

Research Professional Data Task Questions

Please answer the following questions and provide any relevant code. You can write out the code if you do not have access to the relevant software. You have 24 hours from the time of opening this document to answer all the questions. For the programming questions in Part 1, please provide us with your code (either as a separate code file or as a text document).

Part 1: Programming questions:

0. Instructions:

- a. Please use SAS and Stata as indicated. If you do not have access to either SAS or Stata or prefer to answer the questions below using a different package, please provide us with your code for that language (e.g., R or Matlab). Please attempt all the questions.**
- b. Reading in raw data in a delimited format (e.g., .csv) is preferred to reading in data as .sas7bdat (SAS data format) or .dta (Stata data format).**
- c. There are two datasets: `bnk_vars.csv` (a subset of Call Reports quarterly data, which contains bank information) and `sod.csv` (a subset of the summary of deposits data, which includes levels of banks' total deposits at the branch level). The datasets are from 1995 to 2004. Below is the description of the variables:**
 - i. `Sod.csv`**
 1. `RSSDID` is a bank identifier.
 2. `BRNUM` is a branch identifier within each bank (`BRNUM = 0` refers to a given bank's headquarters).
 3. `DEPSUMBR` is the total level of deposits in each branch.
 - ii. `Bnk_vars.csv`**
 1. `RSSDID` is a bank identifier.
 2. `DATE` is the end date for each calendar quarter.
 3. `TOTLOAN_NET` stands for Net Total Loans.
 4. `ROA` stands for the Return on Assets (Net Income divided by average total assets).
 5. `CAPITAL_RATIO` stands for Total Equity as a proportion of Total Assets.

1. (Please use SAS to answer this question.)
 - a. Read in the data from the file “sod.csv,” find duplicates and drop them.
 - b. Fill in the missing values in **DEPSUMBR** with the average of **DEPSUMBR** across branches within the bank in the missing year.
 - c. Add up **DEPSUMBR** at the bank level by year and rename it **TOT_DEPOSITS**.
 - d. Combine the dataset with *bnk_vars.csv* based on **RSSDID** and **YEAR**. You will have to read in *bnk_vars.csv* file, deal with duplicates, and create a new variable **YEAR** to merge the two datasets.

2. (Please use Stata to answer these questions.)
 - a. Read in the data from the file “bnk_vars.csv.” Check for any data issues (e.g., formatting or duplicates); generate a *Stata* date (use %td as the calendar date format) and a *Stata* year-quarter (convert the date variable to a year-quarter; use %tq as the quarter format); create a binary variable **REGULATION** by assigning one for the period from 1999Q3 to 2004Q4, and zero otherwise.
 - b. Remove the whole time series for banks with the capital ratio lower than 4.5% at least once in their history.
 - c. Please create a new variable **TREATMENT** by assigning one to banks that had an event (**EVENT** = 1) at any point in their existence. Otherwise, assign zero.
 - d. Run a regression to see if the regulation you defined in question (a) has an impact on treatment banks’ Capital Ratio by comparing the treatment group (**TREATMENT** = 1) with the control group (**TREATMENT** = 0).
 - e. Add year-quarter fixed effects in the regression above.
 - f. Output your two regression results into a formatted Excel table that could be used in an academic article or presentation.

3. Using the same dataset (*bnk_vars.csv*) please answer the following questions:
 - a. Describe the distribution of the **ROA** variable over time (by year-quarter) and check if there are any outliers.
 - b. Create a new variable **MONTH** using the **DATE** variable.
 - c. Create a change variable for **TOTLOAN_NET**. Use the following definition (you will have to check for duplicates before creating the variable):

$$CH_TOTLOAN_{it} = \frac{TOTLOAN_NET_{it} - TOTLOAN_NET_{i,t-1}}{TOTLOAN_NET_{i,t-1}},$$

where $i = RSSDID, t = year_quarter$

- d.** Compute the means of the new variable *CH_TOTLOAN*, by quarter (not by year-quarter). You can use the variable *MONTH* you created in (b), which accounts for the end month of each quarter.

Part 2: Econometrics/ Statistics questions:

4. If your dependent variable is binary (which only takes the value of 1 or 0), could you suggest what model(s) you might like to use other than the OLS and describe the model(s)? What are the strengths and weaknesses of each?
5. You have a sample of all US banks from 1985 to 2005. You want to estimate the effect of a change in the regulatory regime in 1989 on bank risk-taking. How would you design such a study (i.e., what would your model be)? Do you need a control sample? If yes, how would you construct one? What is the counterfactual and how would you go about establishing this counterfactual? Would you test the market impact of this policy and if so, how would you do that?