Synthetic Controls...What are they really?

Graduate Statistics Club

Demetrios Papakostas

Arizona State University

Introduction

- Hi everyone! Thanks for making it
- My name is Demetri and I am co-president and 5th year PhD student in statistics. In store for today are hijinks, shenanigans, laughs, tears, and some difficult conversations!
- Please sign in
- Thanks to Andrew Herren and Richard Hahn, reading their discussions about synthetic controls was very helpful

Hellenic Mist

Terminology

- Y is an outcome, D is a treatment indicator (did a policy get enacted in that state or not etc.)
- Our focus is on the panel data setting
- From wikipedia, "In statistics and econometrics, panel data and longitudinal are both multi-dimensional data involving measurements over time. Panel data is a subset of longitudinal data where observations are for the same subjects each time."
- "Time series and cross-sectional data can be thought of as special cases of panel data that are in one dimension only (one panel member or individual for the former, one time point for the latter)."

Hellenic Mist

Difference in Difference

One of the seminal papers is Card and Krueger 1994, see [1], which studied the effect of minimum wage increase on employment by studying the impact on fast-food employment in neighboring New Jersey and Pennsylvania. New Jersey recently had raised their minimum wage, whereas Pennsylvania did not, so including Pennsylvania as a control and assuming parallel trends

Formalizing

We can formalize the model as

$$Y_{it} = \alpha + \beta_i D_i + \gamma t_i + \delta(D_i \cdot t_i) + \varepsilon_{it}$$

where α is a constant describing the mean of the outcome in the pre-intervention units across both groups, γ represents the difference in control group between pre and post time periods, β represents the difference in control and treated group in the pre-treatment period, and δ is the average treatment effect on the treated.

Working out the regression coefficients yields

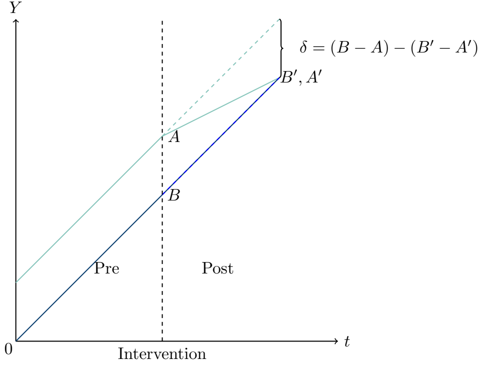
$$\delta = \left[E(Y \mid D = 1, \text{post}) - E(Y \mid D = 1, \text{pre})\right] - \left[E(Y \mid D = 0, \text{post}) - E(Y \mid D = 0, \text{pre})\right]$$

Difference in Difference

Assume parallel trends, i.e.

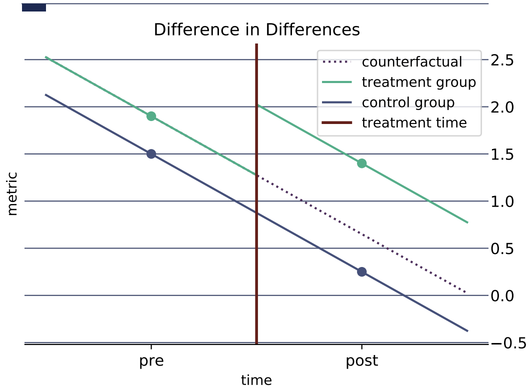
$$E\Big(Y_t^0 - Y_{t-1}^0 \mid D = 1\Big) = E\Big(Y_t^0 - Y_{t-1}^0 \mid D = 0\Big)$$

The units prior to treatment, in both the treatment and the control, follow the same trend. Strong and untestable assumption!



Easier to Visualize Picture

The difference in difference framework comes from a model, but is nested on a very strong assumption.



Hellenic Mist

Synthetic Controls: The idea

- Synthetic Controls are similar to difference in difference estimators, but the control group is estimated differently
- Have been around since 2003, with the three seminal papers being: [2],[3], [4]
- As some might say, it is "one of the most influential ideas of the last 15 years" in observational data treatment effect estimation
- Widely used in academic circles, but also industry and government.

