Bayesian Synthetic Control with a Soft Simplex Constraint

Presenter: Quan Zhou

Department of Statistics

Texas A&M University

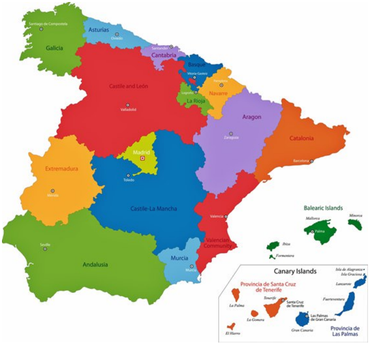
Economic Analysis of Basque Country

The Economic Costs of Conflict: A Case Study of the Basque Country

Alberto Abadie

Javier Gardeazabal

AMERICAN ECONOMIC REVIEW
VOL. 93, NO. 1, MARCH 2003
(pp. 113–132)



From `orangesmile.com/travelguide/spain`

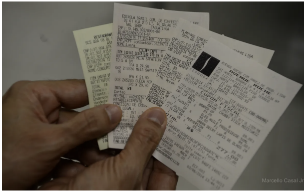
Synthetic control: construct a synthetic *counterfactual* variable and use it as the "control" for estimating treatment effect when no true control data is available.

Anti-tax Evasion Program in Brazil

ELSEVIER
Journal of Econometrics
Volume 207, Issue 2, December 2018, Pages 352-380

ArCo: An artificial counterfactual approach for high-dimensional panel time-series data

Carlos Carvalho [a b 1], Ricardo Masini [c], Marcelo C. Medeiros [b 2]



From cnnbrasil.com.br

Made by ChatGPT

Did Nota Fiscal Paulista raise product prices?

Synthetic Control & Simplex Constraint

In Abadie and Gardeazabal (2003), the counterfactual Y (e.g., GDP of the Basque country in 1980 if there were no terrorism activities) is constructed as

$$Y = \sum_{i=1}^{p} w_i X_i,$$

where $X_1, \ldots, X_p$ are donor units (e.g., the GDP of p other regions in 1980), and $(w_1, \ldots, w_p)$ must satisfy the *simplex* constraint:

$$w_i \geq 0 \text{ for each } i, \quad \sum_{i=1}^{p} w_i = 1.$$

- $w_1, \ldots, w_p$ are often estimated via (weighted) least squares using training data (e.g. GDP data before the terrorism activities started).
- Sparsity is often assumed: most weights are set to zero.

Synthetic Control & Simplex Constraint

Essentially, synthetic control is a sparse linear regression problem with simplex constraint on the regression coefficients.

On the use of simplex constraint:

- *Advantage:* it makes the regression coefficients easier to interpret and avoids extrapolation.
- *Disadvantage:* it may not hold in practice.
- A longstanding debate in the econometrics community [King and Zeng, 2006, Hsiao et al., 2012, Doudchenko and Imbens, 2016, Xu, 2017, Li, 2020, Abadie, 2021, Goh and Yu, 2022, Martinez and Vives-i-Bastida, 2022].

Bayesian Variable Selection

The standard spike-and-slab variable selection model typically assumes:

$$\begin{aligned} Y &\sim N(Xw, \phi^{-1} I), \\ \phi &\sim \mathrm{Gamma}(\kappa_1/2, \kappa_2/2), \\ \gamma_i &\overset{\text{i.i.d.}}{\sim} \mathrm{Bernoulli}(\theta), \\ w_\gamma \mid \gamma, \mu_\gamma, \tau, \phi &\sim N(\mu_\gamma,\ (\tau/\phi) I), \\ w_i \mid \gamma_i = 0 &\sim \delta_0. \end{aligned}$$

Often one uses $\mu_\gamma = 0$ and chooses τ according to prior knowledge.

