

Financial Econometrics – Cross Section & Panel Data

SFI Léman PhD Program, Fall 2025 – Andreas Fuster, EPFL

Problem Set 1 – “Solutions”

- (a) Download data – straightforward. For those using Stata, it is useful to “compress” the data (and re-save) first, as that saves a lot of memory.
- (b) Merge Bank-HHI data – straightforward. Note that Call Rep data is a panel by RSSDID – quarter, not cert-quarter, and there were 8 duplicate observations at the cert-quarter level so need to drop those first (or manually fix them, but I simply dropped them).
- (c) Download Fed Funds target rate data from FRED – straightforward, though then wasn’t sure whether to use quarterly average or end (chose the latter and that seems to be what they did).
- (d) Regressions in Table VIII: in first attempt, did not come close to their numbers. Realized that partly had to do with variable construction (which I then fixed, hopefully) but also that could be due to outliers and them dropping certain observations. After looking at summary stats of the outcomes, realized that likely have to winsorize them in order to avoid excessive influence. The other likely filter is that given the panel structure, one may want to drop banks that strongly change characteristics from one quarter to the next due to M&A etc. – so then remove from the regression sample bank-quarter observations where assets change by more than 25% from one quarter to the next (since this is unlikely to happen “organically”). Also, assume they multiply the log changes by 100 so did that as well. And voilà, come pretty close, both in terms of coefficients and R2 (with exception of dep. spread, where my R2 is much higher), even though my #observations is still higher.
 - a. Note: you can find their replication package at <https://doi.org/10.7910/DVN/KHNXYJ> -- limited info on construction of the regression dataset though. (But they definitely drop obs. based on changes in assets.)
- (e) Bank FE remove permanent differences across banks; quarter FE remove the effects of aggregate shocks at a point in time. These are quite uncontroversial I would say. Removing the quarter FE changes the results little (if anything, makes them larger). The bank FE matter more but I think it’s uncontroversial that one would put them in. It’s less clear to me what the justification for the post-08 X bank FE is – of course one can argue that “the world has changed” but it’s still a bit arbitrary. And these matter quantitatively for some of the coefficients (though mostly wouldn’t overturn the story, except for log(loans), at least in my code – haven’t checked in their replication package).
- (f) Using only data pre-2008 leaves some of the coefficients similar (e.g. on deposit spreads), but changes the sign on log(loans) for me. This happens even though post-08 is only about one quarter of observations, and there was no FF policy rate move over 2009-13 (only very minor fluctuations in the effective fed funds rate). So the results on loans seem the least robust of all. (Though one would also want to check this based on their own code and data in the QJE replication package, linked above.)

Restricting to top 10 or top 25% reduces statistical power substantially (especially top 10%) even though directionally many effects survive. But this illustrates that in US banking research the ‘small’ banks (which jointly only account for a small share of total assets in the system) are often quite important.