Formalizing the Model

Hellenic Mist

In potential outcome framework:

$$Y_{it} = \begin{cases} Y_{it}^0 & \text{if } D_i = 0 \text{ or } t \leq T_0 \\ Y_{it}^1 & \text{if } D_i = 1 \text{ and } t > T_0 \end{cases}$$

In this framework, $Y_{it} = D_i Y_{it}^1 + (1 - D_i) Y_{it}^0$. Further, define:

$$\begin{pmatrix} Y_{11} & Y_{12} & \dots & Y_{1T_0} & Y_{1T} \\ Y_{21} & Y_{22} & \dots & Y_{2T_0} & Y_{2T} \\ \vdots & & & & \vdots \\ Y_{N1} & Y_{N2} & \dots & Y_{NT_0} & Y_{NT} \end{pmatrix} = \left(\begin{array}{cccc|c} X_{11} & X_{12} & \dots & X_{1T_0} & Y_1 \\ \hline X_{21} & X_{22} & \dots & X_{2T_0} & Y_2 \\ \vdots & & & & \vdots \\ X_{N1} & X_{N2} & \dots & X_{NT_0} & Y_N \end{array}\right)$$

Which we re-write as:

$$\left(\begin{array}{c|c} \boldsymbol{X}_{1\cdot} & \boldsymbol{Y}_1 \\ \hline \boldsymbol{X}_{0\cdot} & \boldsymbol{Y}_0 \end{array}\right)$$

Re-Write

- That notation is confusing and gets even more confusing when we talk about auxillary covariates, which we typically think of as being denoted by $\boldsymbol{X}$. So let's re-write it as:

$$\boldsymbol{X}_{0\cdot} \longrightarrow \boldsymbol{Y}_{\text{pre, control}}$$

$$\boldsymbol{X}_{1\cdot} \longrightarrow \boldsymbol{Y}_{\text{pre, treat}}$$

$$\boldsymbol{Y}_{0\cdot} \longrightarrow \boldsymbol{Y}^{0}_{\text{post, control}}$$

$$\boldsymbol{Y}_{1\cdot} \longrightarrow \boldsymbol{Y}^{1}_{\text{post, treat}}$$

- We observe $\boldsymbol{Y}^{1}_{\text{post, treat}}$ and want to estimate $\boldsymbol{Y}^{0}_{\text{post, treat}}$. The difference will be interpreted as the effect of the treatment.

- Following the syntax of [5], we write $\boldsymbol{Y}^{\text{obs}} = \left\{\boldsymbol{Y}^{0}_{\text{pre, treat}}, \boldsymbol{Y}^{0}_{\text{pre, control}}, \boldsymbol{Y}^{0}_{\text{post, control}}\right\}$ and $\boldsymbol{Y}^{\text{mis}} = \boldsymbol{Y}^{0}_{\text{post,treat}}$.

Creating the Synthetic Control

- Take weighted sum of the control units in the pre-intervention period, i.e.:

$$\hat{Y}^0_{1t} = w_1 + \sum_{i=2}^{N} w_i Y_{it}$$

for $t \in$ Pre-intervention. Impose convex hull restraint, i.e. for $\boldsymbol{w} = \{w_1, \ldots, w_N\}^T$

$$\boldsymbol{w} \in \mathcal{W}_{\text{conv}} = \left\{ \boldsymbol{w} : w_1 = 0, w_2 \geq, \ldots, w_N \geq 0, \text{ and } \sum_{i=2}^{N} w_i = 1 \right\}$$

- Minimize:

$$\underset{\boldsymbol{w}}{\text{argmin}} \left\| \boldsymbol{Y}^0_{\text{pre, treat}} - \boldsymbol{Y}^0_{\text{pre, control}} \boldsymbol{w} \right\|$$

That is:

$$\left(\boldsymbol{Y}^0_{\text{pre, treat}} \sum\nolimits_{i \in \text{control}}^{N} w_i \boldsymbol{Y}_{\text{pre, control}} \right)^T \boldsymbol{V} \left(\boldsymbol{Y}^0_{\text{pre, treat}} - \sum\nolimits_{i \in \text{control}}^{N} w_i \boldsymbol{Y}_{\text{pre, control}} \right)$$

subject to $\boldsymbol{w} \in \mathcal{W}_{\text{conv}}$ where $\boldsymbol{V}$ is a positive definite matrix.