A New Bayesian Variable Selection Model

We propose a new model, BVS-SS (Bayesian Variable Selection with Soft Simplex constraint):

$$
\begin{aligned}
Y &\sim N(Xw, \phi^{-1} I), \\
\phi &\sim \mathrm{Gamma}(\kappa_1/2, \kappa_2/2), \\
\gamma_i &\overset{\text{i.i.d.}}{\sim} \mathrm{Bernoulli}(\theta), \\
\tau &\sim \mathrm{Gamma}(a_1, a_2), \\
\mu_\gamma \mid \gamma &\sim \textsf{sym-Dirichlet}(\alpha), \\
w_\gamma \mid \gamma, \mu_\gamma, \tau, \phi &\sim N(\mu_\gamma,\, (\tau/\phi) I), \\
w_i \mid \gamma_i = 0 &\sim \delta_0.
\end{aligned}
$$

Remarks on BVS-SS

- μ_γ satisfies the simplex constraint a.s.
- $p(w_\gamma \mid \gamma, \mu_\gamma, \tau, \phi)$ interpolates between the simplex constraint and the unconstrained setting.
- If the data suggests that the simplex constraint is likely to be satisfied, the posterior of τ should concentrate around 0.
- If the data suggests that the simplex constraint is significantly violated, τ should stay away from 0 in the posterior.
- Our simulation study indicates that τ can be reasonably estimated even with a relatively small sample size.

Posterior Distribution

The joint posterior distribution of $(\gamma, \mu, \tau, \phi)$ can be computed by

$$p(\gamma, \mu, \tau, \phi \mid y) \propto p(y \mid \gamma, \mu, \tau, \phi) p(\mu_\gamma \mid \gamma) p(\gamma) p(\tau) p(\phi)$$

where $p(y \mid \gamma, \mu, \tau, \phi)$ can be computed in closed-form since

$$Y \mid \gamma, \mu_\gamma, \tau, \phi \sim N\left(X_\gamma \mu_\gamma,\, \phi^{-1}(\tau X_\gamma X_\gamma^\top + I)\right).$$

Add-delete-swap Samplers

For standard Bayesian spike-and-slab variable selection with fixed μ_γ and τ, Metropolis–Hastings algorithms can be used to approximate $p(\gamma, \phi \mid y)$, where γ is updated by add-delete-swap proposals.

For BVS-SS, $p(y \mid \gamma, \tau, \phi)$ involves an integral over the simplex and thus has no closed-form expression. Therefore, the standard marginal add-delete-swap scheme cannot be applied directly.

A New Metropolis-within-Gibbs Scheme

We devise a Metropolis-within-Gibbs scheme targeting $p(\gamma, \mu, \tau, \phi \mid y)$.

- ϕ can be drawn from $p(\phi \mid y, \gamma, \mu, \tau)$ (Gamma).
- τ is updated by a Metropolis–Hastings step with a random walk proposal on the log scale.
- Since $\gamma_i = \mathbf{1}\{\mu_i \neq 0\}$, the inclusion indicator and coefficient are updated jointly. Due to the simplex constraint, a one-coordinate update is degenerate.
- Draw $(\gamma_i, \gamma_j, \mu_i, \mu_j)$ (with $i \neq j$) from its full conditional posterior. Its probabilities can be evaluated using the CDF of a univariate normal distribution.

Selection Consistency under the Simplex Constraint

Assume $\tau = 0$. Conditional on ϕ,

$$p(\gamma \mid y, \phi) \;\propto\; p(\gamma)\Gamma(|\gamma|) \int_{\Delta^{|\gamma|-1}} \exp\left\{-\frac{\phi}{2}\|y - X_\gamma \mu_\gamma\|_2^2\right\} \mathrm{d}\mu_\gamma,$$

where Δ^k denotes the k-dimensional simplex.

Theorem

Under mild high-dimensional conditions,

$$p(\gamma^* \mid y) \to 1, \text{ in probability,}$$

with respect to the true data-generating probability measure.

