Topic 6: Factor Models

ECON 5783 — University of Arkansas

Prof. Kyle Butts

December 2024

Imputation Estimator review

The last set of slides, we introduced an "imputation estimator" for panel data treatment effects:

1. Estimate model for $y_{it}(\infty)$ using observations with $d_{it} = 0$ and get fitted values for full sample, $\hat{y}_{it}(\infty)$
2. Regress $y_{it} - \hat{y}_{it}(\infty)$ on d_{it} or event-study indicators to estimate treatment effects
    - $\rightarrow$ Estimating the overall effect, ATT, or dynamic effects of being treated for ℓ periods, ATT^ℓ respectively

Imputation Estimator review

The last set of slides, we introduced an "imputation estimator" for panel data treatment effects:

1. Estimate model for $y_{it}(\infty)$ using observations with $d_{it} = 0$ and get fitted values for full sample, $\hat{y}_{it}(\infty)$
2. Regress $y_{it} - \hat{y}_{it}(\infty)$ on d_{it} or event-study indicators to estimate treatment effects
   - $\rightarrow$ Estimating the overall effect, ATT, or dynamic effects of being treated for ℓ periods, ATT^ℓ respectively

This topic will extend this procedure to "factor models" that will allow more general trending behavior

Factor Models

Synthetic Control

More General Factor Model Imputation

Factor Model

Untreated potential outcomes are given by a factor model:

$$y_{it}(0) = \sum_{r=1}^{p} f_{t,r} * \gamma_{i,r} + u_{it}$$

- $f_{t,r}$ is the r-th **factor** (macroeconomic shock) at time t.
- $\gamma_{i,r}$ is unit i's **factor loading** (exposure) to the r-th factor.

Factor Model Example

$$y_{it}(0) = \sum_{r=1}^{p} f_{t,r} * \gamma_{i,r} + u_{it}$$

If we are thinking about housing prices, y_{it}:

- γ_i are characteristics of neighborhood / house
- f_t are demand shocks in each period

Factor Model Example

$$y_{it}(0) = \sum_{r=1}^{p} f_{t,r} * \gamma_{i,r} + u_{it}$$

If we are thinking about wages, y_{it}:

- γ_i are worker's latent skills (e.g. computer skills)
- f_t reflect changing firm's demand for skills

Factor Model Example

$$y_{it}(0) = \sum_{r=1}^{p} f_{t,r} * \gamma_{i,r} + u_{it}$$

If we are thinking about county employment, y_{it}:

- γ_i are characteristics of a county (e.g. their manufacturing share)
- f_t reflect national shocks to the economy (e.g. the "China shock")

Two-way Fixed Effect vs. Factor Model

The factor model is a generalization of the TWFE model. If $\boldsymbol{f}_t = (\lambda_t, 1)'$ and $\boldsymbol{\gamma}_i = (1, \mu_i)'$, then our factor model becomes the TWFE model:

$$y_{it}(0) = \boldsymbol{f}_t' \gamma_i + u_{it} = \lambda_t + \mu_i + u_{it}$$

We can add unit and/or time fixed-effects as ‘known’ factors if we want

Factor Model and Parallel Trends

Say you have a single treatment timing and two periods. Let D_i be out treated group indicator. Then

$$\begin{aligned}\mathbb{E}[\Delta y_i \mid D_i = d] &= \mathbb{E}[y_{i1} - y_{i0} \mid D_i = d] \\ &= \Delta \boldsymbol{f}\, \mathbb{E}[\boldsymbol{\gamma}_i \mid D_i = d]\end{aligned}$$

Under a factor model, the average change in y_{it} for group $D_i = d$ is the change in factor shocks $\boldsymbol{f}$ times the average exposure to those shocks

Factor Model and Parallel Trends

$$\mathbb{E}[\Delta y_i \mid D_i = d] = \Delta \boldsymbol{f}\, \mathbb{E}[\boldsymbol{\gamma}_i \mid D_i = d]$$

Say the treated group has higher exposure to a shock than the control group

- $\implies$ the trends differ by treatment status

Factor Model and Parallel Trends

$$\mathbb{E}[\Delta y_i \mid D_i = d] = \Delta \boldsymbol{f} \, \mathbb{E}[\boldsymbol{\gamma}_i \mid D_i = d]$$

Say the treated group has higher exposure to a shock than the control group

- $\implies$ the trends differ by treatment status

That is, a factor model allows for “non-parallel trends” based on difference in exposures to shocks

Example

$$y_{it}(0) = \sum_{r=1}^{p} f_{t,r} * \gamma_{i,r} + u_{it}$$

Say we are thinking about neighborhood housing prices, y_{nt}. We are interested in some treatment D_n, e.g. access to subways.

