Data Recipe #107:
Panel Regression with Two Levels of Time Aggregation

Jesse M. Shapiro

The Problem. For each of the 50 states, you have data on the month in which the state adopted a law change. You have data on pollution measured every three months. Your professor asks you to estimate a linear model in which the growth in pollution in a given month is greater after than before the law change. But you don't measure pollution in every month. Should you try to impute pollution between the measurement dates? If so, what kind of imputation method should you use? Or should you just "smudge" the data by assuming all the law changes happen at quarter's end?

The Solution. You can estimate the same linear regression model on quarterly data that you can estimate on monthly data. The reason is that the average of a linear function is a linear function of averages. Therefore, under standard assumptions, a panel regression model can be estimated at various levels of time aggregation.

The Math. Pollution y_{it} in state i and month t evolves according to

$$\begin{aligned} \Delta y_{it} &= \alpha_i + \delta_t + \beta z_{it} + \varepsilon_{it} \\ 0 &= \mathrm{E}(\varepsilon_{it} | \alpha_i, \delta_t, \{z_{it}\}_{\forall t}) \end{aligned}$$

where α_i is a state fixed effect, δ_t is a month fixed effect, $z_{it} \in \{0,1\}$ is an indicator for whether the law is in effect, and ε_{it} is an unobserved error. The Δ is a first-difference operator. The second equation is the strict exogeneity assumption common in panel fixed effects models.

Now imagine that we only observe pollution at the end of every quarter q. Let $\overline{x}_{iq} = \frac{1}{3}\sum_{t \in q} x_{it}$ be the quarterly average of some variable x_{it} and observe that according to the above assumptions

$$\begin{aligned} \overline{\Delta y}_{iq} &= \alpha_i + \overline{\delta}_q + \beta \overline{z}_{iq} + \overline{\varepsilon}_{iq} \\ 0 &= \mathrm{E}\left(\overline{\varepsilon}_{iq} | \alpha_i, \overline{\delta}_q, \left\{\overline{z}_{iq}\right\}_{\forall q}\right) \end{aligned}$$

where the second equation follows from the strict exogeneity assumption given above.[1]

Because $\overline{\Delta y}_{iq} = \frac{1}{3}\Delta y_{iq}$, the average growth in pollution over a quarter can be calculated given pollution at the beginning and end of the quarter. And of course $\overline{z}_{iq} \in \left\{0, \frac{1}{3}, \frac{2}{3}, 1\right\}$ can be calculated as the share of months in the quarter in which the law is in place. The latter has nothing to do with z_{it} being an indicator–it will work for any type of z_{it}, you just need to calculate its average.

The Example. Gentzkow et al. (2015) wish to estimate the effect of whether the state government is controlled by the Republican party z_{it} on the evolution of the share y_{it} of the state's newspapers that are

[1]

$$\begin{aligned} \mathrm{E}\left(\varepsilon_{it} | \alpha_i, \overline{\delta}_q, \{z_{it}\}_{\forall t}\right) &= \mathrm{E}\left(\mathrm{E}\left(\varepsilon_{it} | \alpha_i, \overline{\delta}_q, \{z_{it}\}_{\forall t}, \delta_t\right)\right) \\ &= \mathrm{E}(\mathrm{E}(\varepsilon_{it} | \alpha_i, \delta_t, \{z_{it}\}_{\forall t})) \\ &= 0 \end{aligned}$$

and so on for $\overline{z}_{iq}$. Since the expectation is a linear operator we therefore have

$$0 = \mathrm{E}\left(\overline{\varepsilon}_{iq} | \mu_i, \delta_q, \left\{\overline{z}_{iq}\right\}_{\forall q}\right).$$

Republican. They only collect data on the newspaper market during presidential election years. But state elections often happen between those years. They estimate their model in four-year averages.

The Recipe. This is trivial to do:

```
use law_change.dta, clear
collapse (mean) post_law_change, by(state quarter)
merge state quarter 1:1 using pollution.dta, keep(pollution)
xtset state quarter
gen growth_in_pollution = (1/3)*D.pollution
reg growth_in_pollution post_law_change i.state i.quarter
```

Hope it comes out right!