What about other Covariates

- The SCM allows for other covariates to be “stapled on”
- Essentially, for each unit (state) we have T units, then we include the vector of auxillary covariates at the end. If we have p of these, we take the average over time and add these p covariates to the state outcome vector
- Therefore, we must average over the auxillary covariates over time.

Model Assumptions

For now, we assume we have just one treated unit

- There is no intercept term in the calculation of the weights
- The weights sum up to 1
- The weights are non-negative

Can think of DiD as having an intercept term but all weights are $1/N_{\text{control}}$ Further, it is necessitated in the model that the relation between the "donors" and the treated unit stays the same in the entire post-treatment period

Limitations of Time Stagnant Covariates

- As a motivating example, imagine Connecticut passed a tourism law in May 2021. We conduct a SCM analysis, with the weights of donor states being 25% New York, 20% Massachusetts, and 55% New Hampshire. However, an unusual bombogenesis hit Connecticut in October 2021, throwing a wrench in the tourism season, though it weaked substantially before hitting Massachusetts and New Hampshire. In this contrived example, the SCM would likely place all the fall in tourism on the law, when in fact it was not the case!

- Not a huge deal, but feels a little dis-satisfying that a method that is clearly time dependent cannot account for something like this, as it is certainly plausible

This is a little Odd

- So the SCM actually ASSUMES there exists a set of weights for the control potential outcome of the treated unit! That is, with weights being denoted by w_i for unit (or state) i, [6] pretty good overview...

$$\mu_0 = \sum_{i \in \text{donor}} w_i \mu_i$$

$$Y_{it} = \begin{cases} Y_{it}^0 = \mu_0^T \lambda_t + \varepsilon_{0t} & \text{when } t \leq T_0 \\ Y_{it}^1 = Y_{it}^0 + \beta_t & \text{when } t > T_0 \end{cases}$$

- Notice,the treated unit's "0" potential outcome then is defined entirely in terms of the other observed outcomes of the other units...weird

- Essentially, the outcomes are described by some unobserved variables

I Think I See what You're Getting At

- The issue isn't really the **estimator** (which much research focuses on), but really with the **estimand**, or lack thereof!
- The idea is slightly odd, because we need the other units outcomes (or a combination of them more specifically) to account for **all** the unobserved confounding
- How do we even simulate from a synthetic control model? We should at least know it works in the best-case data generating process, and then break it from there

Okay...so what's the point?

- In a lot of practical applications, the general model is probably not unreasonable
- The interpret-ability is cool and allows for nice applications
- Panel data is super common and of a lot interest, particularly with respect to policy decisions

Fine, I Buy it

For now...

- So perhaps the general idea is a little "silly" as some might say, how are synthetic controls used in practice?

- There is active in research in relaxing restrictions. Dropping the no intercept[7], the convexity restraint[8], performing regression adjustments ([9], [10], suggested in[8]), using a Gaussian Process to model the SCM [5], how to account for multiple treated units[11], how to account for time varying treatments, staggered adoption, how to pool estimates, and work on confidence intervals for estimates

Gaussian Processes, you say?

