Topics in Applied Econometrics for Public Policy

TA Session 8

Sergi Quintana

Barcelona School of Economics

June 20, 2023

Plan for today

- Correct PS
- Order of convergence
- Bias of the kernel

Group Fixed Effects

- The paper from Bonnhome and Manresa 2015 provide a methodology to estimate time varying group effects.
- They tackle the incidental parameter problem, since fixed effects are usally poorly estimated.
- Allows for clustered time patterns of unobserved heterogeneity that are common within groups of individuals. The group-specific time patterns and individual group membership are left unrestricted, and are estimated from the data.

Grouped Fixed Effects. Estimation

In the model of the PS solves:

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\varrho}) = arg\min \sum_{i=1}^{N} \sum_{t=1}^{T} (h_{it} - \alpha(a_{it}, \eta_{gi}) - \beta(a_{it}) m_{it} + x_{it}'\gamma)^2$$

And it considers all possible values of parameters and all possible groups. The algorithm to solve it uses the following two equations:

$$\hat{g}_i(\alpha, \beta, \theta) = arg\min \sum_{t=1}^{T} (h_{it} - \alpha(a_{it}, \eta_g) - \beta(a_{it}) m_{it} + x_{it}'\gamma)^2$$

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = arg\min \sum_{i=1}^{N} \sum_{t=1}^{T} (h_{it} - \alpha(a_{it}, \eta_{\hat{g}_i(\alpha,\beta,\theta)}) - \beta(a_{it}) m_{it} + x_{it}'\gamma)^2$$

AKM Model

The idea of the model is to quantify the contributions of workers and firms to earnings inequality.

$$y_{it} = X_{it}'\beta + \alpha_i + \psi_{j(i,t)} + \epsilon_{it}$$

Model allows for high-wage workers may be more likely to move to higher-paying firms than low-wage workers.
In this model, we focus on the contributions of firm effects and sorting in the following variance decomposition:

$$Var(Y_{it} - X_{it}'\beta) = Var(\alpha_i) + Var(\psi_{j(i,t)} + 2Cov(\alpha_i, \psi_{j(i,t)}) + Var(\epsilon_{it})$$

Studies find that sorting is weak and negative.

AKM - Limited Mobility Bias

- Limited mobility bias is due to the large number of firm-specific parameters that are solely identified from workers who move across firms.
- The bias arises from an insufficient number of job movers in the firm.
- As a result of limited mobility bias, the FE variance of firm effects tends to be overstated. In turn, the co-variance between worker and firm effects tends to be negatively biased,since worker effects and firm effects enter model additively.
- Once bias is accounted for,firm effects dispersion matters less for earnings inequality, and worker sorting becomes always positive and typically strong.

AKM - Bias Correction Methods

- Andrews (2008).
  - Provide a closed form expression for the bias.
  - Under the assumption that errors are homoskedastic.
  - Computationally difficult to implement
- Kline, Saggio, and Sølvsten (2020)
  - Generalize the bias to heretorkedastic errors.
  - Uses the leave-one-out set that remains connected when any (i, t) obsevation has been taken out.
  - Also computationally difficult to implement.

AKM - Bias Correction Methods

Correlated Random-Effects (CRE).

- Developed in Bonhomme, Lamadon and Manresa (2019).
- The idea is the following, to reduce the incidental parameter problem we can cluster firms.
- In their set up, they model firms as discrete fixed effects and workers as (discrete or continuous) random effects correlated with firm classes. This alleviates incidental parameters, specially in short panels.
- Finally, they model the means and covariances of worker and firm effects.
- The CRE model still has many fewer parameters than the AKM fixed effects model.

AKM - Bias Correction Methods

- They propose a static and a dynamic model, and both allow for interactive effects between firms and workers.

CRE implementation follows a two-step approach:

- **Step 1**: Classification step, they group firms into classes using k-means clustering. Clustering can be based on observables, mobility patterns, etc.
- **Step 2**: Estimation step, they estimate the model by maximum likelihood, conditional on the estimated firm classes.

Synthetic Controls

Initially developed by Abadie and Gardeazabal (2003).

- The idea is to generate a fictitious conterfactual using untreated observations.
- Synthetic controls models optimally choose a set of weights which when applied to a group of corresponding units produce an optimally estimated counterfactual to the unit that received the treatment.
- An advantage is that weights make explicit what each unit is contributing to the conterfactual. Regressions also weight the data, but we don't see the weights.