- Say γ_n is the walk-ability of the neighborhood
- f_t are demand shocks for walkable neighborhoods

If new subways are built in more walkable neighborhoods, then we do not believe parallel trends hold in this setting

Factor Loadings are ‘non-observed X_i’

In the worker’s wage example, we might suspect that something like computer-skill might be unobservable and have a time-varying impact.

That is both X_i and β_t are unobserved!

- If the job-training program attracts people with more computer-skills (e.g. excel workshop), then PTs does not hold.
- We can not ‘compare two individuals with the same value of X_i’ since we do not observe them.

Factor Models

$$Y_{it} = \sum_{r=1}^{\rho} \lambda_{i,r} f_{t,r} + \varepsilon_{it}$$

Let's give an example using county-level aggregate employment.

- λ_i might consist of (i) manufacturing share and (ii) share of college-educated
- In each period, shocks to the national economy change manufacturing demand and (ii) technological change drives returns to college degree

Factor Models

$$Y_{it} = \sum_{r=1}^{\rho} \lambda_{i,r} f_{t,r} + \varepsilon_{it}$$

Let's give an example using county-level aggregate employment.

- λ_i might consist of (i) manufacturing share and (ii) share of college-educated
- In each period, shocks to the national economy change manufacturing demand and (ii) technological change drives returns to college degree

We might have ideas of what are the primary characteristics are, but we might not have data on it. If you do, then we are back in X_i land.

Imputation and (Linear) Factor Models

We have a more-general model for $Y_{it}(0)$ now that allows some forms of non-parallel-trends:

$$Y_{it}(0) = \mu_i + \eta_t + \sum_{r=1}^{\rho} \lambda_{i,r} f_{t,r} + \varepsilon_{it}$$

Can we estimate this and use our imputation procedure:

- The short-answer is yes. There is a bunch of different approaches but they're not as simple as TWFE.
- The factor model is much more data-hungry than fixed effects, usually requiring both a large number of units *and* a large number of time-periods.

'Extensions' of the Synthetic Control Model

Generalized Imputation Estimator

Xu (2017) and Gobillon and Magnac (2016) both discuss using an imputation estimator under a *factor model*

Intuition is to estimate $\lambda_{i,r}$ and $f_{t,r}$ using untreated/not-yet-treated observations and then impute:

$$\hat{Y}_{it}(0) = \hat{\mu}_i + \hat{\lambda}_t + \sum_{r=1}^{\rho} \hat{\lambda}_{i,r} \hat{f}_{t,r}$$

- We estimate the non-parallel trending via the factor model and then subtract it off
  - $\rightarrow$ Can implement this using the package `gsynth`

Factor Model Imputation

$$Y_{it}(0) = \mu_i + \eta_t + \sum_{r=1}^{\rho} \lambda_{i,r} f_{t,r} + \varepsilon_{it}$$

Unlike imputation of the two-way fixed effect model, this approach is very data hungry:

- Requires both a large number of time-periods and units

Intuitively, you need a long number of time-periods to estimate a unit's $\lambda_{i,r}$. You need a large number of units to estimate a time-period's $f_{t,r}$

Problem with large-T

Estimation with a factor model is more robust than standard DID, but rely on having access to many years of data. In a lot of applied work, the data is just not available

Problem with large-T

Estimation with a factor model is more robust than standard DID, but rely on having access to many years of data. In a lot of applied work, the data is just not available

But I think there is a more subtle problem at play with this assumption:

- Data from many years ago might not be very useful at understanding the underlying confounders at play in this economy
- Imagine saying "I use data from 1960 to inform me which counties would be a good control group for housing prices in 2000". A lot has happened since then !!!!!!

Factor Models

Synthetic Control

More General Factor Model Imputation

Synthetic Control

The standard synthetic control method considers a single treated unit (country, state, firm, etc.). We will let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iT})'$ be the vector of outcomes for unit i.

- The treated unit is $i = 0$. We have N control units

Synthetic Control

The standard synthetic control method considers a single treated unit (country, state, firm, etc.). We will let $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{iT})'$ be the vector of outcomes for unit i.

