Homework for Chapter 15: Simulation

1. Which of the following describes why a simulation can be used to check whether a given estimator produces the correct result on average?
   1. Because we created the data, so know what the correct result is to check against
   2. Because if the simulation produces the same result consistently, that’s a sign that it is producing the correct result
   3. Because we can select the estimator and know the properties of that estimator, and whether it will produce a correct result given the data generating process we chose
   4. A simulation can’t be used for this purpose
2. You have created a data-generating function where the variable is created using the equation . You pick an estimation method, a new matching estimator you think might work, and test it using a simulation with that data-generating function. Your average estimate from the simulation is that a one-unit increase in increases by 3. What conclusion can you draw from the results of your simulation?
   1. The true effect of a one-unit increase in on is 3
   2. The true effect of a one-unit increase in on is 2
   3. There is a back door between and
   4. The estimator is incorrect on average
3. What problems would arise if we chose to run our simulation only once rather than many, many times?
4. Finish these sentences by filling in the appropriate power analysis-related term from the chapter:
   1. If we have a sample size of A, there is B variation in our treatment variable, C non-treatment-related variation in the outcome, and we want to have D statistical power, then we need the true effect of treatment to be at least as large as the \_\_\_\_
   2. If we have A variation in our treatment variable, B non-treatment-related variation in the outcome, want to have C statistical power, and think the true effect of treatment is D, then we need our sample size to be at least as large as the \_\_\_\_
   3. If our estimator performs a hypothesis rest against a null value, and the true value in the data-generating distribution is not equal to the null, then the proportion of the time that the estimator rejects the null is the \_\_\_\_
   4. If we want to know how big our sample size needs to be to get a certain level of statistical power, it’s important that we have a guess about how big the \_\_\_\_\_ is, because the closer that thing is to zero, the harder it will be to distinguish that thing from zero.
5. Why do we need to change our bootstrap resampling method when we’re faced with non-independent data like panel data or autocorrelation? You’ll want to use the phrase “sampling distribution” in your answer.
6. Coding