Remarks on the Consistency Result

High-dimensional conditions we require include:

- The columns of X are normalized, and the minimum eigenvalues of $X_\gamma^\top X_\gamma/n$ are bounded away from zero for sparse models.
- The true regression coefficients satisfy the simplex constraint and a beta-min condition of order $\sqrt{L\log(p)/n}$, where L bounds the model size.
- The prior inclusion probability is sufficiently small so that the prior penalizes large models.
- The sample size is sufficiently large relative to the true model size, L, and $\log p$.

Estimation and Prediction Consistency

Under the same high-dimensional conditions,

$$\mathbb{E}_y\|\mu - \mu^*\|_2^2 = O_p\left(\max\left\{\frac{\sigma^2 L \log p}{\lambda^2 n},\ p^{-L/(\ell^*)^2}\right\}\right),$$

where $\ell^* = |\gamma^*|$ and λ controls the restricted eigenvalues of the design.

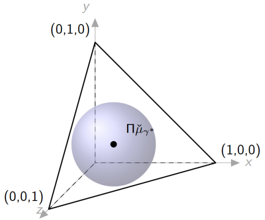
- The posterior expected prediction loss converges at the corresponding rate for a new design matrix.
- Consequently, the posterior estimator of the average treatment effect is consistent.

Proof Idea for Selection Consistency

For any μ_γ satisfying the simplex constraint,

$$\begin{aligned} \|y - X_\gamma \mu_\gamma\|_2^2 = \|y - X_\gamma \hat{\mu}_\gamma\|_2^2 + b_\gamma \\ + \|X_\gamma(\mu_\gamma - \check{\mu}_\gamma)\|_2^2. \end{aligned}$$

- $\hat{\mu}_\gamma$ is the unconstrained least-squares estimator.
- $b_\gamma \geq 0$ measures its deviation from the sum-to-one constraint.
- $X_\gamma \check{\mu}_\gamma$ is its projection onto the corresponding affine space.



A concentration ball around the projected estimator lies within the simplex.

When $\tau \to \infty$

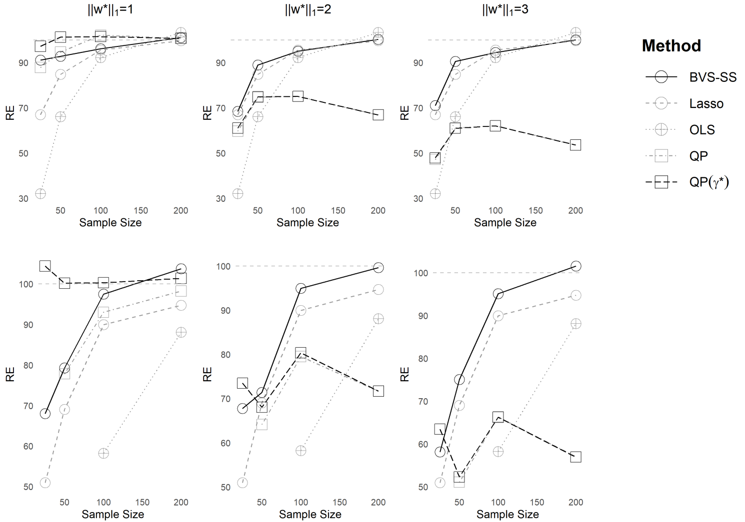
Theorem

Let $\tilde{p}$ denote the posterior distribution under the unconstrained spike-and-slab model. Then, for fixed data,

$$\lim_{\tau \to \infty} \frac{p(\gamma \mid y, \tau)}{\tilde{p}(\gamma \mid y, \tau)} = 1.$$

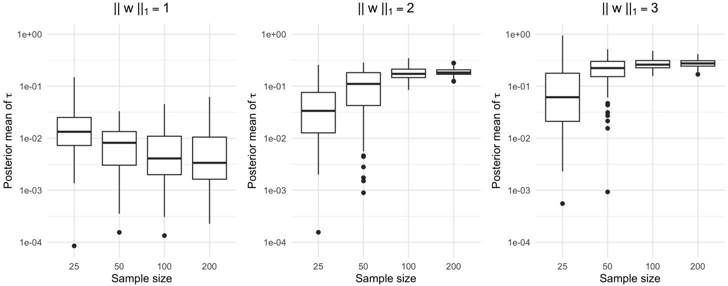
Thus, when τ is large, the BVS-SS posterior approaches its unconstrained counterpart.