- The treated unit is $i = 0$. We have N control units

Synthetic control imputes $\boldsymbol{Y}_0(0)$ for the treated unit using a weighted average of the control units:

$$\hat{\boldsymbol{Y}}_0(0) = \sum_{i=1}^{N} w_i \boldsymbol{Y}_i$$

- Take a part of this state, a part of that state, and then average them together into a 'synthetic control' unit
- In most cases, we use convex weights $1 \geq w_i \geq 0$.

Choosing weights

We want our synthetic control unit to do a good job at approximating the pathway of outcomes for the treated unit: $\boldsymbol{Y}_0(0)$.

Choosing weights

We want our synthetic control unit to do a good job at approximating the pathway of outcomes for the treated unit: $\boldsymbol{Y}_0(0)$.

We only observe $Y_{0t}(0)$ for the treated unit up until period T_0 which is the period prior to treatment.

- Synthetic control should try to match the treated unit's outcome path during the pre-period and *HOPEFULLY* that will mean the synthetic control will do a good job in the post-period. It's a leap of faith

Identifying assumption

The identifying assumption can be said as: “The synthetic control unit approximates the counterfactual trend that the treated unit would be on had they not been treated.”

- This is like a parallel trends assumption between the treated unit and the synthetic control unit

Choosing weights

More formally, the weights are selected by trying to minimize the following:

$$\underset{\{w_i\}}{\text{argmin}} \sum_{t=1}^{T_0} \left(Y_{0t} - \sum_{i=1}^{N} w_i Y_{it} \right)^2$$

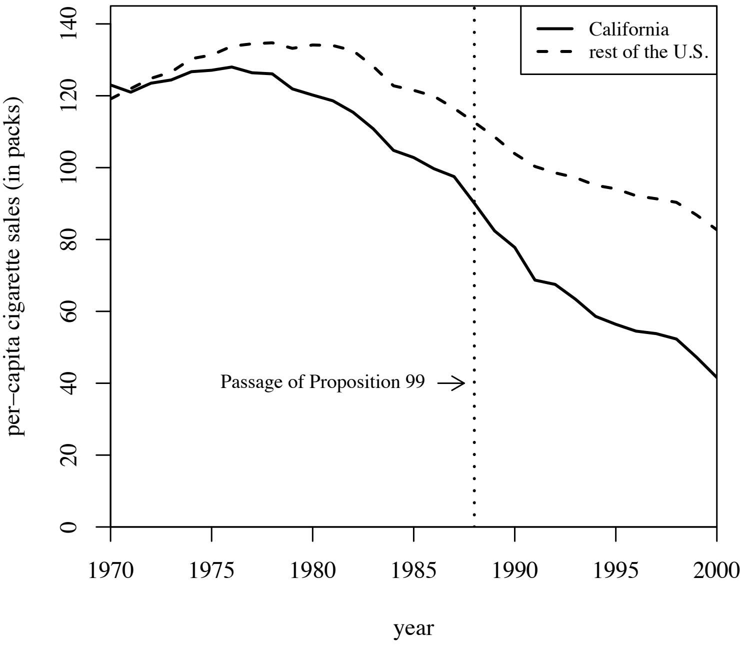
- Minimizing the *pre-treatment* sum of squared prediction errors between treated unit and the synthetic control unit.
- With convex weights, $1 \geq w_i \geq 0$, we add these as a constrained optimization problem
  - $\rightarrow$ Convex weights avoid having '-0.2 Minnesota'

Example: California's Proposition 99

In 1988, California first passed comprehensive tobacco control legislation:

- increased cigarette tax by 25 cents/pack
- earmarked tax revenues to health and anti-smoking budgets
- funded anti-smoking media campaigns
- spurred clean-air ordinances throughout the state
- produced more than $100 million per year in anti-tobacco projects

Other states that subsequently passed control programs are excluded from donor pool of controls (AK, AZ, FL, HI, MA, MD, MI, NJ, OR, WA, DC)



California
rest of the U.S.
Passage of Proposition 99
per–capita cigarette sales (in packs)
year
0
20
40
60
80
100
120
140
1970
1975
1980
1985
1990
1995
2000