- [5] represent the SCM weighting in terms of a (multi unit) Gaussian process.
- This is a natural analogue and a nice way to think about the problem by explicitly modeling time and unit effects in the covariance, alongside some of the benefits that come with Gaussian Processes (such as uncertainty estimates)

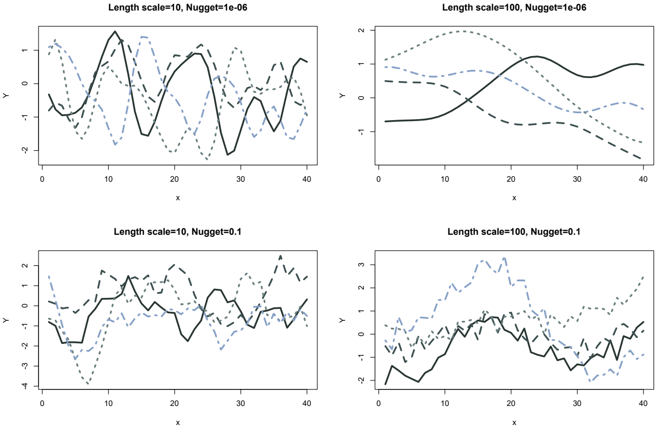
Wait slow down guy, whats a Gaussian Process

- A Gaussian Process is a set of random variables whose joint distribution is multivariate normal
- Gaussian Processes have several nice properties, such as ensuring interactions between every point and guarantee smoothness.
- However, the computation is intensive because of matrix inversions, and we have to specify the covariance kernel!
- BUT...we can make pretty plots

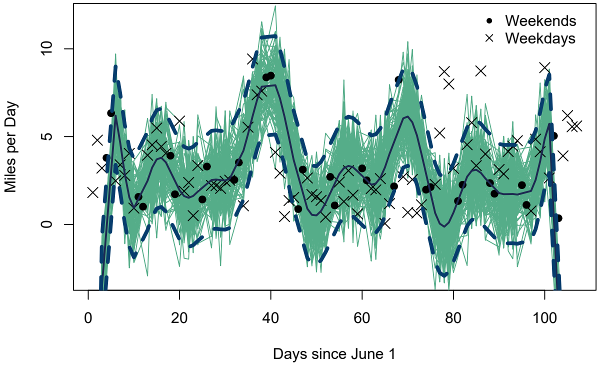
Hellenic Mist

Gaussian Process Prior

$$GP\left(\mathbf{0}, \sigma^2_{\text{sill}} \exp\left(\frac{-\mathrm{d}(t, t')}{\ell}\right) + \sigma^2_{\text{nugget}}\right)$$



June 1, 2022 to September 15, 2022. "Trained" on weekend data with squared exponential kernel with maximum likelihood estimation for sill, nugget, and range. Another nice pic



Back to Business

- [5] stipulate the following model:

$$Y_{it}^0 = f_{it} + \varepsilon_{it}$$
$$\boldsymbol{f} \sim GP(0, k)$$
$$\varepsilon_{it} \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

where $\boldsymbol{f} = (f_{11}, \ldots, f_{NT})$. The kernel is decomposed into separable unit and time kernels, meaning (with i denoting unit and t denoting time)

$$k\big((i,t),(i',t')\big) = k_{\text{unit}}(i,i') \times k_{\text{time}}(t,t') \quad (1)$$

- This assumption implies that the covariance across units is constant over time and time covariance is constant across units.
- Over-parameterized, use Bayesian model to reduce rank of unit covariance

More Details of GP SCM

With the separable kernel structure and reduced rank structure of the unit covariance, multi-unit (treated+control units) GP synthetic control analog can be written as:

$$\begin{pmatrix} \boldsymbol{Y}^{\text{mis}} \\ \boldsymbol{Y}^{\text{obs}} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{K}_{\text{mis}} & \boldsymbol{K}_{\text{mis, obs}} \\ \boldsymbol{K}_{\text{obs,mis}} & \boldsymbol{K}_{\text{obs}} \end{pmatrix} + \sigma^2 \boldsymbol{I}\right]$$

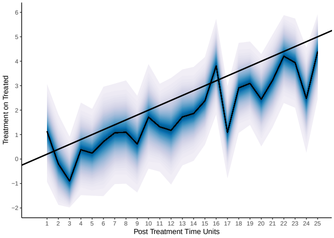
Since GP's are jointly multivariate normal, we note that the posterior mean and covariance are:

