Kyle Butts

Papers Open Source Teaching Blog

Factor Models and Synthetic Control

When I first saw the synthetic control method, I had a reall understanding when it would work and when it would not. Sure a lot of pre-periods. But what else? “Good pre-treatment fit thinking about a research question, how can I decide if I’ll pre-treatment fit a-priori? Certainly, looking at the result isn’t a good idea (paging Jon Roth).

I sort of stumbled onto factor models in a completely differ context. But, after I learnt this model, I began seeing it p and again in papers about synthetic control. They made such difference in my understanding that I wanted to share what I learned.

First, I’ll share what a factor model is and then I’ll circl how it connects to synthetic controls. As an important discl synthetic control estimator is consistent under a wide varie generating processes. However, the factor model is a leading it is pedagogically useful.

Factor Models

The simplest factor model is given by

$$y_{it} = \boldsymbol{f}_t' \boldsymbol{\gamma}_i + u_{it},$$

where $\boldsymbol{f}_t$ is a $p \times 1$ vector of *unobservable* “factors” and $\boldsymbol{\gamma}_i$ i
vector of *unobservable* “factor loadings”. Notice that this k
looks like a time fixed-effect and a unit fixed-effect, but
multiplied together.

We can view the factors $\boldsymbol{f}_t$ as macroeconomic shocks with fact
$\boldsymbol{\gamma}_i$ denoting a unit’s exposure to the shocks. The product of
depics how a unit is affected by the macroeconomic shock. Fo
imagine y is GDP. A factor-loading could be a counties’ shar
manufacturing employment. The factor could be a common natio
to manufacturing employment (e.g. import competition or othe
to the national economy). The product of the two would be th
the national shocks on a given counties’ GDP.

Another possibility lets the $\boldsymbol{\gamma}_i$ represent time-invariant
characteristics with a marginal effect on the outcome $\boldsymbol{f}_t$ tha
over time. For example, imagine y is wages. For example, ima
worker’s wage in a given year. A factor-loading could be a w
education, $\boldsymbol{\gamma}_i$. The factor, $\boldsymbol{f}_t$, could be a common national sh
returns to education (e.g. changes in the national college p
The product of the two would be the effect of the national s
given worker’s wage.

However, these are just examples of what could be in there.
of a factor model is that we don’t need to be able to actual
$\boldsymbol{\gamma}_i$ or $\boldsymbol{f}_t$ to control for them. Just like with unit and time f
effects, you might have ideas on what goes in them, but you
actually need to *measure them*.

Note that the two-way fixed effect is actually an example of
model! Take $\boldsymbol{\gamma}_i = (1, \mu_i)$ and $\boldsymbol{f}_t = (\lambda_t, 1)$; then the model become
$\mu_i + u_{it}$. However, the factor model allows for units to be on
differential trends (based on their exposure to national sho

Since the factor model nests the standard TWFE model, why no them in difference-in-differences models?? Well, of course, free lunch and this model is no exception. The problem is th long panels in order to consistently estimate the factors an factor-loadings seperately (you can't within-transform them you can't with TWFE model). It turns out, that one powerful think about synthetic control is through factor models.

Synthetic Control

The basic idea of synthetic control is to identify a set of units that "well-approximate" the outcome for the treated un pre-treatment periods. The synthetic control estimator is th weighted average of the control units. The original syntheti paper shows that when the outcome model is a factor-model, t synthetic control estimator's bias goes to 0 when the number periods are large (to be clear, the synthetic control estima consistent under other data-generating processes).

Under a factor model, the synthetic control estimate for the unit's untreated potential outcome is formed as the weighted the control unit's outcomes $\hat{y}_{it}^{sc} = \sum_j w_j y_{jt}$, where w_j are foun synthetic control procedure. We label the treated unit as $i =$ Ignoring the error term, we can write this as

$$\hat{y}_{it}^{sc} = \sum_j w_j y_{jt} = \sum_j w_j \boldsymbol{f}'_t \boldsymbol{\gamma}_j = \boldsymbol{f}'_t \sum_j w_j \boldsymbol{\gamma}_j = \boldsymbol{f}'_t \sum_j \boldsymbol{\gamma}^{sc}.$$

The last term shows the insight, the synthetic control unit' is generated affected by the same factors and with factor-lo given as the weighted average of the control units' factor-l

When does the synthetic control do a good job? Well, when th

when does the synthetic control do a good job? Well, when th
control units’ factor-loadings look similar to the treated u
factor-loadings, $\boldsymbol{\gamma}^{sc} = \boldsymbol{\gamma}_0$! If we satisfy this condition, ther
predicted counterfactual outcome in the post-periods is good
Intuitively, the synthetic control and the treated unit are
the same factor shocks and they have the same “exposure” to
shocks.

I find this interpretation really useful. As a researcher, I
thinking of these conditions *a-priori*. First, I think of fac
are important in my setting (e.g. in the GDP example, I thin
important confounding shock may be shocks to industries over
Second, I think of whether my treated units’ factor-loadings
their industrial mix) can be well approximated by the donor
would fail in our ongoing example if my treated county is an
terms of their industrial mix.

I’m not sure I’ve seen this recommended anywhere, but it str
a good check to compute weighted averages of important chara
that you think might be confounders with treatment selection
by baseline covariates X) and see if the synthetic control
approximate the treated units weights, $\sum w_j X_j \approx X_0$.

