

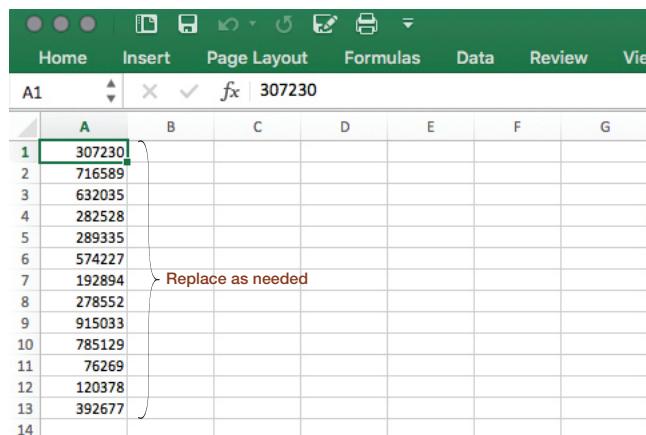
Instructions for Chapter 9 Class Assessment Data Generation

We provide an R script (*script_DataGeneration.R*) and an example set of student IDs (*studentid.csv*) to help you with generating data sets for the class assessment in Chapter 9 (pages 320–322).

With these two files, one can generate separate data sets for each student to attempt the class assignment. The true underlying models for Y1 and Y2 are the same for all data sets (reported in lines 11-16 in *script_DataGeneration.R*), however the data will look different, due to added random noise, for each student. Note that sampling uncertainty will affect the estimated coefficients, however the data generator checks that when the complete training sample (49 observations) is used, the variables of the true model can be identified at 5% significance level. Other variables (up to second lag) will be found to be insignificant.

To generate the data sets, first, adjust file *studentid.csv* to contain IDs for each student. We include an example *studentid.csv* with 13 ID numbers. Replace as appropriate. IDs do not have to be numeric, although only valid characters for file names should be used. Each student ID should be in a single row on the first column, as shown in Figure 1. Save the file as *csv* and proceed to the next step to generate the data sets using the *script_DataGeneration.R* file.

Figure 1 Screenshot of example *studentid.csv* file. It contains 13 student IDs, one at each row.



	A	B	C	D	E	F	G
1	307230						
2	716589						
3	632035						
4	282528						
5	289335						
6	574227						
7	192894						
8	278552						
9	915033						
10	785129						
11	76269						
12	120378						
13	392677						
14							

Load the R script and execute it. The *studentid.csv* should be in its path (placing both files in the R folder of R should suffice). The script will automatically generate as many *csv* files as the number of provided student IDs and store them in the same directory. (The path that the files are saved is printed on screen by the script.)

Note that the script runs several checks to ensure that the data sets will have different data, yet similar solution (the true data generating model can be identified). This can be computationally expensive, as any data sets that do not pass all tests will be rejected and new ones will be generated. The progress is reported on screen. Once the execution of the script is completed, the relevant data sets are stored as *csv* files. Each file contains a data set for a single student, containing the values of Y1, Y2, X1, X2 and X3 for the 49 in-sample periods. The out-of-sample values for X1, X2 and X3 are also provided (14 periods).

Finally, the R script can be easily adjusted to produce data sets with different data generating processes, in terms of coefficients, standard error and variables included.