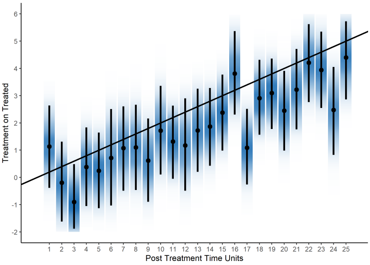
$$\boldsymbol{\mu} = \boldsymbol{K}_{\text{mis,obs}}\left(\boldsymbol{K}_{\text{obs}} + \sigma^2 \boldsymbol{I}\right)^{-1}(\boldsymbol{Y}^{\text{obs}})$$

$$\boldsymbol{\Sigma} = \left(\boldsymbol{K}_{\text{mis}} + \sigma^2 \boldsymbol{I}\right) - \boldsymbol{K}_{\text{mis,obs}}\left(\boldsymbol{K}_{\text{obs}} + \sigma^2 \boldsymbol{I}\right)^{-1}(\boldsymbol{K}_{\text{obs,mis}})$$

Flawed Attempt at a “2-stage GP-SCM”

- Run BART on all observations of DGP with time varying treatment effect aliasing covariates, then do a (horizontally) stacked GP regression on the residuals with all the control pre units to get $Y^0_{\text{post,treat}}$. Subtract from observed $Y^1_{\text{post,treat}}$
- Works fine for one treated unit...However need seperable time and unit kernels for multiple treated units!



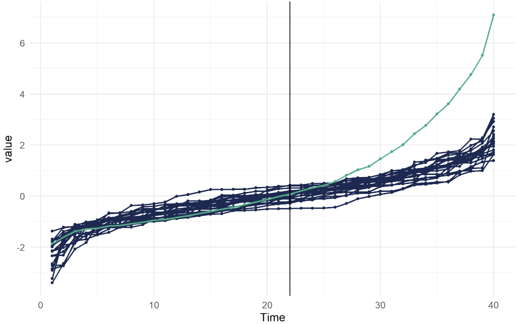


Simulation

Show False Treatment Effect

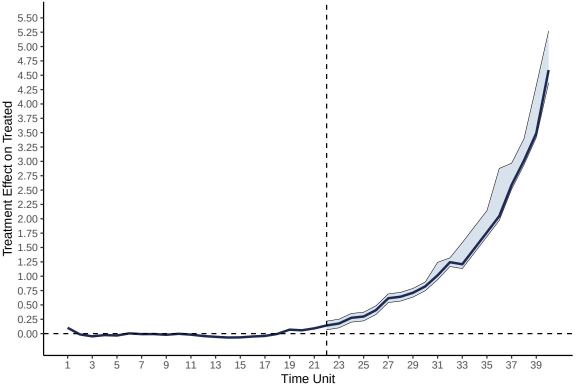
Scenario where we are entirely in the convex hull in the pre-treatment time setting, but are not in the post-treatment, even w/o a treatment effect. We have 25 control units, 1 treated, 40 time units, and a treatment that takes place at the 22nd time unit. The controls are all drawn standard normal. The treated unit is:

$$Y_{it}^{\text{control}} = \text{sort}\left[N(0,1)\right]$$
$$Y_{1t}^{\text{treatment}} = \left[(2.5/40) * t + 0.75\right] * \text{mean}(Y_{it}^{\text{control}})$$



Hellenic Mist

These data look like typical SCM data, but knowing the DGP we know it will be ill-suited. However, there isn't really any of this stress testing done because the method is defined in terms of the estimator



Applications

Applications

- My favorite! Where are these used!

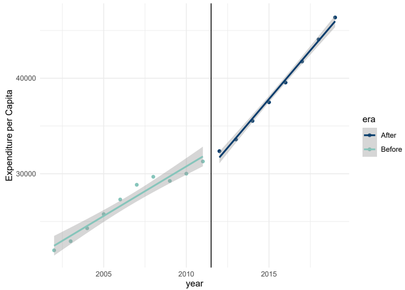
- Synthetic controls have been used to study right-to-carry laws [12], legalized prostitution [13], corporate political connections [14], Basque terrorism,[2] cigarette taxes[3], West Germany reunification [4], minimum wage increases[15], [16], Kansas tax cuts[7], firearm removal [5], and many more.

