control units is larger than the total observed time periods before the intervention;
 - Debate over the simplex constraint: all coefficients are restricted to be non-negative and sum-to-one.
- We propose a novel extension of SCM, using Bayesian variable selection (spike-and-slab prior) with a soft simplex constraint, BVS-SS;
 - Contribution 1: BVS-SS solves the high dimension problem while still preserves the simplex constraint;
 - Contribution 2: BVS-SS introduces a notion of “**soft**” simplex constraint, asking data’s advice on whether the constraint is proper.

SCM Setup

- T_0 is # of pre-treatment periods; T_1 is # of post-treatment periods;
- $X_{1,t}$ is the outcome of the treated unit; $X_{0,t}$ is the $N \times 1$ outcome vector of the control units;

$$\hat{w} = \underset{w \in \Delta^{N-1}}{\operatorname{argmin}} \frac{1}{T_0} \sum_{t=1}^{T_0} (X_{1,t} - X_{0,t}^\top w)^2, \quad (1)$$

$$\Delta^{N-1} = \left\{ u \in \mathbb{R}^N : u_i \geq 0 \text{ for each } i, \text{ and } \sum_{i=1}^N u_i = 1 \right\}$$

where Δ^{N-1} denotes the $(N-1)$ -simplex.

- The average treatment on the treated:

$$\widehat{\text{ATT}} = \frac{1}{T_1} \sum_{t=T_0+1}^T (X_{1,t} - X_{0,t}^\top \hat{w})$$

Challenges

- How to solve the optimization (1) when X_0 is high-dimensional ($N \gg T_0$)?
 - Panel data approach: Elastic net ([Doudchenko and Imbens, 2016](#)), Lasso ([Carvalho et al., 2018](#); [Hollingsworth and Wing, 2020](#)), forward selection ([Shi and Huang, 2023](#)), spike-and-slab ([Kim et al., 2020](#)) etc.;
 - Trade-off: dropping the simplex constraint;

Challenges

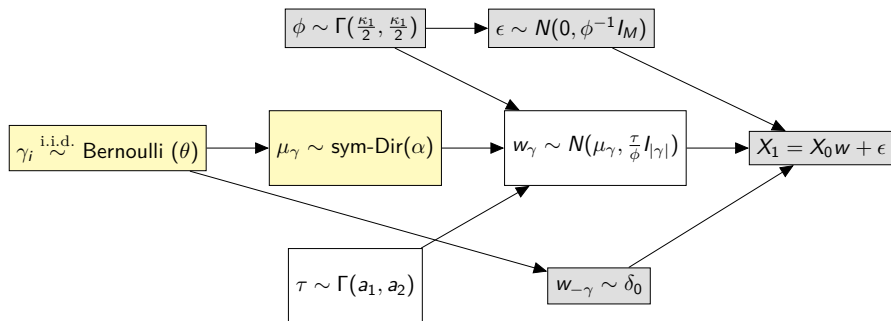
- How to solve the optimization (1) when X_0 is high-dimensional ($N \gg T_0$)?
 - Panel data approach: Elastic net ([Doudchenko and Imbens, 2016](#)), Lasso ([Carvalho et al., 2018](#); [Hollingsworth and Wing, 2020](#)), forward selection ([Shi and Huang, 2023](#)), spike-and-slab ([Kim et al., 2020](#)) etc.;
 - Trade-off: dropping the simplex constraint;
- Is the simplex constraint necessary?
 - Pros: no hedging (e.g., $w_1 = 300\%$ and $w_2 = -250\%$); the counterfactual has economic interpretation;
 - Cons: too rigid (may have bad fitting); high-dimensional problem.

Summary

We propose a novel Bayesian SCM framework, based on Bayesian variable selection with a soft simplex constraint(BVS-SS). It

- ① deals with the high-dimensional control group, while still keeping the simplex constraint;

Summary



Remarks:

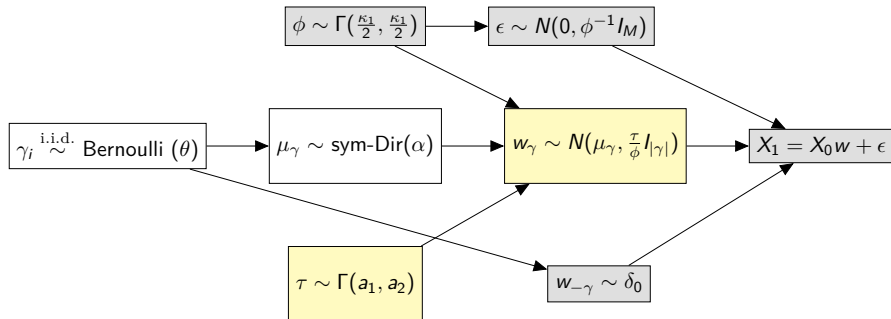
- A small θ penalizes the large models;
- Main diff from spike-and-slab: $\mu_\gamma \in \Delta^{|\gamma|-1}$;

Summary

We propose a novel Bayesian SCM framework, based on Bayesian variable selection with a soft simplex constraint(BVS-SS). It

- ① deals with the high-dimensional control group, while still keeping the simplex constraint;
- ② learns the attitude of data on the constraint case by case.