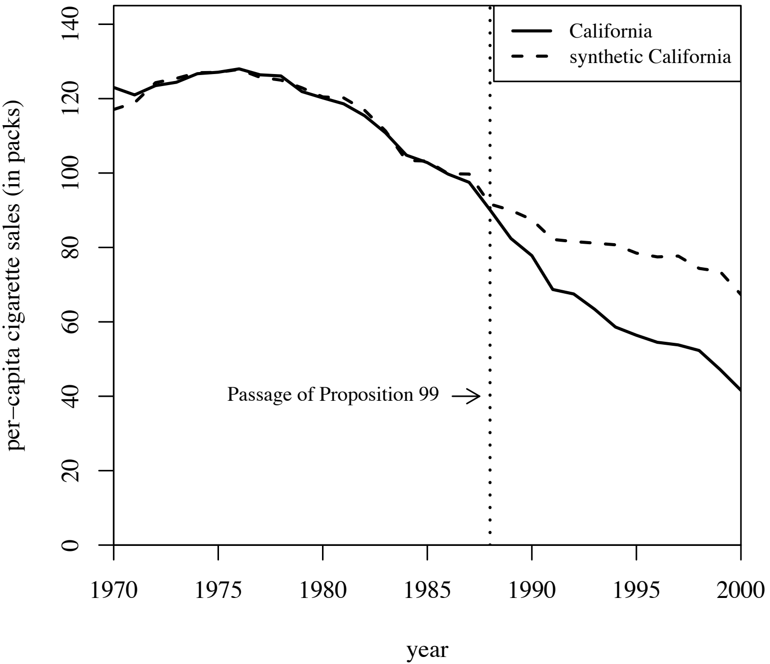

California
synthetic California
Passage of Proposition 99
per−capita cigarette sales (in packs)
year
0
20
40
60
80
100
120
140
1970
1975
1980
1985
1990
1995
2000

Table 2. State weights in the synthetic California

State	Weight	State	Weight
Alabama	0	Montana	0.199
Alaska	–	Nebraska	0
Arizona	–	Nevada	0.234
Arkansas	0	New Hampshire	0
Colorado	0.164	New Jersey	–
Connecticut	0.069	New Mexico	0
Delaware	0	New York	–
District of Columbia	–	North Carolina	0
Florida	–	North Dakota	0
Georgia	0	Ohio	0
Hawaii	–	Oklahoma	0
Idaho	0	Oregon	–
Illinois	0	Pennsylvania	0
Indiana	0	Rhode Island	0
Iowa	0	South Carolina	0
Kansas	0	South Dakota	0
Kentucky	0	Tennessee	0
Louisiana	0	Texas	0
Maine	0	Utah	0.334
Maryland	–	Vermont	0
Massachusetts	–	Virginia	0
Michigan	–	Washington	–
Minnesota	0	West Virginia	0
Mississippi	0	Wisconsin	0
Missouri	0	Wyoming	0

Identifying assumption

The identifying assumption can be said as: “The synthetic control unit approximates the counterfactual trend that the treated unit would be on had they not been treated.”

We select our weights to make the pre-treatment trends very similar between the treated and synthetic control

- Unlike the ‘pre-trends’ test in DID, we are sort of ‘cheating’ here in that we are making the synthetic control do that.

Identifying assumption

The identifying assumption can be said as: "The synthetic control unit approximates the counterfactual trend that the treated unit would be on had they not been treated."

We select our weights to make the pre-treatment trends very similar between the treated and synthetic control

- Unlike the 'pre-trends' test in DID, we are sort of 'cheating' here in that we are making the synthetic control do that.

But, the idea is that as the number of pre-periods grow, you can only do a good job at matching the treated y in *all* pre-periods if your synthetic control "*looks similar* to treated unit"

Beware of Overfitting

Intuitively, synthetic control is ‘believable’ when the synthetic control unit does a great job at approximating the outcome in the per-period

- We think that the synthetic control must be picking up on underlying economic structure in order to co-move with the treated state.

E.g. if you see a set of runners whose times go up and go down during the same races, you might think they train together.

- You wouldn’t think that chance alone made them have the same ups and downs.

Beware of Overfitting

The number one concern you should have when reading or writing a paper that uses a synthetic control method is that of overfitting.

- If I have 1000 control units and 4 pre-periods, I can probably well approximate Y_{0t} in the pre-periods by just random chance.

Synthetic Control and Factor Model

Recall our factor model:

- There are a set of characteristics that the units have and in each period a set of macroeconomic shocks that change the marginal effect of those characteristics
- If we could observe these characteristics, we would want to match on them or use them in conditional PTs

Synthetic Control and Factor Model

Recall our factor model:

- There are a set of characteristics that the units have and in each period a set of macroeconomic shocks that change the marginal effect of those characteristics
- If we could observe these characteristics, we would want to match on them or use them in conditional PTs

Synthetic control, under conditions we will discuss, will create a synthetic control unit that has the same average characteristics as the treated unit!

- Since they are subject to the shocks the same amount, we think their outcome evolutions will match.