The last connection I’ll make is about the importance of lon
for estimating synthetic control. In the above work, notice
the error term. That is because, *in long panels*, the error t
zero. But this isn’t true in short panels because the synthe
procedure ends up matching on noise instead of matching on f
loadings. That makes the $\boldsymbol{\gamma}^{sc} \neq \boldsymbol{\gamma}_0$.

Examples of factor-models in synthetic control paper

To see how prevelant the use of factor models are, I’ll end examples of their appearance throughout the econometric lite

1. Ferman and Pinto - Synthetic Control with Imperfect Pre-T Fit

The authors consider what happens when the pre-treatment fit imperfect deriving results and a new estimator under a facto

> ”We consider the properties of the SC and related estimato linear factor model setting, when the pre-treatment fit is We show that, in this framework, the SC estimator is gener if treatment assignment is correlated with the unobserved heterogeneity, and that such bias does not converge to zer the number of pre-treatment periods is large.”

2. Gobillon and Magnac - Regional Policy Evaluation: Interac Effects and Synthetic Controls and Xu - Generalized Synthet Method

Propose an imputation treatment effect estimator that direct estimates the factor-loadings and the factors themselves. Th similar to the imputation estimator from Borusyak et. al. (2 Gardner (2021).

3. Ben-Michael, Feller, and Rothstein - The Augmented Synthe Control Method

Considers how to use regularization on the weights to preven fitting in shorter panels. This helps better approximate the treated unit’s factor-loadings in a factor-model.

4. Athey, Bayati, Doudchenko, Imbens and Khosravi - Matrix c methods for causal panel data models

They impute untreated potential outcomes by estimating a low factor model using nuclear-norm regularization to prevent ov

Short panels

All of these papers mentioned *all* rely on long-panels to pro identify the underlying factor structure. Over the last two been working on a research agenda that proposes estimators t valid in short-panels.

The key insight my coauthors and I have found is that treatm estimators only require consistent estimation of $\boldsymbol{f}_t$ and *not* of $\boldsymbol{\gamma}_i$ so long as you have many treated units. Using this ins have "unlocked" a large econometric literature that shows wh be estimated consistently in short-panels and derived a way t these estimators into an imputation estimator.

For example, consistently estimates of $\boldsymbol{f}_t$ can be estimated b instrumental variables (quasi-long differencing) or by using that are affected by the same factor structure (common corre effects). Both of these can then be plugged into an estimato out a nice event-study graph. Here's an example from our pap

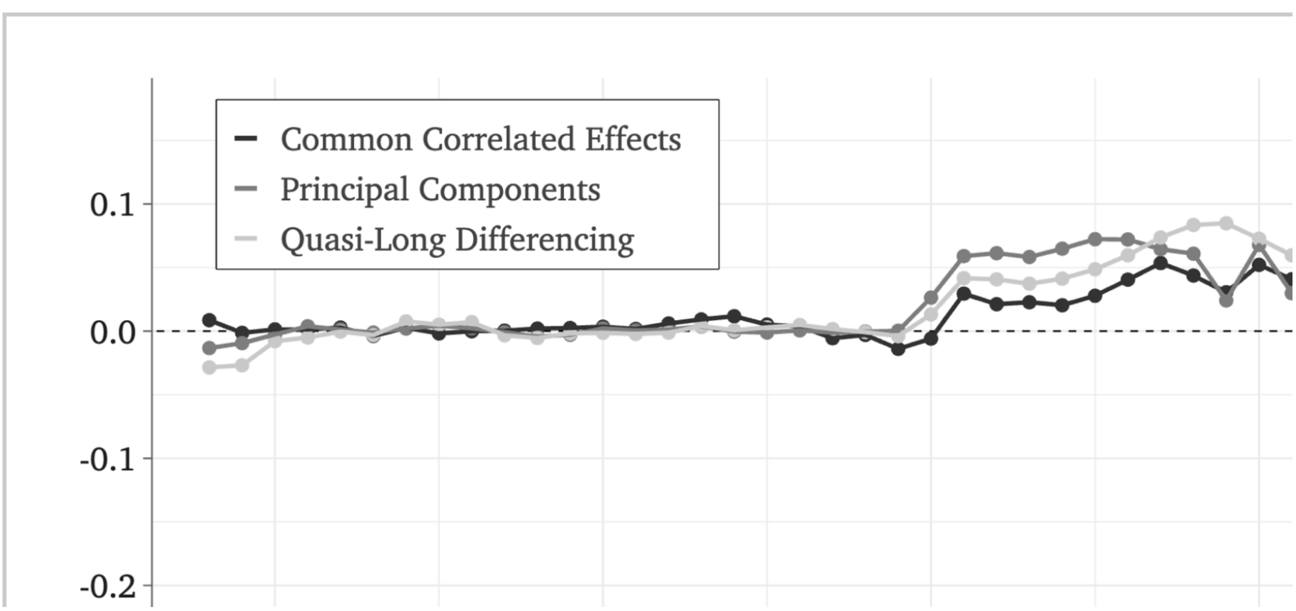

Figure: Example of plug-in estimators for the effect of Wa
local retail employment