Summary



Remark:

- Variance τ : quantifying the deviation from simplex (mean μ_γ).

Hindrance

- Once we have samples from the posterior $p(\mu_\gamma, \gamma, \tau, \phi \mid X_1)$ [Details](#), we can predict the counterfactual by

$$\mathbb{E}_p[\tilde{X}_0 w] = \mathbb{E}_p \left[\mathbb{E}_p(\tilde{X}_0 w \mid \mu_\gamma, \gamma, \tau, \phi) \right]; \widehat{\text{ATT}} = \frac{1}{\tilde{T}} \sum_{t=1}^{\tilde{T}} \left(\tilde{X}_{1,t} - \mathbb{E}_p[\tilde{X}_0 w]_t \right).$$

- How to generate the samples from the posterior distribution $p(\mu_\gamma, \gamma, \tau, \phi \mid X_1)$? A standard Gibbs sampling would fail due to the simplex constraint;
- What is the (non-asymptotic) property of the posterior distribution $p(\mu_\gamma, \gamma, \tau, \phi \mid X_1)$?

A Gibbs Scheme with the Simplex Constraint

How to generate the samples from the posterior distribution $p(\mu_\gamma, \gamma, \tau, \phi \mid X_1)$?

- Gibbs sampling: Fix μ_{-j} , update μ_j according to $p(\mu_j \mid Y, \mu_{-j}, \tau, \phi)$;
- $\mu_j = 1 - \sum_{i \neq j} \mu_i$ almost surely; $p(\mu_j \mid Y, \mu_{-j}, \tau, \phi)$ is degenerate;
- Standard methods like [Geyer \(1992\)](#); [George and McCulloch \(1993\)](#); [Gilks et al. \(1995\)](#); [George and McCulloch \(1997\)](#); [Brooks \(1998\)](#); [Carlo \(2004\)](#) don't work.

Update μ_i, μ_j (with $i \neq j$) simultaneously from the full conditional posterior $p(\mu_i, \mu_j \mid X_1, \mu_{-(i,j)}, \tau, \phi)$.

- 1 Selection based on $p(\gamma_i, \gamma_j \mid X_1, \mu_{-(i,j)}, \tau, \phi)$;
- 2 Sampling based on $p(\mu_i, \mu_j \mid X_1, \gamma_i, \gamma_j, \mu_{-(i,j)}, \tau, \phi)$

► Details →

Non-asymptotic Bounds ($\tau = 0$)

Fix $\tau = 0$. The true model strictly satisfies the simplex constraint. We show the convergence rate of the posterior probability of selecting the true model γ^* .

Theorem 1

Under mild conditions, it is true that

$$p(\gamma^* \mid Y) \geq 1 - c_3 N^{-1},$$

with probability at least $1 - c_1 N^{-c_2 L}$.

Corollary 1

Under the setting of Theorem 1,

$$\mathbb{E} [\|\mu - \mu^*\|_2^2 \mid Y] = O_p \left(C_4 \frac{L \log N}{T_0} \right).$$

More Results

- We also show the property when the simplex constraint is substantially violated ($\tau \rightarrow \infty$);
- Several simulation studies show the advantages of BVS-SS;
- Revisit an empirical example, with 87 control units but only 35 pre-intervention periods. [▶ Details →](#)

Concluding Remarks

- We construct a novel Bayesian synthetic control framework to solve the high dimension issue (BVS-SS);
 - The posterior of the weight shows an empirical attitude on the simplex constraint;
 - [Li \(2020\)](#); [Chernozhukov et al. \(2021\)](#); [Martinez and Vives-i-Bastida \(2022\)](#); [Goh and Yu \(2022\)](#);
- We propose a novel Markov chain Monte Carlo (MCMC) sampler for our BVS-SS posterior;
 - to select the “proper” control units and estimate the weights and counterfactuals through our posterior;
 - [Geyer \(1992\)](#); [George and McCulloch \(1993\)](#); [Gilks et al. \(1995\)](#); [George and McCulloch \(1997\)](#); [Brooks \(1998\)](#); [Carlo \(2004\)](#);
- We derive non-asymptotic bounds for our Bayesian estimations;
 - on control units selection and posterior mean of coeffs/weights;
 - [George and McCulloch \(1993, 1997\)](#); [Brown et al. \(1998\)](#); [Li and Li \(2008\)](#); [Casella et al. \(2009\)](#); [Choi et al. \(2010\)](#); [Farcomeni \(2010\)](#); [Guan and Stephens \(2011\)](#); [Yang et al. \(2016\)](#); [Zhou and Guan \(2019\)](#).