Synthetic Control and Factor Model

Okay, so I’ve pained a very rosy picture of synthetic control. When it works well, we fix the problem of non-parallel trends and we are able to impute the untreated potential outcome well.

However, I want to make clear that this method is **not a panacea**. The original paper is very general and makes it hard to know when it works and when it does not.

- More recent advancements all base discussion on a factor model for outcomes
- Hollingsworth and Wing working paper has great discussion

Bias of Synthetic Control

The original Abadie, Diamond and Hainmueller paper derive the bias of synthetic control when outcomes are generated by a linear factor-model.

The bias arises from **over-fitting on noise**. It is more common to over-fit on data when:

- There are fewer pre-treatment periods T_0
- There are many control units
- The 'convex hull' assumption is unlikely to ˙old

‘Convex Hull’ Assumption

When constraining the weights to be convex ($0 \leq w_i \leq 1$), the synthetic control assumption requires the ‘convex hull’ assumption:

- This is basically an assumption that says ‘we can approximate $\boldsymbol{Y}_0$ using a convex weighted average of control $\boldsymbol{Y}_i$’

‘Convex Hull’ Assumption

When constraining the weights to be convex ($0 \leq w_i \leq 1$), the synthetic control assumption requires the ‘convex hull’ assumption:

- This is basically an assumption that says ‘we can approximate $\boldsymbol{Y}_0$ using a convex weighted average of control $\boldsymbol{Y}_i$’

With a factor model, we can make this assumption a lot clearer:

- The ‘convex hull’ assumption is equivalent to the assumption that the treated unit’s ‘factor loadings’ are a convex average of the other units’ ‘factor loadings’
  - $\rightarrow$ That is, the treated unit can not have an extreme value in any of the $\lambda_{i,r}$ (e.g. huge manufacturing share)

Inference in Synthetic Control

Inference in the classical synthetic control setting is really difficult

- Only one treated unit; you’re not averaging over units so the estimate is subject to random shocks

Inference in Synthetic Control

Inference in the classical synthetic control setting is really difficult

- Only one treated unit; you're not averaging over units so the estimate is subject to random shocks

Two forms of randomization inference are typically used:

- Randomly shuffle treatment to control units and re-estimate synthetic control; want the treated unit to look more extreme than the placebo estimates. The so-called 'spaghetti plot'
- Randomly shuffle time-periods around for treated unit and re-estimate synthetic control

Implementing Synthetic Control

My recommendation for synthetic control is `scpi` package (on R, Stata, and Python)

- Covers all the basic method and include inference methods
- Journal of Statistical Software paper is super readable: https://nppackages.github.io/references/Cattaneo-Feng-Palomba-Titiunik_2024_JSS.pdf

Factor Models

Synthetic Control

More General Factor Model Imputation

‘Extensions’ of the Synthetic Control Model

Synthetic Control with Lasso/Ridge Penalty

People seem to like when the synthetic control is made up of a few units

- Makes the control unit more ‘interpretable’

‘Extensions’ of the Synthetic Control Model

Synthetic Control with Lasso/Ridge Penalty

Can modify the weights optimization problem to penalize weights being too large:

$$\underset{\{w_i\}}{\text{argmin}} \sum_{t=1}^{T_0} \left(Y_{0t} - \sum_{i=1}^{N} w_i Y_{it} \right)^2 + \lambda \|w\|_k$$

- Add a term that punishes when weights are non-zero; λ is a ‘tuning-parameter’ to chose how much to punish
- Can add convex-weights constraint
- Lasso is $k = 1$; Ridge is $k = 2$

‘Extensions’ of the Synthetic Control Model

Augmented Synthetic Control

Augmented control is a method to help with imperfect pre-treatment fit by estimating a bias and subtracting it off.

- Looks similar to regression adjustment. We estimate the trend using covariates and then use that model to bias correct.

'Extensions' of the Synthetic Control Model

Augmented Synthetic Control

1. Calculate the standard synthetic control method weights (or with lasso)
2. Estimate a model of $m_t(Z_i) = \mathbb{E}[Y_{it} \mid Z_i]$ using untreated units where Z_i is pre-treatment characteristics (lagged Y or covariates)

Form synthetic control estimate as

$$\left(Y_{0t} - \sum_{i=1}^{N} w_i Y_{it}\right) - \underbrace{\left(m_t(Z_0) - \sum_{i=1}^{N} w_i m_t(Z_i)\right)}_{\text{`bias correction'}}$$