

Financial Econometrics – Cross Section & Panel Data

Swiss Finance Institute PhD Program, Fall 2024 – Andreas Fuster, EPFL

Problem Set 1 – “Solutions”

Preliminary: I use CPI data downloaded from FRED (St Louis Fed) to turn assets into real assets. I retain banks if they have a chartertype of 200 or 300-340 (this gets me quite close to the Kashyap et al. results, although not exactly; with only chartertype 200 there were too few banks). For those banks that are not part of a larger BHC, I create an artificial BHC ID before summing things at the BHC-quarter level. In my output files, “samp1” is the original 1992-1996 sample, while “samp2” is the 1997-2013 sample period.

1. See the code for variable construction. My equivalent of Table II (for samp1) comes quite close to the original in the paper, except that I have slightly more banks (9,347 instead of their 9,262) and that the DEPRAT distribution for large banks is shifted by about 0.02-0.03 (while all other values are within 0.01 of the ones in the paper).
2. See Tables_III_IV_samp1.csv. The columns that should be most comparable to those in the papers are the ones with Controls = Y (but keep in mind that we don't have one of the controls that they used, namely the Federal Reserve district dummies – although it's not clear these should matter much). Overall, I would say that the results are reasonably close to theirs, and vary in similar ways across the groups, although there are some differences, esp. in the smaller subsamples of large and medium banks (which suggests that they probably used a slightly different ranking). Adding controls usually brings the results a bit closer to zero, which raises the question of whether with additional controls (that could capture differences in business models etc. across banks) the results might weaken further.
3. I did this with 'binsreg' and controlling also for log(real assets) and the different loan ratios, as in the regressions. Results are shown below. I would say that the relationships overall look fairly linear and are reasonably precisely estimated, although for LIQRAT and especially SECRAT the relationship flattens for DEPRAT above 0.3 or so. For COMRAT, which I think is the most interesting dependent variable, the relationship is remarkably linear.
The advantage relative to a regular scatter plot should be obvious – in a regular scatter with this many observations, it's very difficult to see any relationship. Also, binsreg allows you to control for other variables, which you can't do with a regular scatter plot.
4. The justification for time-averaging is given on pp. 55/56, and it is to rule out “spurious” correlations. I think this is indeed defensible, although one could alternatively lag the deposit ratio by one or several quarters to take care of the issue they are worried about (and indeed, on p. 56 they describe that they did this). In our case, we do not lag DEPRAT and just study the contemporaneous correlations with different fixed effects.

I then run 6 versions of each regression – with and without controls X {no FE, time FE only, time FE and bank FE}. Results are in Tables_III_IV_quarterly_samp1.csv. For the banks outside the top 600, the quarter FE essentially don't matter, which is not very surprising given that the main sample period isn't very long (5 years) and most banks are there throughout and probably don't adjust these ratios all that much. Adding bank FEs has more substantial effects, reducing the magnitude of the coefficients by 20-50%. This also isn't very surprising, since adding the bank fixed effects means we are only using within-bank variation, and there may be more of a 'noise' component across quarters. As discussed in the lecture, it is useful to assess how much variation is left after the FE are added. One way to do this is with the “sumhdfe” command provided by deHaan (and discussed in the paper mentioned in the lecture). The key bit here, for the example of COMRAT and DEPRAT, is:

Panel D: Residual variation after partialling-out

Variable	N*	Std. Dev.			R2 by fixed effect		R2 Overall
		Pooled	Within*	Ratio (%)	dateq	bhcid	
comrat	156,374	.0710804	.0304671	42.86	0.010	0.805	0.816
deprat	156,374	.1132507	.0293985	25.96	0.007	0.924	0.933

This shows that the within quarter & bank variation in DEPRAT is only about 1/4 as large as the total variation. (This is still relatively ok though – in many settings the decrease is much more extreme.) Another way of looking at this is to study the distribution of residuals in these variables after fixed effects are added. E.g. for DEPRAT, the table below shows descriptive statistics for the ‘raw’ DEPRAT (column 1), the residuals of DEPRAT after absorbing quarter FE (column 2), and the residuals of DEPRAT after absorbing quarter FE and bank FE (column 3). Notice how the SD and the difference between the 5th and 95th percentiles are much smaller in the last column but the min and max are further apart than in the raw data (where the maximum distance is 1), and are very far from the 5th and 95th percentile. This means that there are a few BHCs for which DEPRAT changes substantially over time, and those will be much more important in the estimation with bank FEs than without (they have “higher leverage”).

Stats	deprat	res_qFE	res_~bFE
Mean	0.287	-0.000	-0.000
SD	0.113	0.113	0.029
Min	0.000	-0.304	-0.619
p5	0.114	-0.173	-0.041
p50	0.280	-0.007	-0.001
p95	0.468	0.180	0.043
Max	1.000	0.732	0.867

In the sample of the top 600 banks, effects for LIQRAT and SECRAT are generally insignificant whether or not FEs are added. For COMRAT, effects are quite strongly significant without bank FE, and actually even survive with bank FE (although only when controls are also added).

5. See Tables_III_IV_samp2.csv and Tables_III_IV_quarterly_samp2.csv. The results for LIQRAT and SECRAT are reasonably close to the earlier sample period, but the result for COMRAT in the time-averaged data very much depends on whether controls are included or not. However, note that time-averaging over such a long period (17 years) seems like a pretty bad idea, given that many banks are not in the sample over the entire period. Also, you may have noticed that the loan ratios that are used as controls are no longer in the data from 2010, so that the averages of those ratios may be a bit misleading.

Looking at the quarterly versions with different fixed effects, the relationship between COMRAT and DEPRAT seems fairly robust for the banks outside the top 600, while for the top 600 the relationship seems weak or absent. There may just not be enough statistical power for this subsample though, esp. since the way I defined it is based on average assets over the entire period and therefore in any quarter this subsample only contains around 350 banks

6. (...)
7. I would say that for their main sample, results appear robust. For the out-of-sample exercise, the relationship between DEPRAT and COMRAT (which I consider the most interesting result in the paper overall) appears fragile. But this may not be surprising given that the second sample period is very

long and contains the financial crisis period and its aftermath. One would have to dig a bit deeper to see whether/when the relationship changes over the course of this period.

Binned scatter plots (item 3):



