Econometrics 2

Empirical Assignment 2

An Introduction to Data Management in Stata

Due: Tuesday 1400/01/24, 23:00

Please send your answers to: m.abd1391@gmail.com

Note 1: *This exercise aims to introduce data management in Stata.*

Note 2: *You should submit your do file and a pdf file that contains summary statistics table.*

A Brief Introduction to Datasets Used

Please download the data by clicking here.

In this assignment, we use three different datasets: A quarterly bank-level dataset for US commercial banks that contains banks financial statements items, and two quarterly state-level data for home price index and personal income growth. Our data covers the period from beginning of 2003 to the end of 2014.

There are 15 excel files in the data folder. 12 files are for Banks data that their names start with **US_DATA**. **HPI_EXP_state.xls** contains home price index and **personal_income_growth.xls** contains personal income growth data. There is also an excel file named **state_abbrev.xlsx** that contains US states names abbreviations list.

We want to import the data, reshape it in desired format, merge datasets, define some new variables and get summary statistics. In each step we saved the changed dataset. We also want to record all Stata commands and output in this session using a log file.

Introduction to Panel Data

As you know, there are three types of economic data: cross-sectional data, time series and panel(or longitudinal) data. A panel dataset consists of a time series for each cross-sectional member in a dataset. Here we have panel datasets. The bank-level data and the personal income growth are in *wide* format and the home price index data is in *long* format. (You can read about wide and long format in Stata here)

Data Management in Stata

Define Paths

1. In your main working directory, create three folders with names: ***DATA***, ***DO*** and ***OUT***. Insert the excel files in ***DATA***. Write your do file in ***DO***. ***OUT*** folder is for your output table and log files.

Importing Bank-Level Data

1. Open a text log file named **data_management_session** for recording what we will be done in this session (Use `log` command).

2. Define a local macro for all **.xlsx** bank-level data files (As mentioned above, their names start with **US_DATA**). Using for loop in Stata, import each excel file and save the data in **.dta** format with same name (in ***DATA***).

Merging Data

1. Now Use **US_DATA_1_Assets_1.dta** as master dataset and merge it with other bank-level datasets based on **SNL_ID**.(Like the previous step, do that in a for loop on datasets you saved there)
2. Save the result in **US_DATA_1.dta**.

Reshaping Data

1. Reshape **US_DATA_1.dta** to long, based on **SNL_ID** as `idvar` and **quarter** as `tvar`.

Dropping Extra Variables

1. Drop theses variables: **GL**, **GM**, **GN**, **EP**, **-** and **GK** .
2. Save the result in **US_DATA_3.dta**.

Importing State-Level Data: Home Price Index

1. Import the excel file **HPI_EXP_state.xls** and save it in **.dta** format.

Generating Matched Variables and Merging State-Level Home Price Index

1. Declare **US_DATA_3.dta** dataset to be panel by `xtset` command. As mentioned above, we have 48 quarters. quarter 1 refers to 2003Q1 and quarter 48 refers to 2014Q4.
2. Now based on **quarter** values, define a new variable **qtr** that gets values 1, 2,3 or 4. For example for **quarter==45**, we have **qtr=1**. (Using `dofq` and `quarter` commands can be helpful).
3. Also define variable **year** based on the **quarter**.
4. Generate a new variable **state** that is equal to **STATE**.
5. Save the result in **US_DATA_4.dta**

6. Merge this dataset with home price index based on **state**, **year** and **qtr**.

7. Save the merged dataset in **US_DATA_5.dta**.

Importing State-Level Data: Personal Income Growth

1. Import the excel file **personal_income_growth.xls**.

2. Reshape the data to long with **STATE** and **quarter** as identifier.

3. Save this dataset in **.dta** format.

Merging State-Level Data

1. Now use **US_DATA_5** as master dataset and merge with **personal_income_growth.dta** matching on **STATE** and **quarter**.

2. Save the result dataset in **US_DATA_6.dta**.

Cleaning the Dataset and Constructing Variables

Now we want to define some new variable. Some of them are commonly used bank-specific ratios that their definition can be found here.

1. Declare **US_DATA_6.dta** to be panel data to Stata.

2. Drop observations with missing **SNL_ID** value.

3. We want to limit our sample to 2010-2014; So drop the observations before 2010.

4. Save the result in **US_DATA_7.dta**.

5. Using **US_DATA_7.dta**, define a new variable **Loan_Asset_Ratio** that is equal to net loan leases to total assets ratio (in percentage).

6. Define a new variable **Avg_Loan** that is equal to the average to net loan leases over one previous quarter (You can use lag operator in Stata).

7. Define a new variable **LLPAGL** that is equal to loan loss provision to average loans ratio (as percentage).

8. Define a new variable **Loan_Growth** as the net loan leases growth rate relative to the past quarter (You can use lag operator in Stata).

9. Define a new variable **Capital_Asset_Ratio** that is equal to total equity capital to total assets ratio (as percentage).

10. Define a variable **Inefficiency** as the ratio of non-interest expense to net income (as percentage).

Summary Statistics

1. Summarize the variables you defined in previous step. Report number of observations, mean, standard deviation and median. We highly recommand you to use **estout** package and `estpost` and `estout` commands for generating tables in tex or word format.

2. Close the log file.