- We will study the effect of legalizing marijuana on personal expenditure per capita

- Also have GDP data, but we do not look at this. Don't look post 2020, as COVID would for sure violate the lack of time-based covariates aliasing the treatment effect! Especially given the politics of it!

Marijuana Legalization

- We study the effect of marijuana legalization in Colorado. Look at 2002 to 2019, and don't include other states that fully legalized between 2012 and 2019
- Also, we do not include any auxillary covariates, and do not drop the convex hull restraint (non-negativity and weights summing to 1) (for the time being)
- Is this even well motivated...ehhh?

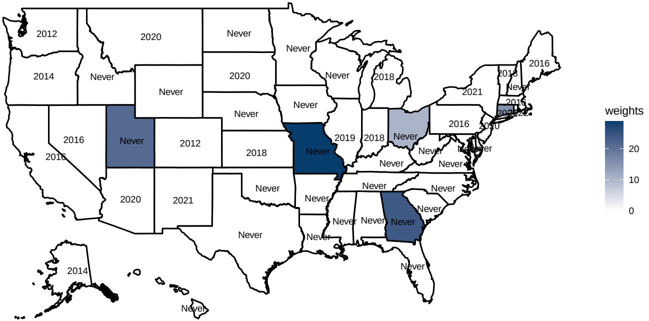


Hellenic Mist

Do these Weights Make Sense?

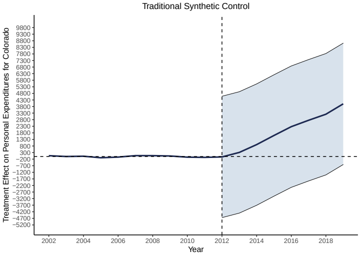
Using the traditional SCM estimator

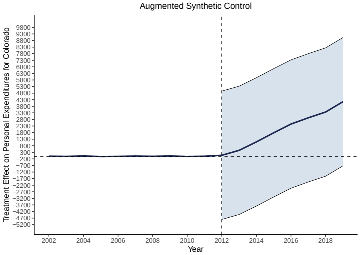
- We get Colorado as 10.4% Ohio, 21.4% Utah, 28.8% Missouri, 24.1% Georgia, 15.4% Connecticut, with ATT=$1,878. The ATT the sum of the effects over each year divided by the number of post-intervention time points.



Using the Augmented Synthetic Control Method

Now, we use the augmented synthetic control method and add the log of a states population as an auxiliary covariate. ATT=$2,035 Looks about the same!



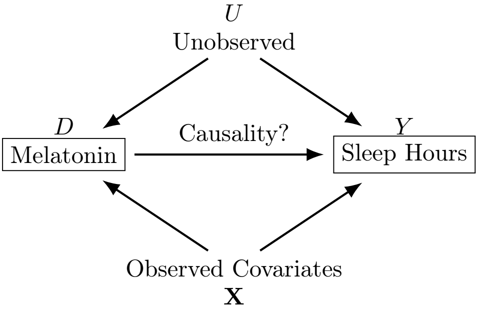


Goofing Off

Fun Example

- This is a FAKE example
- Imagine we measure Sam, Antonio, Andrew, and Demetri's sleep hours in 2021 and 2022. These are all friends and graduate students and avid GSC members
- In March of 2022, Sam decided to start taking melatonin to see if it helped her sleep totals. In a sense, this situation is a natural experiment.
- The data is presented here in this Flourish App
- So, how would we analyse this situation?

Say D indicates whether or not a treatment was taken, Y is sleep hours, and $\mathbf{X}$ indicates a matrix of observed covariates, and U indicates a set of unobserved variables

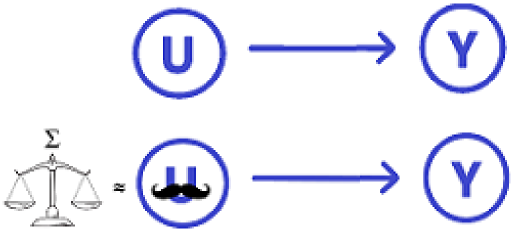


Controlling for $\mathbf{X}$, and hoping there isn't anything in U means we can estimate the treatment effect of melatonin on sleep!