Simulation Studies: Treatment Effect Estimation



Top row: $p = 20$, $|\gamma^*| = 5$. Bottom row: $p = 50$, $|\gamma^*| = 10$. Relative efficiency is measured against OLS(γ^*).

Simulation Studies: Posterior Mean of τ



Distribution of the posterior mean of τ across 100 replicates with $p = 50$ and $|\gamma^*| = 10$.

Anti-tax Evasion Program in Brazil

For this data set, $p = 8, n = 33$.

Table: NFP Impact on Food Inflation

	ATT	τ	ϕ	$\lvert\gamma\rvert$
Mean	0.288	0.14	3.85	2.29
95% credible interval	(0.141, 0.419)	(0, 1.44)	(1.94, 6.05)	(1, 4)

For comparison, Carvalho et al. [2018] estimated ATT to be 0.4478. Their approach does not assume simplex constraint.

Anti-corruption Campaign in China

For this data set, $p = 87, n = 35$.

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)

For comparison, Shi and Huang [2023] estimated ATT to be -0.0309 using a forward selection algorithm without simplex constraint.

Concluding remarks

- Sparse linear regression with constraints on the regression coefficients is used in applied domains, but its theoretical properties are not well understood.
- Bayesian formulation offers a simple solution to the debate over the use of simplex constraint.
- With $\tau = 0$, BVS-SS consistently selects the true variables in high-dimensional settings and achieves consistent coefficient estimation and prediction.
- Future directions:
  - Consistency theory for estimating τ.
  - MCMC methods that are more robust to highly correlated predictors.
  - Extensions that include time-invariant characteristics.
  - Robustness to model misspecification and heterogeneous errors.
  - Shrinkage effect of the simplex constraint?

Thank you!

References I

Alberto Abadie. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425, 2021. doi: 10.1257/jel.20201129.

Carlos Carvalho, Ricardo Masini, and Marcelo C Medeiros. Arco: An artificial counterfactual approach for high-dimensional panel time-series data. *Journal of econometrics*, 207(2): 352–380, 2018.

Nikolay Doudchenko and Guido W Imbens. Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Technical report, National Bureau of Economic Research, 2016.

Gyuhyeong Goh and Jisang Yu. Synthetic control method with convex hull restrictions: a bayesian maximum a posteriori approach. *The Econometrics Journal*, 25(1):215–232, 2022.

Cheng Hsiao, H Steve Ching, and Shui Ki Wan. A panel data approach for program evaluation: measuring the benefits of political and economic integration of hong kong with mainland china. *Journal of Applied Econometrics*, 27(5):705–740, 2012.

Gary King and Langche Zeng. The dangers of extreme counterfactuals. *Political analysis*, 14(2): 131–159, 2006.

Kathleen T Li. Statistical inference for average treatment effects estimated by synthetic control methods. *Journal of the American Statistical Association*, 115(532):2068–2083, 2020.

Ignacio Martinez and Jaume Vives-i-Bastida. Bayesian and frequentist inference for synthetic controls. *arXiv preprint arXiv:2206.01779*, 2022. URL https://arxiv.org/abs/2206.01779.

Zhentao Shi and Jingyi Huang. Forward-selected panel data approach for program evaluation. *Journal of Econometrics*, 234(2):512–535, 2023.

Yiqing Xu. Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76, 2017. doi: 10.1017/pan.2016.2.