Thank You!

References I

- S. Brooks. Markov chain monte carlo method and its application. *Journal of the royal statistical society: series D (the Statistician)*, 47(1):69–100, 1998.
- P. J. Brown, M. Vannucci, and T. Fearn. Multivariate Bayesian variable selection and prediction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3):627–641, 1998.
- C. M. Carlo. Markov chain monte carlo and gibbs sampling. *Lecture notes for EEB*, 581(540):3, 2004.
- C. Carvalho, R. Masini, and M. C. Medeiros. Arco: An artificial counterfactual approach for high-dimensional panel time-series data. *Journal of econometrics*, 207(2):352–380, 2018.
- G. Casella, F. J. Girón, M. L. Martínez, and E. Moreno. Consistency of Bayesian procedures for variable selection. *The Annals of Statistics*, 37(3):1207–1228, 2009.
- V. Chernozhukov, K. Wüthrich, and Y. Zhu. An exact and robust conformal inference method for counterfactual and synthetic controls. *Journal of the American Statistical Association*, 116(536):1849–1864, 2021.

References II

- N. H. Choi, W. Li, and J. Zhu. Variable selection with the strong heredity constraint and its oracle property. *Journal of the American Statistical Association*, 105(489):354–364, 2010.
- N. Doudchenko and G. W. Imbens. Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Technical report, National Bureau of Economic Research, 2016.
- A. Farcomeni. Bayesian constrained variable selection. *Statistica Sinica*, 20(3): 1043–1062, 2010. URL <https://www.jstor.org/stable/24309479>.
- E. I. George and R. E. McCulloch. Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889, 1993.
- E. I. George and R. E. McCulloch. Approaches for Bayesian variable selection. *Statistica sinica*, pages 339–373, 1997.
- C. J. Geyer. Practical markov chain monte carlo. *Statistical science*, pages 473–483, 1992.
- W. R. Gilks, S. Richardson, and D. Spiegelhalter. *Markov chain Monte Carlo in practice*. CRC press, 1995.

References III

- G. Goh and J. Yu. Synthetic control method with convex hull restrictions: a bayesian maximum a posteriori approach. *The Econometrics Journal*, 25(1): 215–232, 2022.
- Y. Guan and M. Stephens. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *The Annals of Applied Statistics*, 5(3):1780, 2011.
- A. Hollingsworth and C. Wing. Tactics for design and inference in synthetic control studies: An applied example using high-dimensional data. *Available at SSRN 3592088*, 2020.
- S. Kim, C. Lee, and S. Gupta. Bayesian synthetic control methods. *Journal of Marketing Research*, 57(5):831–852, 2020.
- C. Li and H. Li. Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics*, 24(9):1175–1182, 2008. doi: 10.1093/bioinformatics/btn081.

References IV

- K. T. Li. Statistical inference for average treatment effects estimated by synthetic control methods. *Journal of the American Statistical Association*, 115(532):2068–2083, 2020.
- I. Martinez and J. Vives-i-Bastida. Bayesian and frequentist inference for synthetic controls. *arXiv preprint arXiv:2206.01779*, 2022. URL <https://arxiv.org/abs/2206.01779>.
- Z. Shi and J. Huang. Forward-selected panel data approach for program evaluation. *Journal of Econometrics*, 234(2):512–535, 2023.
- R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.
- Y. Yang, M. J. Wainwright, and M. I. Jordan. On the computational complexity of high-dimensional Bayesian variable selection. *The Annals of Statistics*, 44(6):2497–2532, 2016.
- Q. Zhou and Y. Guan. Fast model-fitting of Bayesian variable selection regression using the iterative complex factorization algorithm. *Bayesian analysis*, 14(2):573, 2019.

Motivation 1: High Dimensional X (Contd)

To see why no frequentist implementations with simplex, consider the standard Lasso regression ([Tibshirani, 1996](#)).

$$\hat{w} = \underset{w}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^T (X_{1,t} - X_{0,t}^\top w)^2 \quad \text{such that } \|w\|_1 \leq \lambda$$

where λ determines the degree of regularization.

The constraint in Lasso contradicts the simplex constraint Δ^{N-1} !