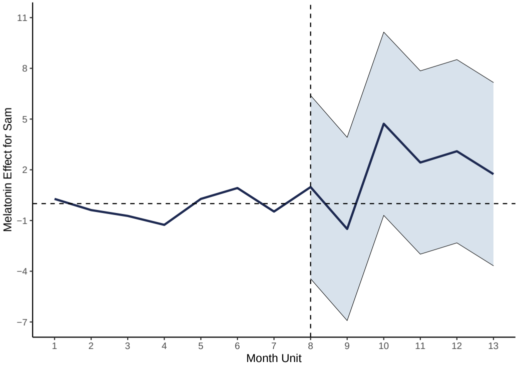
The Synthetic Control Approach

Well the picture was an effort...

- Because we only have 1 treated unit (Sam), this approach doesn't really work...so people tend to look at synthetic controls as the next step
- The synthetic control model stipulates Sam's sleep patterns depend on some underlying, unobserved variable (or set of variables).
- What's odd, however, is that we assume that this variable can be found by some weighted combination of Antonio, Andrew, and Demetri's sleep patterns

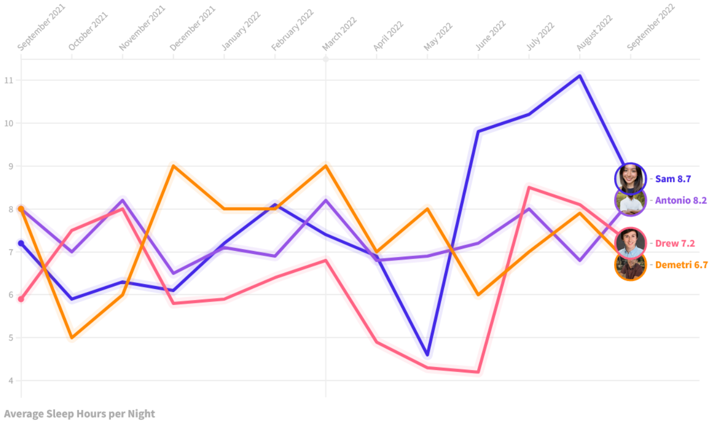


Pic 1



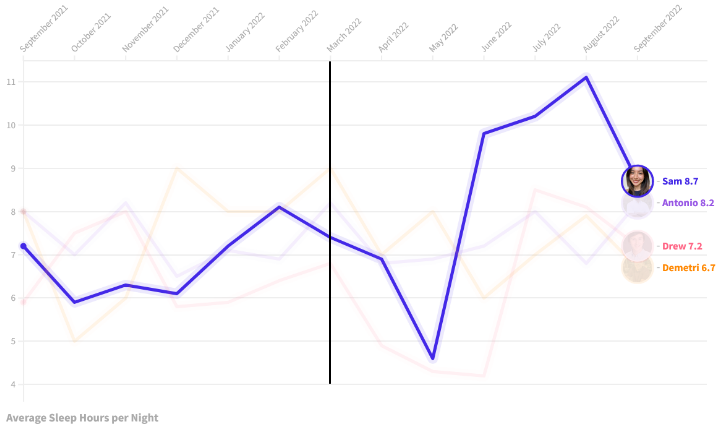
Hellenic Mist

Pic 2


September 2021
October 2021
November 2021
December 2021
January 2022
February 2022
March 2022
April 2022
May 2022
June 2022
July 2022
August 2022
September 2022
11
10
9
8
7
6
5
4
Sam 8.7
Antonio 8.2
Drew 7.2
Demetri 6.7
Average Sleep Hours per Night