Synthetic Controls

Synthetic control weights minimize:

$$\sum_{m=1}^{k} v_m \left(X_{1m} - \sum_{j=2}^{J+1} w_j X_{jm} \right)^2$$

where ω_j are the synthetic control weights and v_m are the weights that we put to each covariate. Usually, they are chosen by minimizing the mean squared prediction error.

$$\sum_{t=1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^*(V) Y_{jt} \right)^2$$

Stata Code

Order of Magnitudes

Definition 1 *(Convergence in probability to zero)* X_n *converges in probability to zero, written* $X_n = o_p(1)$ *or* $X_n \xrightarrow{p} 0$, *if for every* $\varepsilon > 0$

$$P(|X_n| > \varepsilon) \to 0 \text{ as } n \to \infty$$

Definition 2 *(Boundedness in probability) The sequence* $\{X_n\}$ *is bounded in probability, denoted as* $X_n = O_p(1)$, *if for every* $\varepsilon > 0$ *there exists* $\delta(\varepsilon) \in (0, \infty)$ *such that*

$$P(|X_n| > \delta(\varepsilon)) < \varepsilon \text{ for all } n$$

Clearly if $X_t = o_p(1)$, then $X_t = O_p(1)$.

Order of Magnitudes

Definition 1: The sequence $\{X_n\}$ of real numbers is said to be at most of order n^k and is denoted by

$$X_n = O(n^k), \quad \text{if } \frac{X_n}{n^k} \to c$$

as $n \to \infty$ for some constant $c > 0$. Further if $\{X_n\}$ is a sequence of r.v.s then

$$X_n = O_p(n^k) \quad \text{or} \quad O_{a.s.}(n^k)$$

if, as $n \to \infty$,

$$\frac{X_n}{n^k} - c_n \to 0 \quad \text{in prob.} \quad \text{or} \quad \text{a.s.,}$$

Order of Magnitudes

Definition 2: The sequence $\{X_n\}$ of real numbers is said to be of smaller order than n^k and is denoted by

$$X_n = o(n^k), \quad \text{if } \frac{X_n}{n^k} \to 0$$

as $n \to \infty$. Further if $\{X_n\}$ is stochastic then

$$X_n = o_p(n^k) \quad \text{or} \quad o_{\text{a.s.}}(n^k)$$

if

$$\frac{X_n}{n^k} \to 0 \quad \text{in prob. or a.s.}$$

Order of Magnitudes

- We use Big-O and small-o to establish bounds in the rate of convergence.
- $f = O(g)$ means that f's asymptotic growth is no faster than g's.
- $f = o(g)$ means that f's asymptotic growth is strictly slower than g's".
- It's like $\leq$ versus $<$.

Order of Magnitudes

As an example consider a nonstochastic sequence

$$\{X_n\} = \frac{1}{n+4}. \tag{A.57}$$

It is easy to verify that for $k = -1$ in definition 1,

$$\frac{X_n}{n^{-1}} = \frac{n}{n+4} \to 1 \tag{A.58}$$

as $n \to \infty$. Thus $X_n = O(n^{-1}) = O(1/n)$.

The sequence $X_n = 1/(n+4)$ is also $o(1) = o(n^0)$. This is because, for $k = 0$, using Definition 2, $X_n/n^0 = X_n = 1/n + 4 \to 0$ as $n \to \infty$.

Order of Magnitudes

Some interesting results about probability orders

Definition 3 *(Algebra of probability orders)*

i) $X_n = o_p(a_n)$ *iff* $a_n^{-1} X_n = o_p(1)$,

ii) $X_n = O_p(a_n)$ *iff* $a_n^{-1} X_n = O_p(1)$.

- Exercise: True of false. $\hat{\beta_n}$ is the OLS estimator of β under the usual hypothesis.

a) $\hat{\beta} = o_p(1)$

b) $\hat{\beta} - \beta = o_p(1)$

c) $\hat{\beta} - \beta = O_p(n^{.5})$

d) $\hat{\beta} - \beta = o_p(n^{-.25})$

e) $\hat{\beta} - \beta = o_p(n^{-.5})$

f) $\hat{\beta} - \beta = O_p(n^{-.5})$

Bias Kernel

The Kernel estimate is:

$$\widehat{f}(x_0) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x_i - x_0}{h}\right)$$

And it is known to be biased:

$$b(x_0) = \mathrm{E}[\widehat{f}(x_0)] - f(x_0) = \frac{1}{2} h^2 f''(x_0) \int z^2 K(z) dz.$$

The asymptotic distribution is:

$$\sqrt{Nh}(\widehat{f}(x_0) - f(x_0) - b(x_0)) \xrightarrow{d} \mathcal{N}\left[0,\ f(x_0) \int K(z)^2 dz\right]$$

Bias Kernel

There are some important things to notice from the previous formulas.

- The bias converges at $O_p(h^2)$.
- Since the optimal bandwidth is $O_p(N^{-0.2})$ the bias converges at $O_p(N^{-0.4})$.
- Also, we have that $\sqrt{Nh}$ is $O_p(N^{0.4})$.
- Therefore, once we re-scale $\sqrt{Nh}b(x_o)$ we get $O_p(1)$. Which implies that the bias of the asyptotic distribution does not disappear.