▶ Back ◀

Posterior and ATT

A routine calculation using Woodbury identity yields the marginal likelihood of (μ, τ, ϕ) :

$$p(X_1 \mid \mu, \tau, \phi) \propto \frac{\phi^{T/2}}{\tau^{|\gamma|/2} \det(V_{\gamma, \tau})^{1/2}} \exp \left\{ -\frac{\phi}{2} (X_1 - X_{0, \gamma} \mu_\gamma)^\top \Sigma_{\gamma, \tau} (X_1 - X_{0, \gamma} \mu_\gamma) \right\},$$

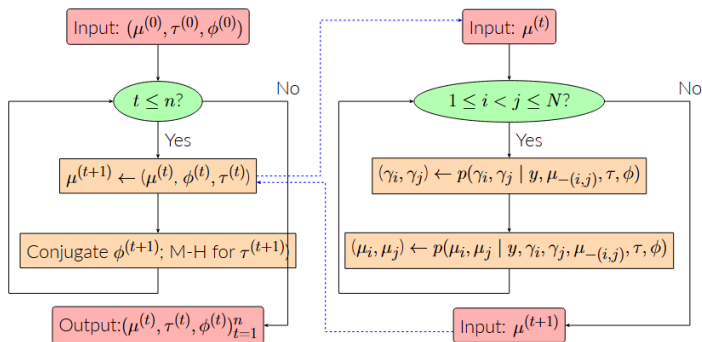
where

$$V_{\gamma,\tau} = X_{0,\gamma}^\top X_{0,\gamma} + \tau^{-1}I, \quad \Sigma_{\gamma,\tau} = I - X_{0,\gamma} V_{\gamma,\tau}^{-1} X_{0,\gamma}^\top.$$

w 's conditional posterior:

$$w \mid X_1, \mu, \tau, \phi \sim N(V_{\gamma, \tau}^{-1}(X_{0, \gamma}^\top X_1 + \tau^{-1} \mu_\gamma), \phi^{-1} V_{\gamma, \tau}^{-1}).$$

Metropolis-within-Gibbs



- When $\alpha = 1$, γ 's conditional posterior can be expressed by CDF of standard normals; given $\gamma_i = \gamma_j = 1$, μ_i and μ_j 's marginal posteriors degenerate to a **truncated Gaussian** distribution.

China's anti-corruption campaign

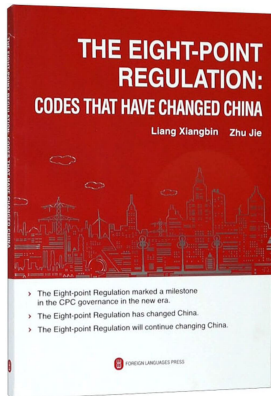


Journal of Econometrics
Volume 234, Issue 2, June 2023, Pages 512-535



Forward-selected panel data approach for program evaluation

Zhentao Shi ^a, Jingyi Huang ^b



- Shi and Huang (2023): China anti-corruption campaign in 2012;
- X_1 : monthly growth rate of luxury watch imports;
- Category: “watches with case of, or clad with, precious metal”;
- X_0 : monthly growth rate of other 87 commodity categories;
- Only $T_0 = 35$ pre-treatment periods.

China's anti-corruption campaign

Table: Anti-corruption Campaign's Impact on Luxury Watches Importation

	ATT	τ	ϕ	$ \gamma $
Mean	-0.021	0.069	20.86	5.09
95% credible interval	(-0.032, -0.008)	(0, 0.641)	(12.22, 32.76)	(1, 20)

Remark: The ratio τ/ϕ is very small \rightarrow empirical support on the validity of SCM type estimator.

Control Units Selection

- Shi and Huang (2023):

- ① knitted or crocheted fabric;
- ② cork and articles of cork;
- ③ salt, sulfur, earth, stone, plaster, lime and cement.

- BVS-SS:

- ① knitted or crocheted fabric (51.4%);
- ② special woven fabrics (lace, tapestries, trimmings, embroidery) (16.6%);
- ③ man-made filaments (14.8%);
- ④ wool, fine or coarse animal hair (12%);
- ⑤ other fabric/textile;
- ⑥ optical, photographic, cinematographic, measuring, checking, medical or surgical instruments and apparatus; parts and accessories" (10.2%).

Remarks: seems a more reasonable selection.

Counterfactual Estimates: Anti-corruption

Import value (normalized)

Time

Legend:

- Observation (solid black line)
- Post Counterfactual (dashed red line)
- Treatment (solid green vertical line)

Time	Observation	Post Counterfactual
2010-Feb	0.8	-
2011-Feb	1.5	-
2012-Jan	2.5	-
2013-Jan	1.8	1.3
2014-Jan	1.2	1.4
2015-Dec	1.1	1.7

Figure: Significant impact right after the Eight-point regulation