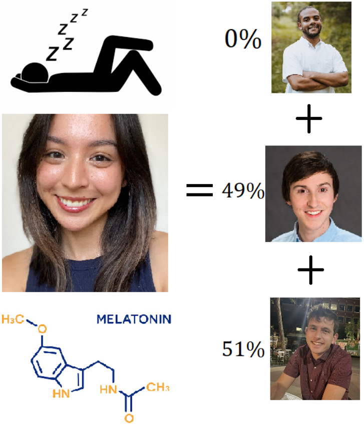

Hellenic Mist

Pic 3


September 2021
October 2021
November 2021
December 2021
January 2022
February 2022
March 2022
April 2022
May 2022
June 2022
July 2022
August 2022
September 2022
11
10
9
8
7
6
5
4
Sam 8.7
Antonio 8.2
Drew 7.2
Demetri 6.7
Average Sleep Hours per Night


Hellenic Mist

The perils of the Screenshot



Conclusions

- A nifty idea that's cool in practice, but a little underwhelming from a methodological standpoint
- Still active work on improving the estimator and expanding use cases

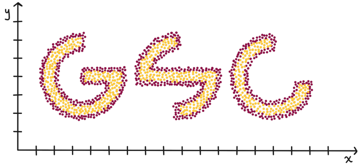


Figure: Thank you Sam!

Bibliography I

1. Card, D. & Krueger, A. Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review* **84,** 772–793 (1994).
2. Abadie, A. & Gardeazabal, J. The economic costs of conflict: A case study of the Basque country. *American Economic Review* **93,** 113–132 (2003).
3. Abadie, A., Diamond, A. & Hainmuller, J. Synthetic control methods for comparative case studies: Estimating the effect of Californias tobacco control program. *Journal of the American Statistical Association* **105,** 493–505 (2010).
4. Abadie, A., Diamond, A. & Hainmuller, J. Comparative politics and the synthetic control method. *American Journal of Political Science* **59,** 495–510 (2015).

Bibliography II

5. Ben-Michael, E., Arbour, D., Feller, A., Franks, A. & Raphael, S. Estimating the effects of a California gun control program with Multitask Gaussian Processes. *arXiv e-prints,* arXiv–2110 (2021).

6. Shi, X., Miao, W., Hu, M. & Tchetgen, E. T. Theory for identification and inference with synthetic controls: a proximal causal inference framework. *arXiv preprint arXiv:2108.13935.* <https://arxiv.org/abs/2108.13935> (2021).

7. Ben-Michael, E., Feller, A. & Rothstein, J. The augmented synthetic control method. *Journal of the American Statistical Association* **116,** 1789–1803 (2021).

8. Doudchenko, N. & Imbens, G. W. *Balancing, regression, difference-in-differences and synthetic control methods: A synthesis* tech. rep. (National Bureau of Economic Research, 2016).

Hellenic Mist

Bibliography III

9. Powell, D. Imperfect synthetic controls: Did the Massachusetts health care reform save lives? *Available at SSRN 3192710* (2018).

10. Broderson, K. H. *et al.* Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics* **9,** 247–274 (2014).

11. Kreif, N. *et al.* Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health economics* **25,** 1514–1528 (2016).

12. Donohue, J. J., Aneja, A. & Weber, K. D. Right-to-carry laws and violent crime: A comprehensive assessment using panel data and a state-level synthetic control analysis. *Journal of Empirical Legal Studies* **16,** 198–247 (2019).

Bibliography IV

13. Cunningham, S. & Shah, M. Decriminalizing indoor prostitution: Implications for sexual violence and public health. *The Review of Economic Studies* **85,** 1683–1715 (2018).

14. Acemoglu, D., Johnson, S., Kermani, A., Kwak, J. & Mitton, T. The value of connections in turbulent times: Evidence from the United States. *Journal of Financial Economics* **121,** 368–391 (2016).

15. Powell, D. Synthetic control estimation beyond case studies: Does the minimum wage reduce employment? *RAND Working Paper WR-1142. Santa Monica, CA: RAND Corporation* (2017).

16. Dube, A. & Zipperer, B. Pooling multiple case studies using synthetic controls. An application to minimum wage policies. *IXA Discussion Paper* (2015).

Hellenic Mist

Thank you